Algebraic & Geometric Topology

Volume 17 (2017)

Issue 3 (pages 1283–1916)

msp

ALGEBRAIC & GEOMETRIC TOPOLOGY

msp.org/agt

See inside back cover or msp.org/agt for submission instructions.

The subscription price for 2017 is US $410/year for the electronic version, and $650/year (+$60, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP. Algebraic & Geometric Topology is indexed by Mathematical Reviews, Zentralblatt MATH, Current Mathematical Publications and the Science Citation Index.

Algebraic & Geometric Topology (ISSN 1472-2747 printed, 1472-2739 electronic) at Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published 6 times per year and continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840.

AGT peer review and production are managed by EditFLOW® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

Algebraic & Geometric Topology 17 (2017) 1283–1321

Grid diagrams and Manolescu's unoriented skein exact triangle for knot Floer homology

C-M MICHAEL WONG

We rederive Manolescu's unoriented skein exact triangle for knot Floer homology over $\mathbb{F}_2$ combinatorially using grid diagrams, and extend it to the case with $\mathbb{Z}$ coefficients by sign refinements. Iteration of the triangle gives a cube of resolutions that converges to the knot Floer homology of an oriented link. Finally, we reestablish the homological σ–thinness of quasialternating links.

57R58; 57M25, 57M27

1 Introduction

Heegaard Floer homology was first introduced by Ozsváth and Szabó [16] as an invariant for 3–manifolds, defined using holomorphic disks and Heegaard diagrams. It was extended by Ozsváth and Szabó [15], and independently by Rasmussen [19], to give an invariant, *knot Floer homology*, for nullhomologous knots in a closed, oriented 3–manifold, which was further generalized by Ozsváth and Szabó [18] to the case of oriented links. Knot Floer homology comes in several flavors; its most usual form, $\widehat{\mathrm{HFK}}(L)$ for an oriented link L, is a bigraded module over $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$ or $\mathbb{Z}$, whose Euler characteristic is the Alexander polynomial. For the purposes of this paper, we shall only consider links in S^3.

A combinatorial description of knot Floer homology over $\mathbb{F}_2$ was given by Manolescu, Ozsváth and Sarkar [11], and Manolescu, Ozsváth, Szabó and Thurston [12], using *grid diagrams*, which are certain multipointed Heegaard diagrams on the torus. In this approach, one can associate a chain complex $\widetilde{\mathrm{GC}}(\mathbb{G})$ to a grid diagram $\mathbb{G}$, and calculate its homology $\widetilde{\mathrm{GH}}(\mathbb{G})$. Sign refinements for the boundary map ∂ are also given in [12] in a well-defined manner, allowing the chain complex to be defined over $\mathbb{Z}$. If $\mathbb{G}$ is a grid diagram for a link L of ℓ components, with *grid number* n, then

$$\widetilde{\mathrm{GH}}(\mathbb{G}) \cong \widehat{\mathrm{GH}}(L) \otimes V^{n-l},$$

where V is a free module of rank 2 over the base ring $R = \mathbb{F}_2$ or $\mathbb{Z}$, and $\widehat{\mathrm{GH}}(L)$ is a link invariant, called the *combinatorial knot Floer homology* or the *grid homology*

Published: 17 July 2017 DOI: 10.2140/agt.2017.17.1283

of L. Over $\mathbb{F}_2$, $\widehat{\mathrm{GH}}(L)$ is isomorphic to $\widehat{\mathrm{HFK}}(L)$; over $\mathbb{Z}$, it has been shown by Sarkar [21] that $\widehat{\mathrm{GH}}(L)$ is isomorphic to $\widehat{\mathrm{HFK}}(L, \mathfrak{o})$ for some orientation system $\mathfrak{o}$.

Ozsváth and Szabó [17] observed that Heegaard Floer homology of the branched double cover $\widehat{\mathrm{HF}}(-\Sigma(L))$, like Khovanov homology $\widetilde{\mathrm{Kh}}(L)$ (see Khovanov [6; 7] or Bar-Natan [3]), satisfies an unoriented skein exact triangle. Manolescu [9] then showed that over $\mathbb{F}_2$, knot Floer homology also satisfies an unoriented skein exact triangle. More precisely, let L_∞ be a link in S^3. Given a planar diagram of L_∞, let L_0 and L_1 be the two resolutions of L_∞ at a crossing in that diagram, as in Figure 1. Denote by ℓ_∞, ℓ_0 and ℓ_1 the number of components of the links L_∞, L_0 and L_1 respectively, and set $m = \max\{\ell_\infty, \ell_0, \ell_1\}$.

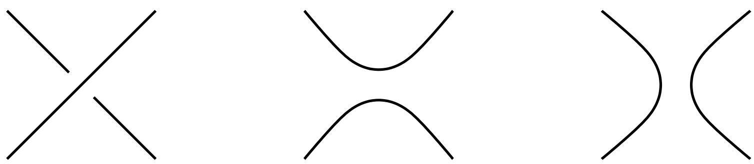

Figure 1: L_∞, L_0 and L_1 near a point

Theorem 1.1 (Manolescu) *There exists an exact triangle*

$$\cdots \to \widehat{\mathrm{HFK}}(L_\infty; \mathbb{F}_2) \otimes V^{m-\ell_\infty} \to \widehat{\mathrm{HFK}}(L_0; \mathbb{F}_2) \otimes V^{m-\ell_0} \\ \to \widehat{\mathrm{HFK}}(L_1; \mathbb{F}_2) \otimes V^{m-\ell_1} \to \cdots,$$

where V is a vector space of dimension 2 over $\mathbb{F}_2$.

Remark 1.2 The arrows in the exact triangle point in the reverse direction from those in [9]; this is caused by a difference in the orientation convention. We follow the convention in [15] and [11; 12], where the Heegaard surface is the oriented boundary of the handlebody in which the α curves bound discs.

Manolescu observed that the exact triangle above is different from that of $\widetilde{\mathrm{Kh}}(L)$ and $\widehat{\mathrm{HF}}(-\Sigma(L))$ in that it does not even respect the homological grading modulo 2, and that it is unclear whether an analogous triangle holds for other versions of knot Floer homology. He also used the exact triangle to show that $\operatorname{rk} \widehat{\mathrm{HFK}}(L; \mathbb{F}_2) = 2^{\ell-1} \det(L)$ for quasialternating links, which explains the fact that $\widetilde{\mathrm{Kh}}$ and $\widehat{\mathrm{HFK}}$ have equal ranks for many classes of knots.

The goal of the present paper is to reprove Manolescu's theorem in elementary terms using grid diagrams, without appealing to the topological theory. The advantages of this approach are threefold.

First, Manolescu's skein exact triangle was proven in [9] to exist over $\mathbb{F}_2$. By assigning signs to the maps between chain complexes, we can obtain an analogous exact triangle in combinatorial knot Floer homology with $\mathbb{Z}$ coefficients, which was not known to exist before. In other words, we obtain the following statement.

Theorem 1.3 *For sufficiently large n, there exists an exact triangle*

$$\cdots \to \widehat{\mathrm{GH}}(L_\infty; R) \otimes V^{n-\ell_\infty} \to \widehat{\mathrm{GH}}(L_0; R) \otimes V^{n-\ell_0} \to \widehat{\mathrm{GH}}(L_1; R) \otimes V^{n-\ell_1} \to \cdots,$$

where $R = \mathbb{F}_2$ or $\mathbb{Z}$, and V is a free module of rank 2 over R.

Second, the exact triangle was iterated by Baldwin and Levine [2] to obtain a cube of resolutions; when using twisted coefficients, this gives a combinatorial description of knot Floer homology, distinct from that provided by grid diagrams. In the present context, knowing explicitly the maps between the chain complexes associated to the grid diagrams, we can likewise iterate the exact triangle to get a cube of resolutions complex $\mathrm{CR}(\mathbb{G})$ over $\mathbb{F}_2$, with untwisted coefficients. The higher terms in the resulting spectral sequence are combinatorially computable.

Corollary 1.4 *The cube of resolutions $\mathrm{CR}(\mathbb{G}; \mathbb{F}_2)$, which has no diagonal maps, gives rise to a spectral sequence that converges to $\widehat{\mathrm{GH}}(L; \mathbb{F}_2) \otimes V^{m-\ell}$.*

Note that the spectral sequence is presumably not a knot invariant; see [2, Remark 7.7].

The technique of spectral sequences was first used by Ozsváth and Szabó in [17], where a spectral sequence from $\widetilde{\mathrm{Kh}}(L)$ to $\widehat{\mathrm{HF}}(-\Sigma(L))$ was shown to exist. More recently, Lipshitz, Ozsváth and Thurston [8] have found a way to compute the higher terms in this spectral sequence using bordered Floer homology. Inspired by this work, Baldwin [1] has found another method of computing these higher terms.

Third, there exists a δ–*grading* on $\widehat{\mathrm{HFK}}$, and Manolescu and Ozsváth [10] investigated the δ–grading changes in the skein exact triangle [10, Proposition 3.9]. This result allowed them to apply the skein exact triangle to *quasialternating links*, to show that such links are *Floer homologically σ–thin* over $\mathbb{F}_2$. The δ–gradings can also be determined in the combinatorial picture; doing so, we prove a generalization of the statement of Floer homological σ–thinness to $\mathbb{Z}$.

Theorem 1.5 *Suppose that $\det(L_0)$, $\det(L_1) > 0$ and $\det(L_\infty) = \det(L_0) + \det(L_1)$. Then with respect to the δ–grading, the exact sequence in Theorem 1.3 can be written as*

$$\begin{aligned}\cdots \to \widehat{\mathrm{GH}}_{*-\frac{1}{2}\sigma(L_1)}(L_1; R) \otimes V^{n-\ell_1} &\to \widehat{\mathrm{GH}}_{*-\frac{1}{2}\sigma(L_\infty)}(L_\infty; R) \otimes V^{n-\ell_\infty} \\ &\to \widehat{\mathrm{GH}}_{*-\frac{1}{2}\sigma(L_0)}(L_0; R) \otimes V^{n-\ell_0} \to \widehat{\mathrm{GH}}_{*-\frac{1}{2}\sigma(L_1)+1}(L_1; R) \otimes V^{n-\ell_1} \to \cdots,\end{aligned}$$

where $R = \mathbb{F}_2$ or $\mathbb{Z}$, and V is a free module of rank 2 over R with grading zero.

Theorem 1.6 *Quasialternating links are Floer homologically σ–thin over $\mathbb{Z}$.*

This paper is organized as follows. We review the definition of knot Floer homology in terms of grid diagrams in Section 2, and reprove the skein exact triangle in Section 3. In these two sections, we will work only over $\mathbb{F}_2$. Sign refinements are then given in Section 4 to establish the analogous result over $\mathbb{Z}$. Next, we discuss how the exact triangle can be iterated to obtain a cube of resolutions over $\mathbb{F}_2$ in Section 5. Finally, we establish the homological σ–thinness of quasialternating links over $\mathbb{F}_2$ and $\mathbb{Z}$ in Section 6.

Acknowledgements The author is very grateful to John Baldwin for his suggestion of the topic of the paper, and is indebted to John for many essential ideas in the text. He also thanks John Baldwin, Robert Lipshitz and Peter Ozsváth for their guidance. He thanks Ciprian Manolescu for a helpful conversation, and Gahye Jeong for pointing out a previously missing case in the proof of Lemma 3.6. Lastly, he thanks the referee for remarkably thorough and useful comments, and for pointing out a mistake in the statement of the main theorem, Theorem 1.3, in an earlier version.

The author was supported in part by the Princeton University Mathematics Department.

2 Grid diagrams

We review the combinatorial description of knot Floer homology in terms of grid diagrams. In this and the next section, we will work only over $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$.

A *planar grid diagram* $\widetilde{\mathbb{G}}$ with *grid number* n is a square grid in $\mathbb{R}^2$ with $n \times n$ cells, together with a collection of Os and Xs, such that

- each row contains exactly one O and exactly one X;
- each column contains exactly one O and exactly one X; and
- each cell is either empty, contains one O, or contains one X.

Given a planar grid diagram $\widetilde{\mathbb{G}}$, we can place it in a standard position on $\mathbb{R}^2$ as follows. We place the bottom left corner at the origin, and require that each cell be a square of edge length one. We can then construct an oriented, planar link projection by drawing horizontal segments from the Os to the Xs in each row, and vertical segments from the Xs to the Os in each column. At every intersection point, we let the horizontal segment be the underpass and the vertical one the overpass. This gives a planar projection of an oriented link L onto $\mathbb{R}^2$; we say that $\widetilde{\mathbb{G}}$ is a *planar grid presentation* of L.

We transfer our grid diagram to the torus $\mathbb{T}$, by gluing the topmost segment to the bottommost one, and gluing the leftmost segment to the rightmost one. Then the horizontal and vertical arcs become horizontal and vertical circles. The torus inherits

its orientation from the plane. The resulting diagram $\mathbb{G}$ is a *toroidal grid diagram*, or simply a *grid diagram*. $\mathbb{G}$ is then a *grid presentation* of L; we also say that $\mathbb{G}$ is a grid diagram for L.

Given a toroidal grid diagram $\mathbb{G}$, we associate to it a chain complex $(\widetilde{\mathrm{GC}}(\mathbb{G}), \partial)$ as follows. The set of generators of $\widetilde{\mathrm{GC}}(\mathbb{G})$, denoted $\boldsymbol{S}(\mathbb{G})$, consists of one-to-one correspondences between the horizontal circles and vertical circles. Equivalently, we can regard the generators as n–tuples of intersection points between the horizontal and vertical circles, such that no intersection point appears on more than one horizontal or vertical circle.

We now define the differential map $\partial\colon \widetilde{\mathrm{GC}}(\mathbb{G}) \to \widetilde{\mathrm{GC}}(\mathbb{G})$. Given $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{S}(\mathbb{G})$, let $\mathrm{Rect}(\boldsymbol{x}, \boldsymbol{y})$ denote the space of embedded rectangles with the following properties. First of all, $\mathrm{Rect}(\boldsymbol{x}, \boldsymbol{y})$ is empty unless $\boldsymbol{x}, \boldsymbol{y}$ coincide at exactly $n-2$ points. An element r of $\mathrm{Rect}(\boldsymbol{x}, \boldsymbol{y})$ is an embedded disk in $\mathbb{T}$, whose boundary consists of four arcs, each contained in horizontal or vertical circles; under the orientation induced on the boundary of r, the horizontal arcs are oriented from a point in $\boldsymbol{x}$ to a point in $\boldsymbol{y}$. The space of empty rectangles $r \in \mathrm{Rect}(\boldsymbol{x}, \boldsymbol{y})$ with $\boldsymbol{x} \cap \mathrm{Int}(r) = \varnothing$ is denoted $\mathrm{Rect}^\circ(\boldsymbol{x}, \boldsymbol{y})$.

More generally, a *path* from $\boldsymbol{x}$ to $\boldsymbol{y}$ is a 1–cycle γ on $\mathbb{T}$ such that the boundary of the intersection of γ with the union of the horizontal curves is $\boldsymbol{y} - \boldsymbol{x}$, and a *domain* p from $\boldsymbol{x}$ to $\boldsymbol{y}$ is a two-chain in $\mathbb{T}$ whose boundary ∂p is a path from $\boldsymbol{x}$ to $\boldsymbol{y}$; the set of domains from $\boldsymbol{x}$ to $\boldsymbol{y}$ is denoted $\pi(\boldsymbol{x}, \boldsymbol{y})$.

Given $\boldsymbol{x} \in \boldsymbol{S}(\mathbb{G})$, we define

$$\partial(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \boldsymbol{S}(\mathbb{G})} \sum_{\substack{r \in \mathrm{Rect}^\circ(\boldsymbol{x}, \boldsymbol{y}) \\ \mathrm{Int}(r) \text{ contains no } O\text{s or } X\text{s}}} \boldsymbol{y} \in \widetilde{\mathrm{GC}}(\mathbb{G}).$$

It is not too difficult to see that indeed $\partial \circ \partial = 0$, and so ∂ is a differential: we have

$$\partial \circ \partial(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \boldsymbol{S}(\mathbb{G})} \sum_{p \in \pi(\boldsymbol{x}, \boldsymbol{z})} N(p) \cdot \boldsymbol{z},$$

where $N(p)$ is the number of ways of decomposing p as a composite of two empty rectangles $p = r_1 * r_2$ with $r_1 \in \mathrm{Rect}^\circ(\boldsymbol{x}, \boldsymbol{y})$ and $r_2 \in \mathrm{Rect}^\circ(\boldsymbol{y}, \boldsymbol{z})$. Let $p = r_1 * r_2$; then r_1 and r_2 either are disjoint, have overlapping interiors or share a corner. If r_1 and r_2 are disjoint or have overlapping interiors, then $p = r_2 * r_1$; if they share a corner, then there exists a unique alternate decomposition of $p = r_1' * r_2'$. In any case, we obtain that $N(p) = 0$ for all $p \in \pi(\boldsymbol{x}, \boldsymbol{z})$.

Moreover, to the complex $\widetilde{\mathrm{GC}}(\mathbb{G})$ we can associate a *Maslov grading* and an *Alexander grading*, determined by the functions $M\colon \boldsymbol{S} \to \mathbb{Z}$ and $S\colon \boldsymbol{S} \to \frac{1}{2}\mathbb{Z}$. For reasons we will see in the next section, we will in general not be concerned with these gradings, unless otherwise specified. We postpone their definitions to Section 6. It can be checked, however, that the differential ∂ decreases the Maslov grading by 1 and preserves the Alexander grading.

We can now take the homology of the chain complex $(\widetilde{\mathrm{GC}}(\mathbb{G}), \partial)$, and define

$$\widetilde{\mathrm{GH}}(\mathbb{G}) = \mathrm{H}_*(\widetilde{\mathrm{GC}}(\mathbb{G}), \partial).$$

It is shown in [11; 12] that, if $\mathbb{G}$ is a grid diagram with grid number n for the oriented link L with ℓ components, then

$$\widetilde{\mathrm{GH}}(\mathbb{G}) \cong \widehat{\mathrm{GH}}(L) \otimes V^{n-\ell},$$

where V is a 2–dimensional vector space over $\mathbb{F}_2$, spanned by one generator in (Maslov and Alexander) bigrading $(-1,-1)$ and another in bigrading $(0,0)$, and $\widehat{\mathrm{GH}}(L)$ is a link invariant, often referred to as the *combinatorial knot Floer homology*, or *grid homology*, that is a vector space isomorphic to $\widehat{\mathrm{HFK}}(L)$. $\widetilde{\mathrm{GH}}(\mathbb{G})$ can also be denoted by $\widetilde{\mathrm{GH}}(L,n)$.

Remark 2.1 While the proof that ∂ is a differential is completely elementary, its method is very useful to what we shall prove in this paper. In general, in order to prove that a map defined by counting certain domains is a chain map, or to prove that it is a chain homotopy, we juxtapose two domains and enumerate all possible outcomes.

3 Manolescu's unoriented skein exact triangle

We now prove our main result over $\mathbb{F}_2$ in purely combinatorial terms.

Before we start our main discussion, we make a change in our notation. In the original description developed in [11; 12], the Os and Xs were used to determine the Alexander and Maslov gradings of the generators. However, given a link L_∞ in S^3 and its two resolutions L_0, L_1 at a crossing, we observe that L_∞, L_0, L_1 do not share a compatible orientation. Since we shall soon combine all three grid diagrams into one, we must forget the orientations of the links; this implies that we must also forget the distinctions between the Os and Xs, and ignore the gradings. Notice that using markers of only one type will not change the definition of $\widetilde{\mathrm{GH}}(L,n)$, since without the gradings, the definition of the chain complex $\widetilde{\mathrm{GC}}(\mathbb{G})$ associated to a grid diagram G is symmetric in the Os and the Xs. Therefore, we shall henceforth replace all Os with Xs.

With this new notation, the first two conditions for a grid diagram become the condition that there are exactly two Xs on each row and each column. We denote by $\mathbb{X}$ the set of Xs on a grid diagram. The differential is then given by

$$\partial(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in\boldsymbol{S}(\mathbb{G})} \sum_{\substack{r\in\mathrm{Rect}^\circ(\boldsymbol{x},\boldsymbol{y})\\ \mathrm{Int}(r)\cap\mathbb{X}=\varnothing}} \boldsymbol{y} \in \widetilde{\mathrm{GC}}(\mathbb{G}).$$

To begin, we position L_∞, L_0, L_1 as in Figure 2, and make sure that their respective grid diagrams $\mathbb{G}_\infty, \mathbb{G}_0, \mathbb{G}_1$ are identical except near the crossing, as indicated in the same figure. Next, we let $C_k = \widetilde{\mathrm{GC}}(\mathbb{G}_k)$ be the chain complex associated with $\mathbb{G}_k$, for each $k \in \{\infty, 0, 1\}$. We endow the set $\{\infty, 0, 1\}$ with an action by $\mathbb{Z}/3\mathbb{Z}$ by identifying ∞ with 2, so that $\infty + 1 = 0$ and $1 + 1 = \infty$.

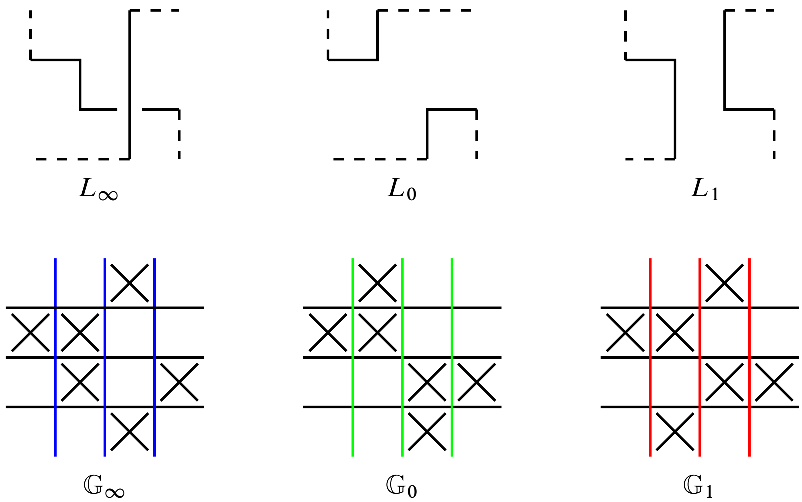

Figure 2: Grid diagrams for L_∞, L_0 and L_1 near a point

Remark 3.1 The leftmost and rightmost columns in each of the diagrams in Figure 2 do not need to be in the form displayed; the markers can be in any row in those two columns, and the proofs in this and the next section do not rely on the positions of the markers. In the figures in this section, we leave the markers there for ease of visualization. However, in Section 5, it will be necessary to have the markers exactly as they appear in Figure 2, to iterate the exact triangle.

Instead of drawing three different diagrams, we can draw $\mathbb{G}_\infty, \mathbb{G}_0$, $\mathbb{G}_1$ all on the same diagram, as in Figure 3. We label by β_k the vertical circle corresponding to $\mathbb{G}_k$ for each $k \in \{\infty, 0, 1\}$, as indicated. These three vertical circles, together with all the other vertical circles, divide $\mathbb{T}$ into a number of components; we let b be the unique component that is an annulus not containing any X in its interior. Also, exactly three of these components are embedded triangles not containing any X in their interior;

we let t_k be the triangle whose β_k arc is disjoint from the boundary of b. Finally, β_k and β_{k+1} intersect at exactly two points; we denote by u_k the intersection point that lies on the boundary of b, and by v_k the other intersection point. Then $u_k = b \cap t_{k+2}$ and $v_k = t_k \cap t_{k+1}$.

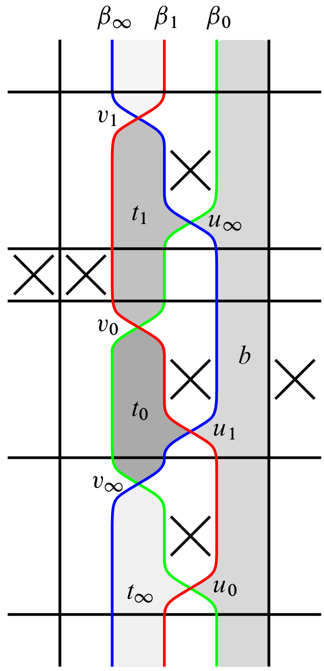

Figure 3: Combined grid diagram for L_∞, L_0 and L_1 near a point. The annulus b and the triangles t_k are shaded. The circles β_k and β_{k+1} intersect at two points; u_k are the ones on the right, and v_k are those on the left.

In this setting and over $\mathbb{F}_2$, Theorem 1.3 follows from our main proposition:

Proposition 3.2 *There exists an exact triangle*

$$\cdots \to \widetilde{\mathrm{GH}}(\mathbb{G}_\infty;\mathbb{F}_2) \to \widetilde{\mathrm{GH}}(\mathbb{G}_0;\mathbb{F}_2) \to \widetilde{\mathrm{GH}}(\mathbb{G}_1;\mathbb{F}_2) \to \cdots .$$

We make use of the following lemma from homological algebra, first used by Ozsváth and Szabó [17], and used also in [9].

Lemma 3.3 *Let $\{(C_k, \partial_k)\}_{k\in\{\infty,0,1\}}$ be a collection of chain complexes over an arbitrary commutative ring, and let $\{f_k\colon C_k \to C_{k+1}\}_{k\in\{\infty,0,1\}}$ be a collection of anti-chain maps such that the following conditions are satisfied for each k:*

(1) *The composite* $f_{k+1} \circ f_k \colon C_k \to C_{k+2}$ *is chain-homotopic to zero, by a chain homotopy* $\varphi_k \colon C_k \to C_{k+2}$*:*

$$f_{k+1} \circ f_k + \partial_{k+2} \circ \varphi_k + \varphi_k \circ \partial_k = 0.$$

(2) *The map* $\varphi_{k+1} \circ f_k + f_{k+2} \circ \varphi_k \colon C_k \to C_k$ *is a quasi-isomorphism. (In particular, if there exists a chain homotopy* $\psi_k \colon C_k \to C_k$ *such that*

$$\varphi_{k+1} \circ f_k + f_{k+2} \circ \varphi_k + \partial_k \circ \psi_k + \psi_k \circ \partial_k = \mathrm{Id},$$

then this condition is satisfied.)

Then we have an exact sequence

$$\cdots \to \mathrm{H}_*(C_k) \xrightarrow{(f_k)_*} \mathrm{H}_*(C_{k+1}) \xrightarrow{(f_{k+1})_*} \mathrm{H}_*(C_{k+2}) \to \cdots .$$

We now define the chain maps $f_k \colon C_k \to C_{k+1}$ by counting pentagons and triangles.

Given $\boldsymbol{x} \in \boldsymbol{S}(\mathbb{G}_k)$ and $\boldsymbol{y} \in \boldsymbol{S}(\mathbb{G}_{k+1})$, let $\mathrm{Pent}_k(\boldsymbol{x}, \boldsymbol{y})$ denote the space of embedded pentagons with the following properties. First of all, $\mathrm{Pent}_k(\boldsymbol{x}, \boldsymbol{y})$ is empty unless $\boldsymbol{x}, \boldsymbol{y}$ coincide at exactly $n-2$ points. An element p of $\mathrm{Pent}_k(\boldsymbol{x}, \boldsymbol{y})$ is an embedded disk in $\mathbb{T}$, whose boundary consists of five arcs, each contained in horizontal or vertical circles; under the orientation induced on the boundary of p, we start at the β_k–component of $\boldsymbol{x}$, traverse the arc of a horizontal circle, meet its corresponding component of $\boldsymbol{y}$, proceed to an arc of a vertical circle, meet the corresponding component of $\boldsymbol{x}$, continue through another horizontal circle, meet the component of $\boldsymbol{y}$ contained in β_{k+1}, proceed to an arc in β_{k+1}, meet the intersection point u_k of β_k and β_{k+1}, and finally, traverse an arc in β_k until we arrive back at the initial component of $\boldsymbol{x}$. Notice that all angles here are at most straight angles. The space of empty pentagons $p \in \mathrm{Pent}_k(\boldsymbol{x}, \boldsymbol{y})$ with $\boldsymbol{x} \cap \mathrm{Int}(p) = \varnothing$ is denoted $\mathrm{Pent}^\circ_k(\boldsymbol{x}, \boldsymbol{y})$.

Similarly, we let $\mathrm{Tri}_k(\boldsymbol{x}, \boldsymbol{y})$ denote the space of embedded triangles with the following properties. $\mathrm{Tri}_k(\boldsymbol{x}, \boldsymbol{y})$ is empty unless $\boldsymbol{x}, \boldsymbol{y}$ coincide at exactly $n-1$ points. An element p of $\mathrm{Tri}_k(\boldsymbol{x}, \boldsymbol{y})$ is an embedded disk in $\mathbb{T}$, whose boundary consists of three arcs, each contained in horizontal or vertical circles; under the orientation induced on the boundary of p, we start at the β_k–component of $\boldsymbol{x}$, traverse the arc of a horizontal circle, meet the component of $\boldsymbol{y}$ contained in β_{k+1}, proceed to an arc in β_{k+1}, meet the intersection point v_k of β_k and β_{k+1}, and finally traverse an arc in β_k to return to the initial component of $\boldsymbol{x}$. Again, all the angles here are less than straight angles. It is clear that all triangles $p \in \mathrm{Tri}_k(\boldsymbol{x}, \boldsymbol{y})$ satisfy $\boldsymbol{x} \cap \mathrm{Int}(p) = \varnothing$. Furthermore, for any generator $\boldsymbol{x}$, there is at most one generator $\boldsymbol{y}$ such that $\mathrm{Tri}_k(\boldsymbol{x}, \boldsymbol{y})$ is not empty.

See Figure 4 for examples of elements of $\mathrm{Pent}^\circ_k(x, y)$ and $\mathrm{Tri}_k(x, y)$.

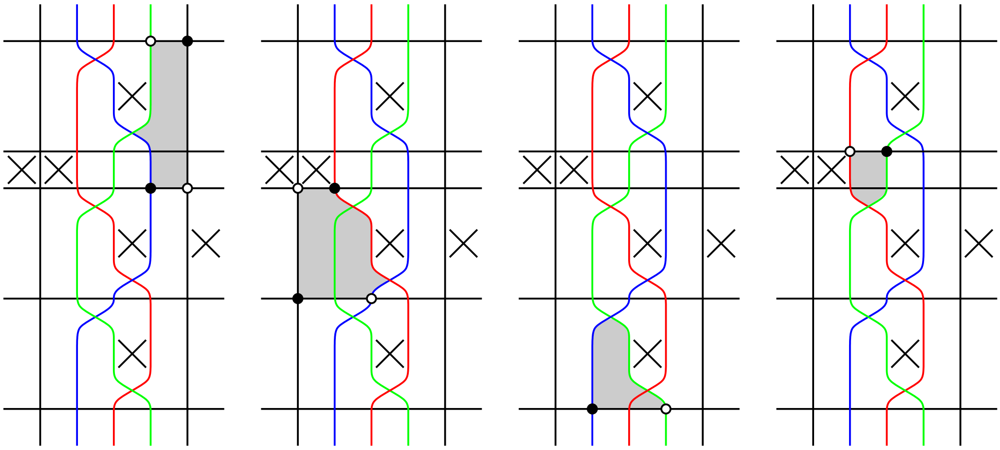

Figure 4: Two allowed pentagons in $\mathrm{Pent}_k^{\circ}(\boldsymbol{x},\boldsymbol{y})$ and two allowed triangles in $\mathrm{Tri}_k(\boldsymbol{x},\boldsymbol{y})$. The components of $\boldsymbol{x}$ are indicated by solid points, and those of $\boldsymbol{y}$ are indicated by hollow ones.

Given $\boldsymbol{x} \in \boldsymbol{S}(\mathbb{G}_k)$, we define elements of C_{k+1},

$$\mathcal{P}_k(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in\boldsymbol{S}(\mathbb{G}_{k+1})} \sum_{\substack{p\in\mathrm{Pent}_k^{\circ}(\boldsymbol{x},\boldsymbol{y}) \\ \mathrm{Int}(p)\cap\mathbb{X}=\varnothing}} \boldsymbol{y},$$

$$\mathcal{T}_k(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in\boldsymbol{S}(\mathbb{G}_{k+1})} \sum_{\substack{p\in\mathrm{Tri}_k(\boldsymbol{x},\boldsymbol{y}) \\ \mathrm{Int}(p)\cap\mathbb{X}=\varnothing}} \boldsymbol{y},$$

$$f_k(\boldsymbol{x}) = \mathcal{P}_k(\boldsymbol{x}) + \mathcal{T}_k(\boldsymbol{x}).$$

Lemma 3.4 *The map f_k is a chain map. In fact, $\mathcal{P}_k$ and $\mathcal{T}_k$ are both chain maps.*

Proof The proof is similar to that of [12, Lemma 3.1]. We consider domains which are obtained as the juxtaposition of a pentagon or a triangle, and a rectangle. There are a few possibilities; in particular, the polygons may be disjoint, their interiors may overlap, or they may share a common corner. If the polygons are disjoint or if their interiors overlap, the domain can be decomposed as either $r * p$ or $p * r$, and so does not contribute to $\partial_{k+1} \circ f_k + f_k \circ \partial_k$. If the polygons share a common corner, the resulting domain always has an alternate decomposition, as shown in Figure 5. In all cases, the domain can be decomposed in two ways, and makes no contribution to $\partial_{k+1} \circ f_k + f_k \circ \partial_k$. Note that each domain has exactly two decompositions, both of which are counted in $\partial_{k+1} \circ f_k + f_k \circ \partial_k$; this is not going to be the case in similar lemmas later. □

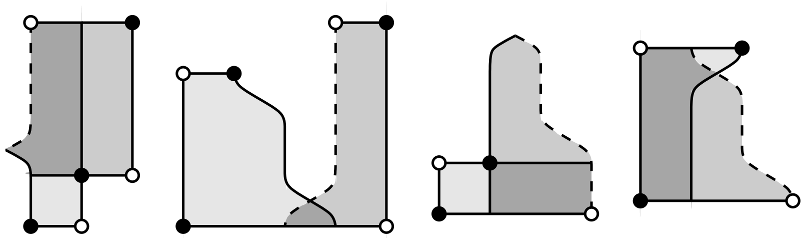

Figure 5: Two typical domains that arise as the juxtaposition of a pentagon and a rectangle, and two that arise as that of a triangle and a rectangle. The β_k curve is solid while the β_{k+1} curve is dotted.

Remark 3.5 In the proof above, every domain that arises as the juxtaposition of a triangle and a rectangle has exactly two decompositions, one contributing to $\partial_{k+1} \circ \mathcal{T}_k$ and one to $\mathcal{T}_k \circ \partial_k$. The analogous statement is not true for $\mathcal{P}_k$.

Henceforth, when considering the composition of two maps that count polygons, we shall ignore the cases where the two polygons are disjoint or have overlapping interiors, since we can always decompose the domain as either $p_1 * p_2$ or $p_2 * p_1$ in these cases.

Next, we define the chain homotopies $\varphi_k \colon C_k \to C_{k+2}$ by counting hexagons and quadrilaterals.

Given $\boldsymbol{x} \in \boldsymbol{S}(\mathbb{G}_k)$ and $\boldsymbol{y} \in \boldsymbol{S}(\mathbb{G}_{k+2})$, let $\mathrm{Hex}_k(\boldsymbol{x}, \boldsymbol{y})$ denote the space of embedded hexagons with the following properties. First, $\mathrm{Hex}_k(\boldsymbol{x}, \boldsymbol{y})$ is empty unless $\boldsymbol{x}$ and $\boldsymbol{y}$ coincide at exactly $n-2$ points. An element p of $\mathrm{Hex}_k(\boldsymbol{x}, \boldsymbol{y})$ is an embedded disk in $\mathcal{T}$, whose boundary consists of six arcs, each contained in horizontal or vertical circles; under the orientation induced on the boundary of p, we start at the β_k–component of $\boldsymbol{x}$, traverse the arc of a horizontal circle, meet its corresponding component of $\boldsymbol{y}$, proceed to an arc of a vertical circle, meet the corresponding component of $\boldsymbol{x}$, continue through another horizontal circle, meet the component of $\boldsymbol{y}$ contained in β_{k+2}, proceed to an arc in β_{k+2}, meet the intersection point u_{k+1} of β_{k+1} and β_{k+2}, traverse an arc in β_{k+1}, meet the intersection point u_k of β_k and β_{k+1}, and finally, traverse an arc in β_k until we arrive back at the initial component of $\boldsymbol{x}$. All the angles here are at most straight angles. The space of empty hexagons $p \in \mathrm{Hex}_k(\boldsymbol{x}, \boldsymbol{y})$ with $\boldsymbol{x} \cap \mathrm{Int}(p) = \varnothing$ is denoted $\mathrm{Hex}^\circ_k(\boldsymbol{x}, \boldsymbol{y})$.

Similarly, we let $\mathrm{Quad}_k(\boldsymbol{x}, \boldsymbol{y})$ denote the space of embedded quadrilaterals with the following properties. $\mathrm{Quad}_k(\boldsymbol{x}, \boldsymbol{y})$ is empty unless $\boldsymbol{x}, \boldsymbol{y}$ coincide at exactly $n-1$ points. An element p of $\mathrm{Quad}_k(\boldsymbol{x}, \boldsymbol{y})$ is an embedded disk in $\mathcal{T}$, whose boundary consists of four arcs, each contained in horizontal or vertical circles; under the orientation induced on the boundary of p, we start at the β_k–component of $\boldsymbol{x}$, traverse the arc

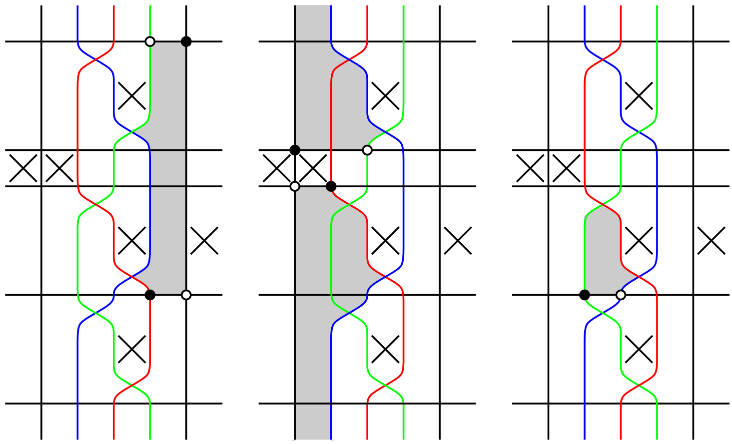

Figure 6: Two allowed hexagons in $\mathrm{Hex}^{\circ}_k(\boldsymbol{x}, \boldsymbol{y})$ and an allowed quadrilateral in $\mathrm{Quad}_k(\boldsymbol{x}, \boldsymbol{y})$. The hexagon in the middle figure is the only allowed empty hexagon that is a left domain.

of a horizontal circle, meet the component of $\boldsymbol{y}$ contained in β_{k+2}, proceed to an arc in β_{k+2}, meet the intersection point u_{k+1} of β_{k+1} and β_{k+2}, proceed to an arc in β_{k+1}, meet the intersection point v_k of β_k with β_{k+1}, and finally traverse an arc in β_k to return to the initial component of $\boldsymbol{x}$. All the angles here are at most straight angles. It is clear that all quadrilaterals $p \in \mathrm{Tri}_k(\boldsymbol{x}, \boldsymbol{y})$ satisfy $\boldsymbol{x} \cap \mathrm{Int}(p) = \varnothing$.

See Figure 6 for examples of elements of $\mathrm{Hex}^{\circ}_k(x, y)$ and $\mathrm{Quad}_k(x, y)$.

Given $\boldsymbol{x} \in \boldsymbol{S}(\mathbb{G}_k)$, we define elements of C_{k+2} by

$$\mathcal{H}_k(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \boldsymbol{S}(\mathbb{G}_{k+2})} \sum_{\substack{p \in \mathrm{Hex}^{\circ}_k(\boldsymbol{x},\boldsymbol{y}) \\ \mathrm{Int}(p) \cap \mathbb{X} = \varnothing}} \boldsymbol{y},$$

$$\mathcal{Q}_k(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \boldsymbol{S}(\mathbb{G}_{k+2})} \sum_{\substack{p \in \mathrm{Quad}_k(\boldsymbol{x},\boldsymbol{y}) \\ \mathrm{Int}(p) \cap \mathbb{X} = \varnothing}} \boldsymbol{y},$$

$$\varphi_k(\boldsymbol{x}) = \mathcal{H}_k(\boldsymbol{x}) + \mathcal{Q}_k(\boldsymbol{x}).$$

We say that a domain p is a *left domain* if $\mathrm{Int}(p) \cap \mathrm{Int}(b) = \varnothing$, and a *right domain* if $\mathrm{Int}(p) \cap \mathrm{Int}(b) \neq \varnothing$.

Lemma 3.6 *The maps f_k and φ_k satisfy condition* (1) *of Lemma 3.3.*

Proof Juxtaposing a triangle and a pentagon appearing in $\mathcal{P}_{k+1} \circ \mathcal{T}_k$, we generically obtain a composite domain that admits a unique alternative decomposition as a quadrilateral and a rectangle, appearing in either $\partial_{k+2} \circ \mathcal{Q}_k$ or $\mathcal{Q}_k \circ \partial_k$, except for one special case, which is described in (1) below. Juxtaposing two pentagons appearing in $\mathcal{P}_{k+1} \circ \mathcal{P}_k$, we generically obtain a composite domain that admits a unique alternative decomposition as a hexagon and a rectangle, appearing in either $\partial_{k+2} \circ \mathcal{H}_k$ or $\mathcal{H}_k \circ \partial_k$, except for one special case, which is described in (3) below.

There are three special cases. We consider domains p arising from juxtapositions of:

(1) A pentagon and a triangle appearing in $\mathcal{T}_{k+1} \circ \mathcal{P}_k$, such that the β_k–component of $\boldsymbol{x}$ does not lie on the boundary of any annular component of $\mathbb{T}$ minus the vertical circles (including β_∞, β_0, β_1). Visually, the β_k–component of $\boldsymbol{x}$ lies on the central vertical axis of the figure. The domain p can only be alternatively decomposed as the triangle t_{k+2} and a pentagon, bounded by β_k, a horizontal arc, a vertical arc, another horizontal arc and β_{k+2}, in its induced orientation. The pentagon is in the opposite orientation as one that would be counted in the map f_{k+2}, and its boundary meets only the intersection point v_{k+2}; such a pentagon is not counted in any map. The triangle t_{k+2} is not counted in any map either. Therefore, p is counted exactly once in $f_{k+1} \circ f_k + \partial_{k+2} \circ \varphi_k + \varphi_k \circ \partial_k$. However, we can replace t_{k+2} with the triangle t_k, and obtain a corresponding domain p' that also connects $\boldsymbol{x}$ to $\boldsymbol{y}$. The new domain p' admits a unique alternative decomposition as a triangle and a pentagon, counted in $\mathcal{P}_{k+1} \circ \mathcal{T}_k$. See Figure 7(1).

(2) A pentagon and a triangle appearing in $\mathcal{T}_{k+1} \circ \mathcal{P}_k$, such that the β_k–component of $\boldsymbol{x}$ lies on the boundary of an annular component of $\mathbb{T}$ minus the vertical circles (including β_∞, β_0, β_1). Visually, the β_k–component of $\boldsymbol{x}$ lies to the left of the central vertical axis of the figure. The domain p can only be alternatively decomposed as the triangle t_{k+2} and a pentagon, bounded by β_k, a horizontal arc, a vertical arc, another horizontal arc and β_{k+2}, in its induced orientation. The pentagon is in the opposite orientation as one that would be counted in the map f_{k+2}, and its boundary meets only the intersection point v_{k+2}; such a pentagon is not counted in any map. The triangle t_{k+2} is not counted in any map either. Therefore, p is counted exactly once in $f_{k+1} \circ f_k + \partial_{k+2} \circ \varphi_k + \varphi_k \circ \partial_k$. However, we can replace t_{k+2} with the triangle t_k, and obtain a corresponding domain p' that also connects $\boldsymbol{x}$ to $\boldsymbol{y}$. The new domain p' admits a unique alternative decomposition as a quadrilateral and a rectangle, counted in $\partial_{k+2} \circ \mathcal{Q}_k$. See Figure 7(2).

(3) Two triangles appearing in $\mathcal{T}_{k+1} \circ \mathcal{T}_k$. The domain p can only be alternatively decomposed as the triangle t_{k+1} and another triangle, bounded by segments of β_k, β_{k+2} and a horizontal circle in its induced orientation, and having u_{k+2} as a corner. Since neither triangle is counted in any of the maps f_k or φ_k, p is counted exactly once in $f_{k+1} \circ f_k + \partial_{k+2} \circ \varphi_k + \varphi_k \circ \partial_k$. However, we can replace t_{k+1} with the annulus b, and

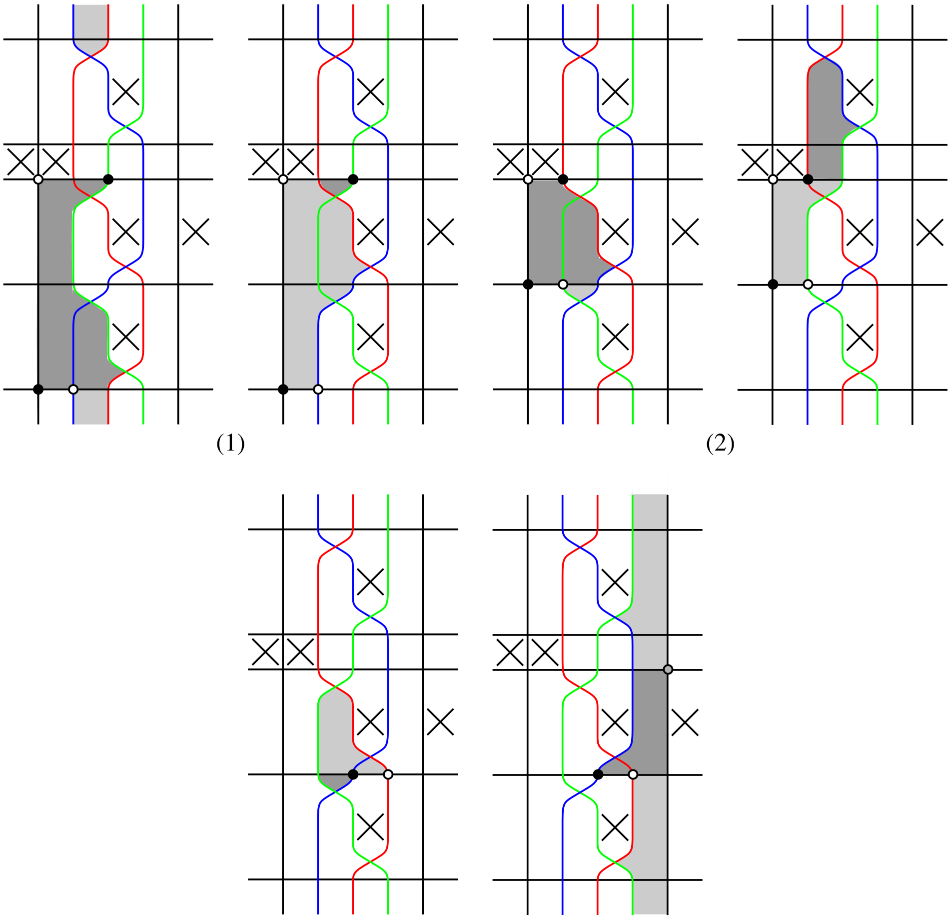

Figure 7: Three special cases. In each case, there are two domains p and p', each counted exactly once in $f_{k+1} \circ f_k + \partial_{k+2} \circ \varphi_k + \varphi_k \circ \partial_k$.

obtain a corresponding domain p' that also connects $\boldsymbol{x}$ to $\boldsymbol{y}$. The situation is similar to the special case in the proof of [12, Lemma 3.1]. Depending on the initial point $\boldsymbol{x}$, the new domain p' admits a unique alternative decomposition as two pentagons or as a hexagon and a rectangle, counted either in $\mathcal{P}_{k+1} \circ \mathcal{P}_k$, in $\partial_{k+2} \circ \mathcal{H}_k$, or in $\mathcal{H}_k \circ \partial_k$. See Figure 7(3).

Finally, the remaining terms in $\partial_{k+2} \circ \mathcal{Q}_k$ cancel with terms in $\mathcal{Q}_k \circ \partial_k$, and the remaining terms in $\partial_{k+2} \circ \mathcal{H}_k$ cancel with terms in $\mathcal{H}_k \circ \partial_k$. Table 1 summarizes how the terms cancel each other. $\square$

We now define the chain homotopy $\psi_k \colon C_k \to C_k$ by counting heptagons.

term in		position	cancels with term in	special case
$f_{k+1} \circ f_k$	$\mathcal{P}_{k+1} \circ \mathcal{P}_k$	left, right ($\not\supset b$)	$\partial_{k+2} \circ \mathcal{H}_k$ or $\mathcal{H}_k \circ \partial_k$	
		right ($\supset b$)	$\mathcal{T}_{k+1} \circ \mathcal{T}_k$	(3)
	$\mathcal{T}_{k+1} \circ \mathcal{P}_k$	left (central x)	$\mathcal{P}_{k+1} \circ \mathcal{T}_k$	(1)
		left (left x)	$\partial_{k+2} \circ \mathcal{Q}_k$	(2)
	$\mathcal{P}_{k+1} \circ \mathcal{T}_k$	left (central y)	$\mathcal{Q}_k \circ \partial_k$	
		left (otherwise)	$\mathcal{T}_{k+1} \circ \mathcal{P}_k$	(1)
		right	$\partial_{k+2} \circ \mathcal{Q}_k$	
	$\mathcal{T}_{k+1} \circ \mathcal{T}_k$	left	$\mathcal{P}_{k+1} \circ \mathcal{P}_k$, $\partial_{k+2} \circ \mathcal{H}_k$ or $\mathcal{H}_k \circ \partial_k$	(3)
$\partial_{k+2} \circ \varphi_k$	$\partial_{k+2} \circ \mathcal{H}_k$	left	$\mathcal{P}_{k+1} \circ \mathcal{P}_k$	
		right ($\not\supset b$)	$\mathcal{P}_{k+1} \circ \mathcal{P}_k$ or $\mathcal{H}_k \circ \partial_k$	
		right ($\supset b$)	$\mathcal{T}_{k+1} \circ \mathcal{T}_k$	(3)
	$\partial_{k+2} \circ \mathcal{Q}_k$	left (central y)	$\mathcal{Q}_k \circ \partial_k$	
		left (otherwise)	$\mathcal{T}_{k+1} \circ \mathcal{P}_k$	(2)
		right	$\mathcal{P}_{k+1} \circ \mathcal{T}_k$ or $\mathcal{Q}_k \circ \partial_k$	
$\varphi_k \circ \partial_k$	$\mathcal{H}_k \circ \partial_k$	left	$\mathcal{P}_{k+1} \circ \mathcal{P}_k$	
		right ($\not\supset b$)	$\mathcal{P}_{k+1} \circ \mathcal{P}_k$ or $\partial_{k+2} \circ \mathcal{H}_k$	
		right ($\supset b$)	$\mathcal{T}_{k+1} \circ \mathcal{T}_k$	(3)
	$\mathcal{Q}_k \circ \partial_k$	left	$\mathcal{P}_{k+1} \circ \mathcal{T}_k$ or $\partial_{k+2} \circ \mathcal{Q}_k$	
		right	$\partial_{k+2} \circ \mathcal{Q}_k$	

Table 1: This table shows how the terms cancel each other in Lemma 3.6. A left domain is indicated as "(central x)" if the β_k–component of $\boldsymbol{x}$ lies on the central vertical axis of the figure, and "(left x)" if it lies to the left of this axis; similarly for $\boldsymbol{y}$. A right domain is indicated as "($\supset b$)" if $\mathrm{Int}(p) \supset \mathrm{Int}(b)$, and "($\not\supset b$)" otherwise. The special cases are shown in Figure 7.

Given $\boldsymbol{x} \in \boldsymbol{S}(\mathbb{G}_k)$ and $\boldsymbol{y} \in \boldsymbol{S}(\mathbb{G}_{k+2})$, let $\mathrm{Hept}_k(\boldsymbol{x}, \boldsymbol{y})$ denote the space of embedded heptagons. First of all, $\mathrm{Hept}_k(\boldsymbol{x}, \boldsymbol{y})$ is empty unless $\boldsymbol{x}, \boldsymbol{y}$ coincide at exactly $n-2$ points. An element p of $\mathrm{Hept}_k(\boldsymbol{x}, \boldsymbol{y})$ is an embedded disk in $\mathbb{T}$, whose boundary consists of seven arcs, each contained in horizontal or vertical circles; under the orientation induced on the boundary of p, we start at the β_k–component of $\boldsymbol{x}$, traverse the arc of a horizontal circle, meet its corresponding component of $\boldsymbol{y}$, proceed to an arc of a vertical circle, meet the corresponding component of $\boldsymbol{x}$, continue through another horizontal circle, meet the component of $\boldsymbol{y}$ contained in β_k, proceed to an arc in β_k, meet the intersection point u_{k+2} of β_k and β_{k+2}, proceed to an arc in β_{k+2},

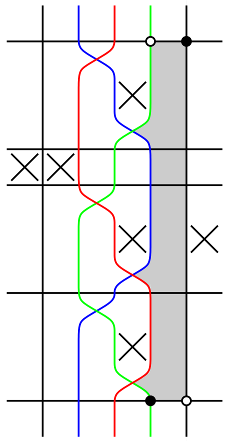

Figure 8: An allowed heptagon in $\mathrm{Hept}_k^\circ(\boldsymbol{x},\boldsymbol{y})$. Note that allowed heptagons are necessarily right domains.

meet the intersection point u_{k+1} of β_{k+1} and β_{k+2}, traverse an arc in β_{k+1}, meet the intersection point u_k of β_k and β_{k+1}, and finally, traverse an arc in β_k until we arrive back at the initial component of $\boldsymbol{x}$. All the angles here are again at most straight angles. The space of empty heptagons $p \in \mathrm{Hept}_k(\boldsymbol{x},\boldsymbol{y})$ with $\boldsymbol{x} \cap \mathrm{Int}(p) = \varnothing$ is denoted $\mathrm{Hept}_k^\circ(\boldsymbol{x},\boldsymbol{y})$.

See Figure 8 for an example of an element of $\mathrm{Hept}_k^\circ(x,y)$.

Given $\boldsymbol{x} \in \boldsymbol{S}(\mathbb{G}_k)$, we define

$$\psi_k(\boldsymbol{x}) = \mathcal{K}_k(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in\boldsymbol{S}(\mathbb{G}_k)} \sum_{\substack{p\in\mathrm{Hept}_k^\circ(\boldsymbol{x},\boldsymbol{y})\\ \mathrm{Int}(p)\cap\mathbb{X}=\varnothing}} \boldsymbol{y} \in C_k.$$

Lemma 3.7 *We have*

$$\varphi_{k+1}\circ f_k + f_{k+2}\circ\varphi_k + \partial_k\circ\psi_k + \psi_k\circ\partial_k = \mathrm{Id},$$

so that the maps f_k and φ_k satisfy condition (2) *of Lemma 3.3.*

Proof First, we see that juxtaposing either a triangle and a hexagon, or a pentagon and a quadrilateral, does not contribute to $\varphi_{k+1}\circ f_k + f_{k+2}\circ\varphi_k$. In other words, we first claim that $\mathcal{P}_{k+2}\circ\mathcal{Q}_k + \mathcal{Q}_{k+1}\circ\mathcal{P}_k + \mathcal{H}_{k+1}\circ\mathcal{T}_k + \mathcal{T}_{k+2}\circ\mathcal{H}_k = 0$.

There are exactly four cases. We consider domains p formed by juxtaposing:

(1) A quadrilateral and a pentagon appearing in $\mathcal{P}_{k+2}\circ\mathcal{Q}_k$, such that p is a right domain. In this case, p admits a unique alternative decomposition as a triangle and a hexagon appearing in $\mathcal{H}_{k+1}\circ\mathcal{T}_k$. See Figure 9(1).

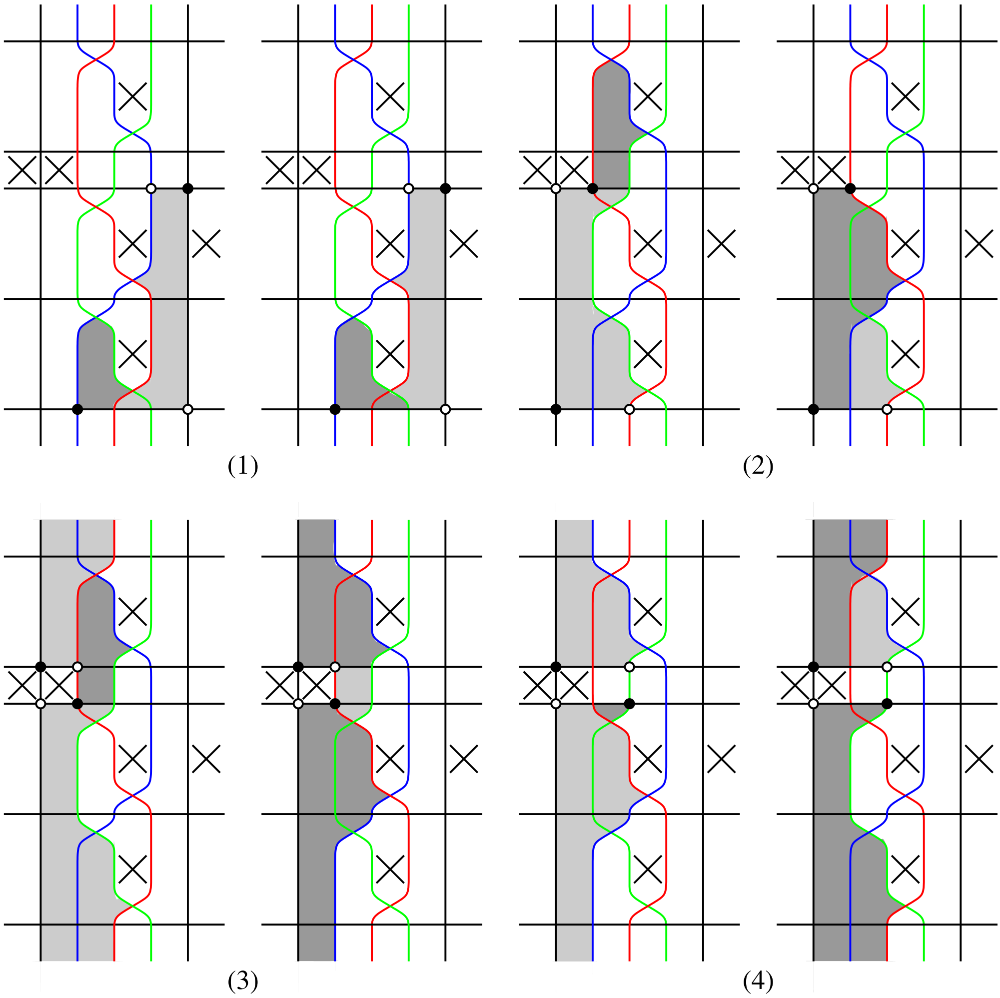

Figure 9: The four cases when juxtaposing a triangle and a hexagon, or a pentagon and a quadrilateral. All terms that arise cancel out with each other.

(2) A quadrilateral and a pentagon appearing in $\mathcal{P}_{k+2} \circ \mathcal{Q}_k$, such that p is a left domain with height less than n. Only this decomposition of p is counted. However, we can replace the triangle t_k inside p by the triangle t_{k+2}, and obtain a corresponding domain p' that can be uniquely decomposed as a pentagon and a quadrilateral appearing in $\mathcal{Q}_{k+1} \circ \mathcal{P}_k$. See Figure 9(2).

(3) A quadrilateral and a pentagon appearing in $\mathcal{P}_{k+2} \circ \mathcal{Q}_k$, such that p is a right domain with height n. This is in fact only possible when $k = 1$. Only this decomposition of p is counted. However, p contains the triangles t_k and t_{k+1}; we can replace t_{k+1}

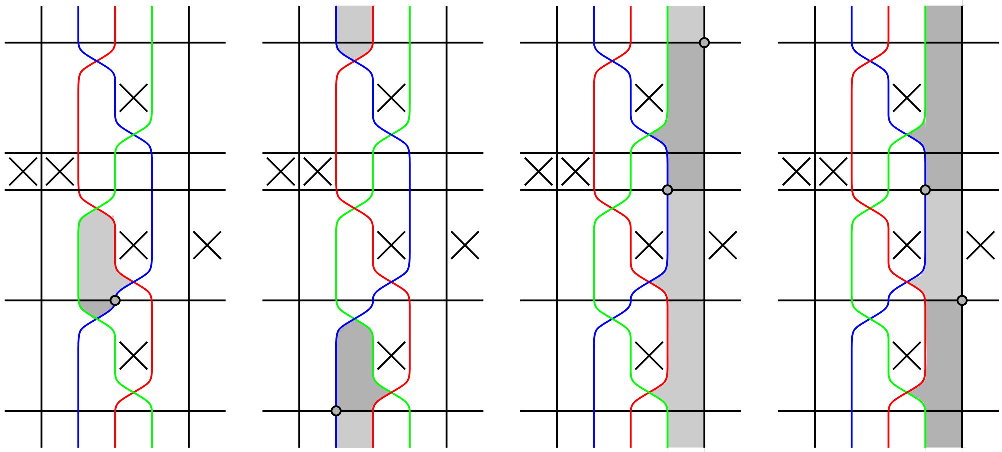

Figure 10: Decomposing the identity map

by the triangle t_{k+2}, and obtain a corresponding domain p' that can be uniquely decomposed as a hexagon and a triangle appearing in $\mathcal{T}_{k+2} \circ \mathcal{H}_k$. See Figure 9(3).

The astute reader may find it strange that in this particular case, in p the triangle t_k is "attached" to the top of the rest of the domain, whereas in p' it is "attached" to the bottom. One way to convince oneself of the validity of the procedure of obtaining p' from p, is to think of it as first replacing t_k by t_{k+2}, and then replacing t_{k+1} by t_k.

(4) A triangle and a hexagon appearing in $\mathcal{H}_{k+1} \circ \mathcal{T}_k$, such that p is a left domain. This is only possible when $k = 0$. Only this decomposition of p is counted. However, we can replace the triangle t_k inside p by the triangle t_{k+2}, and obtain a corresponding domain p' that can be uniquely decomposed as a pentagon and a quadrilateral appearing in $\mathcal{Q}_{k+1} \circ \mathcal{P}_k$. See Figure 9(4).

Juxtaposing a pentagon and a hexagon appearing in $\mathcal{H}_{k+1} \circ \mathcal{P}_k$ or $\mathcal{P}_{k+2} \circ \mathcal{H}_k$, we generically obtain a composite domain that admits a unique alternative decomposition as a heptagon and a rectangle, appearing in $\partial_k \circ \mathcal{K}_k$ or $\mathcal{K}_k \circ \partial_k$, except for one special case discussed below.

Depending on the initial point $\boldsymbol{x}$, there exists exactly one domain p connecting $\boldsymbol{x}$ to itself that admits a unique decomposition, either as a triangle and a quadrilateral in $\mathcal{Q}_{k+1} \circ \mathcal{T}_k$ (in which case p is the triangle t_{k+1}), as a quadrilateral and a triangle in $\mathcal{T}_{k+2} \circ \mathcal{Q}_k$ (p is the triangle t_k), as a pentagon and a hexagon appearing in $\mathcal{H}_{k+1} \circ \mathcal{P}_k$ or $\mathcal{P}_{k+2} \circ \mathcal{H}_k$ (p is the annulus b), or as a rectangle and a heptagon appearing in $\mathcal{K}_k \circ \partial_k$ or $\partial_k \circ \mathcal{K}_k$ (p is again the annulus b). Of course, in this case, the identity map is counted once. See Figure 10.

term in		position	cancels with term in	special case
$\varphi_{k+1} \circ f_k$	$\mathcal{H}_{k+1} \circ \mathcal{P}_k$	right ($\neq b$)	$\partial_k \circ \mathcal{K}_k$ or $\mathcal{K}_k \circ \partial_k$	
		right ($= b$)	Id	
	$\mathcal{Q}_{k+1} \circ \mathcal{P}_k$	left (ht. $< n-1$)	$\mathcal{P}_{k+2} \circ \mathcal{Q}_k$	(2)
		left (ht. $= n-1$)	$\mathcal{H}_{k+1} \circ \mathcal{T}_k$	(4)
	$\mathcal{H}_{k+1} \circ \mathcal{T}_k$	left	$\mathcal{Q}_{k+1} \circ \mathcal{P}_k$	(4)
		right	$\mathcal{P}_{k+2} \circ \mathcal{Q}_k$	(1)
	$\mathcal{Q}_{k+1} \circ \mathcal{T}_k$	left	Id	
$f_{k+2} \circ \varphi_k$	$\mathcal{P}_{k+2} \circ \mathcal{H}_k$	right ($\neq b$)	$\partial_k \circ \mathcal{K}_k$ or $\mathcal{K}_k \circ \partial_k$	
		right ($= b$)	Id	
	$\mathcal{T}_{k+2} \circ \mathcal{H}_k$	left	$\mathcal{P}_{k+2} \circ \mathcal{Q}_k$	(3)
	$\mathcal{P}_{k+2} \circ \mathcal{Q}_k$	left (ht. $< n$)	$\mathcal{Q}_{k+1} \circ \mathcal{P}_k$	(2)
		left (ht. $= n$)	$\mathcal{T}_{k+1} \circ \mathcal{H}_k$	(3)
		right	$\mathcal{H}_{k+1} \circ \mathcal{T}_k$	(1)
	$\mathcal{T}_{k+2} \circ \mathcal{Q}_k$	left	Id	
$\partial_k \circ \psi_k$	$\partial_k \circ \mathcal{K}_k$	right	$\mathcal{H}_{k+1} \circ \mathcal{P}_k$, $\mathcal{P}_{k+2} \circ \mathcal{H}_k$, $\mathcal{K}_k \circ \partial_k$ or Id	
$\psi_k \circ \partial_k$	$\mathcal{K}_k \circ \partial_k$	right	$\mathcal{H}_{k+1} \circ \mathcal{P}_k$, $\mathcal{P}_{k+2} \circ \mathcal{H}_k$, $\partial_k \circ \mathcal{K}_k$ or Id	
Id		n/a	$\mathcal{H}_{k+1} \circ \mathcal{P}_k$, $\mathcal{P}_{k+2} \circ \mathcal{H}_k$, $\partial_k \circ \mathcal{K}_k$, $\mathcal{K}_k \circ \partial_k$, $\mathcal{Q}_{k+1} \circ \mathcal{T}_k$ or $\mathcal{T}_{k+2} \circ \mathcal{Q}_k$	

Table 2: This table shows how the terms cancel each other in Lemma 3.7. The special cases are shown in Figure 9.

Finally, the remaining terms in $\partial_k \circ \mathcal{K}_k$ cancel with terms in $\mathcal{K}_k \circ \partial_k$. Table 2 summarizes how the terms above cancel each other. □

The proof of Proposition 3.2 is completed by combining Lemmas 3.3, 3.4, 3.6 and 3.7.

4 Signs

Sign refinements are given by Manolescu, Ozsváth, Szabó and Thurston [12] to extend the definition of combinatorial knot Floer homology to one with coefficients in $\mathbb{Z}$.

In this section, we shall likewise assign sign refinements to our maps, to prove the analogous statement of Proposition 3.2 with coefficients in $\mathbb{Z}$.

Given a grid diagram, we denote by Rect° the union of $\mathrm{Rect}^\circ(\boldsymbol{x}, \boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y}$. We follow [12] and adopt the following definition.

Definition 4.1 A *true sign assignment*, or simply a *sign assignment*, is a function

$$\mathcal{S}\colon \mathrm{Rect}^\circ \to \{\pm 1\}$$

with the following properties:

(1) For any four distinct $r_1, r_2, r_1', r_2' \in \mathrm{Rect}^\circ$ with $r_1 * r_2 = r_1' * r_2'$, we have

$$\mathcal{S}(r_1) \cdot \mathcal{S}(r_2) = -\mathcal{S}(r_1') \cdot \mathcal{S}(r_2').$$

(2) For $r_1, r_2 \in \mathrm{Rect}^\circ$ such that $r_1 * r_2$ is a vertical annulus, we have

$$\mathcal{S}(r_1) \cdot \mathcal{S}(r_2) = -1.$$

(3) For $r_1, r_2 \in \mathrm{Rect}^\circ$ such that $r_1 * r_2$ is a horizontal annulus, we have

$$\mathcal{S}(r_1) \cdot \mathcal{S}(r_2) = +1.$$

Theorem 4.2 (Manolescu, Ozsváth, Szabó and Thurston) *There exists a sign assignment as defined in Definition 4.1. Moreover, this sign assignment is essentially unique: if $\mathcal{S}_1$ and $\mathcal{S}_2$ are two sign assignments, then there is a function $g\colon \boldsymbol{S}(\mathbb{G}) \to \{\pm 1\}$ such that $\mathcal{S}_2(r) = g(\boldsymbol{x}) \cdot g(\boldsymbol{y}) \cdot \mathcal{S}_1(r)$ for all $r \in \mathrm{Rect}^\circ(\boldsymbol{x}, \boldsymbol{y})$.*

Remark 4.3 In the construction of a sign assignment in [12], the sign of a rectangle does not depend on the positions of the Os and Xs of the diagram. We can view each generator $\boldsymbol{x}$ as a permutation $\sigma_{\boldsymbol{x}}$: if the component of $\boldsymbol{x}$ on the i^{th} horizontal circle lies on the $s(i)^{\text{th}}$ vertical circle, then we let $\sigma_{\boldsymbol{x}}$ be $(s(1)\, s(2) \cdots s(n))$. Then the sign of a rectangle in $\mathrm{Rect}^\circ(\boldsymbol{x}, \boldsymbol{y})$ depends only on $\sigma_{\boldsymbol{x}}$ and $\sigma_{\boldsymbol{y}}$.

The sign assignment from Theorem 4.2 is then used in [12] to construct a chain complex over $\mathbb{Z}$ as follows. The complex $\widetilde{\mathrm{GC}}(\mathbb{G}) = \widetilde{\mathrm{GC}}(\mathbb{G}; \mathbb{Z})$ is the free $\mathbb{Z}$–module generated by elements of $\boldsymbol{S}(\mathbb{G})$. Fixing a sign assignment $\mathcal{S}$, the complex $\widetilde{\mathrm{GC}}(\mathbb{G})$ is endowed with the endomorphism $\partial_{\mathcal{S}}\colon \widetilde{\mathrm{GC}}(\mathbb{G}) \to \widetilde{\mathrm{GC}}(\mathbb{G})$, defined by

$$\partial_{\mathcal{S}}(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \boldsymbol{S}(\mathbb{G})} \sum_{\substack{r \in \mathrm{Rect}^\circ(\boldsymbol{x}, \boldsymbol{y}) \\ \mathrm{Int}(r) \cap \mathbb{X} = \varnothing}} \mathcal{S}(r) \cdot \boldsymbol{y} \in \widetilde{\mathrm{GC}}(\mathbb{G}).$$

One can then see that $(\widetilde{\mathrm{GC}}(\mathbb{G}), \partial_{\mathcal{S}})$ is a chain complex. Indeed, the terms in $\partial_{\mathcal{S}} \circ \partial_{\mathcal{S}}(\boldsymbol{x})$ can be paired off as before, by the axioms defining $\mathcal{S}$. Moreover, given $\mathcal{S}_1$ and $\mathcal{S}_2$, the map $\Phi\colon (\widetilde{\mathrm{GC}}(\mathbb{G}), \partial_{\mathcal{S}_1}) \to (\widetilde{\mathrm{GC}}(\mathbb{G}), \partial_{\mathcal{S}_2})$, defined by

$$\Phi(\boldsymbol{x}) = g(\boldsymbol{x}) \cdot \boldsymbol{x},$$

gives an isomorphism of the two chain complexes. Again, one can take the homology of the chain complex $(\widetilde{\mathrm{GC}}(\mathbb{G}), \partial_{\mathcal{S}})$, and define

$$\widetilde{\mathrm{GH}}(\mathbb{G}) = \mathrm{H}_*(\widetilde{\mathrm{GC}}(\mathbb{G}), \partial_{\mathcal{S}}).$$

It is shown in [12] that

$$\widetilde{\mathrm{GH}}(\mathbb{G}) \cong \widehat{\mathrm{GH}}(\mathbb{G}) \otimes V^{n-\ell}$$

for some $\mathbb{Z}$–module $\widehat{\mathrm{GH}}(\mathbb{G})$ that is a link invariant, where V is a rank 2 free module over $\mathbb{Z}$, spanned by one generator in bigrading $(-1,-1)$ and another in bigrading $(0,0)$. The link invariant $\widehat{\mathrm{GH}}(\mathbb{G})$, also denoted by $\widehat{\mathrm{GH}}(L)$, is shown by Sarkar [21] to be isomorphic to $\widehat{\mathrm{HFK}}(L, \mathfrak{o})$ for some *orientation system* $\mathfrak{o}$ of the link L.

We are now ready to turn to the proof of the analogous statement of Proposition 3.2, with signs.

Proposition 4.4 *There exists an exact triangle*

$$\cdots \to \widetilde{\mathrm{GH}}(\mathbb{G}_\infty; \mathbb{Z}) \to \widetilde{\mathrm{GH}}(\mathbb{G}_0; \mathbb{Z}) \to \widetilde{\mathrm{GH}}(\mathbb{G}_1; \mathbb{Z}) \to \cdots .$$

Our proof is reminiscent of that in [12, Section 4]. We adopt the strategy from Section 3; to do so, we must specify the signs used in defining our various chain maps and chain homotopies, and check that they indeed satisfy Lemma 3.3.

We begin by considering pentagons. First, we define the notion of a *corresponding generator*. For each $\boldsymbol{x} \in C_k$, there exist exactly one $\boldsymbol{x}' \in C_{k+1}$ and one $\boldsymbol{x}'' \in C_{k+2}$ that are canonically closest to $\boldsymbol{x}$; we require that $\boldsymbol{x}$, $\boldsymbol{x}'$ and $\boldsymbol{x}''$ coincide everywhere except on the β curves, and $\boldsymbol{x}'$ and $\boldsymbol{x}''$ are obtained from $\boldsymbol{x}$ by sliding the β_k–component horizontally to the β_{k+1} and β_{k+2} curves respectively. We define the maps $c_k^+\colon C_k \to C_{k+1}$ and $c_k^-\colon C_k \to C_{k+2}$ by

$$c_k^+(\boldsymbol{x}) = \boldsymbol{x}' \quad \text{and} \quad c_k^-(\boldsymbol{x}) = \boldsymbol{x}''.$$

We can now define the straightening maps

$$d_k^{\mathcal{P}}\colon \mathrm{Pent}_k^\circ(\boldsymbol{x}, \boldsymbol{y}) \to \mathrm{Rect}_k^\circ(\boldsymbol{x}, c_k^-(\boldsymbol{y})), \quad e_k^{\mathcal{P}}\colon \mathrm{Pent}_k^\circ(\boldsymbol{x}, \boldsymbol{y}) \to \mathrm{Rect}_{k+1}^\circ(c_k^+(\boldsymbol{x}), \boldsymbol{y}),$$

as follows. Given $p \in \mathrm{Pent}_k^\circ(\boldsymbol{x}, \boldsymbol{y})$, we obtain $d_k^{\mathcal{P}}(p)$ by sliding the β_{k+1}–component of $\boldsymbol{y}$ back to the β_k curve, thereby postcomposing p with a triangle; similarly, we obtain $e_k^{\mathcal{P}}(p)$ by sliding the β_k–component of $\boldsymbol{x}$ to the β_{k+1} curve, thereby precomposing p with another triangle. Notice that by Remark 4.3, we have

$$\mathcal{S}(d_k^{\mathcal{P}}(p)) = \mathcal{S}(e_k^{\mathcal{P}}(p)).$$

We now define

$$\mathcal{P}_k(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in\boldsymbol{S}(\mathbb{G}_{k+1})} \sum_{\substack{p\in\mathrm{Pent}_k^\circ(\boldsymbol{x},\boldsymbol{y})\\ \mathrm{Int}(p)\cap\mathbb{X}=\varnothing}} \epsilon_k^{\mathcal{P}}(p)\cdot\boldsymbol{y} \in C_{k+1},$$

where

$$\epsilon_k^{\mathcal{P}}(p) = \begin{cases} \mathcal{S}(d_k^{\mathcal{P}}(p)) & \text{if } p \text{ is a left pentagon,}\\ -\mathcal{S}(d_k^{\mathcal{P}}(p)) & \text{if } p \text{ is a right pentagon.}\end{cases}$$

Turning to triangles, we again view generators as permutations. The signature $\mathrm{sgn}(\boldsymbol{x})$ of a generator $\boldsymbol{x}$ is defined to be the signature $\mathrm{sgn}(\sigma_{\boldsymbol{x}})$ of the corresponding permutation. Then we define

$$\mathcal{T}_k(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in\boldsymbol{S}(\mathbb{G}_{k+1})} \sum_{\substack{p\in\mathrm{Tri}_k(\boldsymbol{x},\boldsymbol{y})\\ \mathrm{Int}(p)\cap\mathbb{X}=\varnothing}} \epsilon_k^{\mathcal{T}}(p)\cdot\boldsymbol{y} \in C_{k+1},$$

where

$$\epsilon_k^{\mathcal{T}}(p) = \mathrm{sgn}(\boldsymbol{x}).$$

Finally, we define

$$f_k(\boldsymbol{x}) = \mathcal{P}_k(\boldsymbol{x}) + \mathcal{T}_k(\boldsymbol{x}) \in C_{k+1}.$$

Lemma 4.5 *The map f_k is an anti-chain map. In fact, $\mathcal{P}_k$ and $\mathcal{T}_k$ are both anti-chain maps.*

Proof The proof follows from the proof of Lemma 3.4. We first consider the juxtaposition of a pentagon and a rectangle. If the two polygons are disjoint or have overlapping interiors, then the domain can be decomposed as either $r * p$ or $p' * r'$; then by property (1) of Definition 4.1, $\mathcal{S}(r)\cdot\mathcal{S}(d_k^{\mathcal{P}}(p)) = -\mathcal{S}(d_k^{\mathcal{P}}(p'))\cdot\mathcal{S}(r')$, and consequently $\mathcal{S}(r)\cdot\epsilon_k^{\mathcal{P}}(p) = -\epsilon_k^{\mathcal{P}}(p')\cdot\mathcal{S}(r')$. If the two polygons share a corner, then the domain can be decomposed in two ways; straightening the pentagons (with either $d_k^{\mathcal{P}}$ or $e_k^{\mathcal{P}}$) and using property (1) of Definition 4.1, we again see that the terms cancel.

Consider now the juxtaposition of a triangle p and a rectangle r. Notice first that the differential ∂ always changes the signature of a generator; this means that $\mathcal{S}(r)\cdot\epsilon_k^{\mathcal{T}}(p) = -\epsilon_k^{\mathcal{T}}(p')\cdot\mathcal{S}(r')$. By Remark 3.5, such a domain can always be decomposed in two ways, one contributing to $\partial_{k+1}\circ\mathcal{T}_k$, and one to $\mathcal{T}_k\circ\partial_k$; moreover, the two rectangles involved correspond to the same permutation, and so in fact have the same sign. □

We similarly define the straightening maps

$$d_k^{\mathcal{H}}\colon \operatorname{Hex}_k^\circ(\boldsymbol{x},\boldsymbol{y}) \to \operatorname{Rect}_k^\circ(\boldsymbol{x}, c_k^+(\boldsymbol{y})), \quad e_k^{\mathcal{H}}\colon \operatorname{Hex}_k^\circ(\boldsymbol{x},\boldsymbol{y}) \to \operatorname{Rect}_{k+2}^\circ(c_k^-(\boldsymbol{x}),\boldsymbol{y})$$

by sliding the appropriate component, and again notice that $\mathcal{S}(d_k^{\mathcal{H}}(p)) = \mathcal{S}(e_k^{\mathcal{H}}(p))$. We define

$$\mathcal{H}_k(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in \boldsymbol{S}(\mathbb{G}_{k+2})} \sum_{\substack{p\in \operatorname{Hex}_k^\circ(\boldsymbol{x},\boldsymbol{y}) \\ \operatorname{Int}(p)\cap\mathbb{X}=\varnothing}} \epsilon_k^{\mathcal{H}}(p)\cdot \boldsymbol{y} \ \in C_{k+2},$$

where

$$\epsilon_k^{\mathcal{H}}(p) = \mathcal{S}(d_k^{\mathcal{H}}(p)).$$

For quadrilaterals,

$$\mathcal{Q}_k(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in \boldsymbol{S}(\mathbb{G}_{k+2})} \sum_{\substack{p\in \operatorname{Quad}_k(\boldsymbol{x},\boldsymbol{y}) \\ \operatorname{Int}(p)\cap\mathbb{X}=\varnothing}} \epsilon_k^{\mathcal{Q}}(p)\cdot \boldsymbol{y} \ \in C_{k+2},$$

where

$$\epsilon_k^{\mathcal{Q}}(p) = \operatorname{sgn}(\boldsymbol{x}).$$

Lemma 4.6 *The maps f_k and φ_k satisfy condition* (1) *of Lemma 3.3.*

Proof Again the proof follows from that of Lemma 3.6. We say that a domain p is *rectangle-like* if it is an allowed rectangle, pentagon, hexagon or heptagon, and we write $p\in \mathrm{RL}$; we say that it is *triangle-like* if it is an allowed triangle or quadrilateral, and we write $p\in\mathrm{TL}$. We say a domain p is *Type I* if $p\in \mathrm{RL}*\mathrm{RL}$, *Type II* if $p\in\mathrm{RL}*\mathrm{TL}$, *Type III* if $p\in\mathrm{TL}*\mathrm{RL}$, and *Type IV* if $p\in \mathrm{TL}*\mathrm{TL}$. (Recall that $*$ and $\circ$ compose in the opposite order, so a term in $\mathcal{P}_{k+1}\circ\mathcal{T}_k$ is in $\mathrm{TL}*\mathrm{RL}$.) Refer to Table 1. Typically, a domain can be decomposed in two ways; usually, it fits into one of these cases:

(1) **Both decompositions are Type I** In this case, we see that the terms cancel out by straightening the polygons and applying property (1) of Definition 4.1.

(2) **One decomposition is Type II and one is Type III** The domain is a left domain. In this case, the terms cancel out because the two rectangle-like polygons have the same sign, but the two triangle-like polygons have opposite signs.

(3) **Both decompositions are Type III** The domain is a right domain. In this case, the two triangle-like polygons have the same sign. However, the two rectangle-like polygons are both right domains, and one is a rectangle while the other is a pentagon. Since right pentagons have opposite signs as the straightened rectangles, the two rectangle-like polygons are of opposite sign.

The only special cases are those that involve two different domains, which are exactly the special cases in Lemma 3.6.

(1) In this case, the two pentagons have the same sign, but the two triangles have opposite signs.

(2) In this case, the rectangle and the pentagon have the same sign, but the triangle and the quadrilateral have opposite signs.

(3) In one domain, the two triangles have the same sign, and so their composite is positive; in the other domain, the composite of the straightened rectangles is a vertical annulus, and so is negative by property (2) of Definition 4.1.

Since there are no other cases, our proof is complete. □

For heptagons, there is only one straightening map $d_k^{\mathcal{K}}\colon \mathrm{Hept}_k^\circ(\boldsymbol{x},\boldsymbol{y}) \to \mathrm{Rect}_k^\circ(\boldsymbol{x},\boldsymbol{y})$. Since the only allowed heptagons are right heptagons, putting $\epsilon_k^{\mathcal{K}}(p) = -\mathcal{S}(d_k^{\mathcal{K}}(p))$ we define

$$\mathcal{K}_k(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in \boldsymbol{S}(\mathbb{G}_k)} \sum_{\substack{p\in \mathrm{Hept}_k^\circ(\boldsymbol{x},\boldsymbol{y}) \\ \mathrm{Int}(p)\cap\mathbb{X}=\varnothing}} \epsilon_k^{\mathcal{K}}(p)\cdot \boldsymbol{y} \in C_k.$$

Lemma 4.7 *We have*

$$\varphi_{k+1}\circ f_k + f_{k+2}\circ\varphi_k + \partial_k\circ\psi_k + \psi_k\circ\partial_k = \mathrm{Id},$$

so that the maps f_k and φ_k satisfy condition (2) *of Lemma 3.3.*

Proof The typical cases are as in the proof of Lemma 4.6. We check the special cases in Lemma 3.7.

(1) This is actually a typical case; the two decompositions of the domain are both Type III.

(2) In this case, the pentagons have the same sign, but the quadrilaterals have opposite signs.

(3) In this case, the pentagon and the hexagon have the same sign, but the triangle and the quadrilateral have opposite signs.

(4) Also in this case, the pentagon and the hexagon have the same sign, but the triangle and the quadrilateral have opposite signs.

We now check the decomposition of the identity map. If it is decomposed as a triangle and a quadrilateral, we see that they are of the same sign, and so we obtain a positive domain. If it is decomposed as a pentagon and a hexagon, we see that the composite of the straightened rectangles is negative by property (2) of Definition 4.1, but the right pentagon and its straightening have opposite signs; this means that the overall domain is also positive. □

The proof of Proposition 4.4 is completed by combining Lemmas 3.3, 4.5, 4.6 and 4.7.

5 Iteration of the skein exact triangle

In this section, we will work only over $\mathbb{F}_2$. Let $\mathbb{G}_\infty$, $\mathbb{G}_0$, $\mathbb{G}_1$ be grid diagrams that are identical except near a point, as indicated in Figure 2. In Section 3, we constructed the maps $f_k\colon \widetilde{\mathrm{GC}}(\mathbb{G}_k) \to \widetilde{\mathrm{GC}}(\mathbb{G}_{k+1})$ and $\varphi_k\colon \widetilde{\mathrm{GC}}(\mathbb{G}_k) \to \widetilde{\mathrm{GC}}(\mathbb{G}_{k+2})$ that satisfy Lemma 3.3. Now Lemma 3.3 and the five lemma together imply that $\widetilde{\mathrm{GC}}(\mathbb{G}_\infty)$ is quasi-isomorphic to the mapping cone of $f_0\colon \widetilde{\mathrm{GC}}(\mathbb{G}_0) \to \widetilde{\mathrm{GC}}(\mathbb{G}_1)$ (see [17, Lemma 4.2]), where the quasi-isomorphism is given by

$$f_\infty + \varphi_\infty\colon \widetilde{\mathrm{GC}}(\mathbb{G}_\infty) \to \widetilde{\mathrm{GC}}(\mathbb{G}_0) \oplus \widetilde{\mathrm{GC}}(\mathbb{G}_1).$$

We now wish to iterate this quasi-isomorphism to obtain a cube of resolutions that computes the same homology.

Let the crossings of a link L be numbered from 1 to m. Start with a planar projection of L, and convert it into a grid diagram. By applying stabilization and commutation as described in [4; 5; 12], we can require that:

(1) Near every crossing, the diagram is as indicated in Figure 11. For the i^{th} crossing, the associated 6×6 block of cells illustrated is referred to as the i^{th} block.

(2) If $i \neq j$, then the i^{th} block and the j^{th} block occupy disjoint rows and disjoint columns.

Let the resulting grid diagram be $\mathbb{G}$. To each sequence $k_1, k_2, \dots, k_m$ with $k_i \in \{\infty, 0, 1\}$ for $1 \le i \le m$, we associate a grid diagram $\mathbb{G}_{k_1,\dots,k_m}$. The diagram $\mathbb{G}_{k_1,\dots,k_m}$ is obtained from $\mathbb{G}$ by replacing the i^{th} block by the appropriate 6×6 block as in

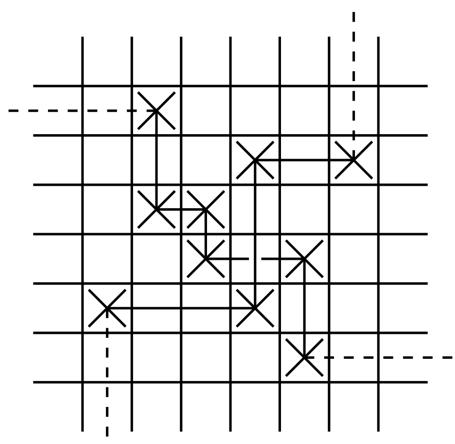

Figure 11: The grid diagram $\mathbb{G}$ of an oriented link L near a crossing. The 6×6 block of cells in the center is the block associated to this crossing.

Figure 2 (where the central 4×4 block is shown), depending on the value of k_i. In particular, $\mathbb{G}_{\infty,\dots,\infty} = \mathbb{G}$. Let $C_{k_1,\dots,k_m}$ be the associated chain complex of $\mathbb{G}_{k_1,\dots,k_m}$, equipped with the differential map. For $1 \le i \le m$, we let the edge map

$$f^i_{k_1,\dots,k_m}\colon C_{k_1,\dots,k_{i-1},k_i,k_{i+1},\dots,k_m} \to C_{k_1,\dots,k_{i-1},k_i+1,k_{i+1},\dots,k_m}$$

be the map f_k defined in Section 3; this makes sense, since the two chain complexes differ only near a crossing. Analogously, we have the map

$$\varphi^i_{k_1,\dots,k_m}\colon C_{k_1,\dots,k_{i-1},k_i,k_{i+1},\dots,k_m} \to C_{k_1,\dots,k_{i-1},k_i+2,k_{i+1},\dots,k_m},$$

which is just the map φ_k defined earlier.

This allows us to define the *big cube of resolutions* of $\mathbb{G}$ to be the complex

$$\begin{aligned}(\mathrm{BCR}(\mathbb{G}), \partial_{\mathrm{BCR}}) = \Biggl(& \bigoplus_{k_i\in\{\infty,0,1\}} C_{k_1,\dots,k_m}, \\ & \sum_{k_i\in\{\infty,0,1\}} \Bigl(\partial_{k_1,\dots,k_m} + \sum_{j\,:\,k_j\in\{\infty,0\}} f^j_{k_1,\dots,k_m} + \sum_{t\,:\,k_t=\infty} \varphi^t_{k_1,\dots,k_m} \Bigr)\Biggr),\end{aligned}$$

and the *(small) cube of resolutions* of $\mathbb{G}$ to be the complex

$$(\mathrm{CR}(\mathbb{G}), \partial_{\mathrm{CR}}) = \Biggl(\bigoplus_{k_i\in\{0,1\}} C_{k_1,\dots,k_m}, \sum_{k_i\in\{0,1\}} \Bigl(\partial_{k_1,\dots,k_m} + \sum_{j\,:\,k_j=0} f^j_{k_1,\dots,k_m} \Bigr)\Biggr),$$

which is a subcomplex of $\mathrm{BCR}(\mathbb{G})$. In the cube of resolutions $\mathrm{CR}(\mathbb{G})$, each vertex is associated to a grid diagram in which all crossings have been resolved. The case when there are two crossings is illustrated below:

$$\begin{tikzcd}
C_{\infty,\infty} \arrow[r, "f^2_{\infty,\infty}"'] \arrow[rr, bend left, "\varphi^2_{\infty,\infty}"] & C_{\infty,0} \arrow[r, "f^2_{\infty,0}"'] \arrow[d, "f^1_{\infty,0}"] \arrow[dd, bend right, "\varphi^1_{\infty,0}"'] & C_{\infty,1} \arrow[d, "f^1_{\infty,1}"'] \arrow[dd, bend left, "\varphi^1_{\infty,1}"] \\
& C_{0,0} \arrow[r, "f^2_{0,0}"'] \arrow[d, "f^1_{0,0}"] & C_{0,1} \arrow[d, "f^1_{0,1}"'] \\
& C_{1,0} \arrow[r, "f^2_{1,0}"'] & C_{1,1}
\end{tikzcd}$$

The cube of resolutions $\mathrm{CR}(\mathbb{G})$ consists of the 2×2 square on the lower right. The big cube of resolutions $\mathrm{BCR}(\mathbb{G})$ consists of the whole diagram, together with $C_{0,\infty}$, $C_{1,\infty}$, $f^1_{\infty,\infty}$, $f^1_{0,\infty}$, $f^2_{0,\infty}$, $f^2_{1,\infty}$, $\varphi^1_{\infty,\infty}$, $\varphi^2_{0,\infty}$ and $\varphi^2_{1,\infty}$, which are not shown. (These would all be at the lower left of the diagram.)

Remark 5.1 Neither $\mathrm{BCR}(\mathbb{G})$ nor $\mathrm{CR}(\mathbb{G})$ contains diagonal maps, eg a map that goes from $C_{\infty,0}$ to $C_{0,1}$ or $C_{1,1}$. This is in stark contrast with cubes of resolutions in many other contexts. For example, in [17], where the technique of spectral sequences was first applied to Heegaard Floer homology, there are diagonal maps in both the big $\{\infty, 0, 1\}^m$ cube and the small $\{0, 1\}^m$ cube. Such diagonal maps are needed because the edge maps f and φ only commute *up to chain homotopy*. Below, Lemma 5.3 will guarantee that our edge maps commute *on the nose*, allowing us to define the diagonal maps to be zero.

In the case with two crossings, observe that $C_{\infty,\infty}$ is quasi-isomorphic to the mapping cone of $f^2_{\infty,0}$ via $f^2_{\infty,\infty}+\varphi^2_{\infty,\infty}$. The fact that the diagram commutes now immediately implies that

(1) the cube of resolutions $\mathrm{CR}(\mathbb{G})$ is a chain complex;

(2) the sum $f^1_{\infty,0}+f^1_{\infty,1}+\varphi^1_{\infty,0}+\varphi^1_{\infty,1}$ is a chain map from the mapping cone of $f^2_{\infty,0}$ to the cube of resolutions $\mathrm{CR}(\mathbb{G})$; and

(3) this chain map is a quasi-isomorphism.

(To see the last statement, observe that if $f\colon C_1 \to C_2$ and $f'\colon C'_1 \to C'_2$ are quasi-isomorphisms, and if there exist maps $g_1\colon C_1 \to C'_1$ and $g_2\colon C_2 \to C'_2$ such that the diagram

$$\begin{tikzcd} C_1 \arrow[r, "f"] \arrow[d, "g_1"'] & C_2 \arrow[d, "g_2"] \\ C'_1 \arrow[r, "f'"] & C'_2 \end{tikzcd}$$

commutes, then $f+f'$ is a quasi-isomorphism between the mapping cone of g_1 and that of g_2.) Thus, we see that $C_{\infty,\infty}$ is quasi-isomorphic to $\mathrm{CR}(\mathbb{G})$.

The general case is similar; to be precise, our claim is the following.

Proposition 5.2 *The cubes of resolutions* $\mathrm{BCR}(\mathbb{G};\mathbb{F}_2)$ *and* $\mathrm{CR}(\mathbb{G};\mathbb{F}_2)$ *are indeed chain complexes. Moreover,* $\mathrm{CR}(\mathbb{G};\mathbb{F}_2)$ *is quasi-isomorphic to* $\widetilde{\mathrm{GC}}(\mathbb{G};\mathbb{F}_2)$. *As a consequence,* $\mathrm{BCR}(\mathbb{G};\mathbb{F}_2)$ *is acyclic.*

As mentioned, the main ingredient in proving Proposition 5.2 is the following lemma.

Lemma 5.3 *All maps involved commute. Precisely,*

$$f^{i_2}_{k_1,\dots,k_{i_1}+1,\dots,k_m}\circ f^{i_1}_{k_1,\dots,k_{i_1},\dots,k_m} = f^{i_1}_{k_1,\dots,k_{i_2}+1,\dots,k_m}\circ f^{i_2}_{k_1,\dots,k_{i_2},\dots,k_m},$$
$$\varphi^{i_2}_{k_1,\dots,k_{i_1}+1,\dots,k_m}\circ f^{i_1}_{k_1,\dots,k_{i_1},\dots,k_m} = f^{i_1}_{k_1,\dots,k_{i_2}+2,\dots,k_m}\circ \varphi^{i_2}_{k_1,\dots,k_{i_2},\dots,k_m}.$$

Proof Inspecting the 6×6 block in Figure 11, we see that there is an X near each corner of the block, and so each allowed polygon defined in Section 3 counted in a chain map or a chain homotopy can only leave the block either horizontally or vertically. If it leaves the block horizontally, then it is contained in the rows that the block spans; if it leaves the block vertically, then it is contained in the columns that the block spans. This shows that if two polygons from two different crossings share a corner, one must be long and horizontal, and the other long and vertical; in such cases, the composite domain always has an obvious alternative decomposition. The cases where the two polygons are disjoint or have overlapping interiors are obvious. □

Proof of Proposition 5.2 Exactly as in the case with two crossings, Lemma 5.3 implies that the cube of resolutions is indeed a chain complex, and that all appropriate maps are chain maps. We proceed by induction: at each step, we claim that the chain complexes

$$\Bigg(\bigoplus_{k_i\in\{0,1\}} C_{\infty,\dots,\infty,k_{t+1},\dots,k_m}, \sum_{k_i\in\{0,1\}}\Big(\partial_{\infty,\dots,\infty,k_{t+1},\dots,k_m} + \sum_{j:k_j=0} f^j_{\infty,\dots,\infty,k_{t+1},\dots,k_m}\Big)\Bigg)$$

and

$$\Bigg(\bigoplus_{k_i\in\{0,1\}} C_{\infty,\dots,\infty,k_t,\dots,k_m}, \sum_{k_i\in\{0,1\}}\Big(\partial_{\infty,\dots,\infty,k_t,\dots,k_m} + \sum_{j:k_j=0} f^j_{\infty,\dots,\infty,k_t,\dots,k_m}\Big)\Bigg)$$

are quasi-isomorphic. The quasi-isomorphism is given by

$$\sum_{k_i\in\{0,1\}} f^t_{\infty,\dots,\infty,k_{t+1},\dots,k_m} + \sum_{k_i\in\{0,1\}} \varphi^t_{\infty,\dots,\infty,k_{t+1},\dots,k_m}.$$

This map is a quasi-isomorphism by the induction hypothesis and by the comment after (3) in the case with two crossings above. □

Corollary 1.4 follows from Proposition 5.2.

Remark 5.4 Over $\mathbb{Z}$, for the cube of resolutions to be a chain complex, the analogue of Lemma 5.3 over $\mathbb{Z}$ should presumably state that the maps involved *anticommute*. However, if we denote by ρ^i (resp. τ^i) the maps defined by the rectangle-like (resp. triangle-like) polygons associated to the i^{th} crossing following the definitions in Section 4, then we have:

(1) $\rho^2 \circ \rho^1 = -\rho^1 \circ \rho^2$
(2) $\rho^2 \circ \tau^1 = -\tau^1 \circ \rho^2$
(3) $\tau^2 \circ \rho^1 = -\rho^1 \circ \tau^2$
(4) $\tau^2 \circ \tau^1 = \tau^1 \circ \tau^2$

Since *only* the maps defined by triangle-like polygons commute, the author has thus far been unable to prove a version of Lemma 5.3, and therefore Proposition 5.2, over $\mathbb{Z}$.

6 Quasialternating links

We now describe an application of the skein exact triangle. To begin, we must first define the δ–grading on knot Floer homology. The δ–grading is closely related to the Maslov and Alexander gradings; in view of that, in this section we revert to the traditional notation with both Os and Xs, and denote the set of Os by $\mathbb{O}$ and that of Xs by $\mathbb{X}$.

The following formulation is found in [12]. Given two collections A and B of finitely many points in the plane, let

$$\mathcal{J}(A,B) = \tfrac{1}{2}\#\big\{\big((a_1,a_2),(b_1,b_2)\big) \in A \times B \mid a_1 < b_1 \text{ and } a_2 < b_2\big\} \\ + \tfrac{1}{2}\#\big\{\big((a_1,a_2),(b_1,b_2)\big) \in A \times B \mid b_1 < a_1 \text{ and } b_2 < a_2\big\}.$$

Treating $\boldsymbol{x} \in \boldsymbol{S}(\mathbb{G})$ as a collection of points with integer coordinates in a fundamental domain for $\mathbb{T}$, and similarly $\mathbb{O}$ and $\mathbb{X}$ as collections of points in the plane with half-integer coordinates, the Maslov grading of a generator is given by

$$M(\boldsymbol{x}) = \mathcal{J}(\boldsymbol{x},\boldsymbol{x}) - 2\mathcal{J}(\boldsymbol{x},\mathbb{O}) + \mathcal{J}(\mathbb{O},\mathbb{O}) + 1,$$

while the Alexander grading is given by

$$A(\boldsymbol{x}) = \mathcal{J}(\boldsymbol{x},\mathbb{X}) - \mathcal{J}(\boldsymbol{x},\mathbb{O}) - \tfrac{1}{2}\mathcal{J}(\mathbb{X},\mathbb{X}) + \tfrac{1}{2}\mathcal{J}(\mathbb{O},\mathbb{O}) - \tfrac{1}{2}(n-1),$$

where n is the size of the grid diagram. Now the δ–grading is just

$$\delta(\boldsymbol{x}) = A(\boldsymbol{x}) - M(\boldsymbol{x});$$

in other words, we have

$$\delta(\boldsymbol{x}) = -\mathcal{J}(\boldsymbol{x},\boldsymbol{x}) + \mathcal{J}(\boldsymbol{x},\mathbb{X}) + \mathcal{J}(\boldsymbol{x},\mathbb{O}) - \tfrac{1}{2}\mathcal{J}(\mathbb{X},\mathbb{X}) - \tfrac{1}{2}\mathcal{J}(\mathbb{O},\mathbb{O}) - \tfrac{1}{2}(n+1).$$

These gradings do not depend on the choice of the fundamental domain; moreover, they agree with the original definitions in terms of pseudoholomorphic representatives.

It has been observed that, for many classes of links, the knot Floer homology over $R = \mathbb{F}_2$ or $\mathbb{Z}$ is a free R–module supported in only one δ–grading, which motivates the following definition by Rasmussen [19; 20] and Manolescu and Ozsváth [10]:

Definition 6.1 Let $R = \mathbb{F}_2$ or $\mathbb{Z}$. A link L is *Floer homologically thin* over R if $\widehat{\mathrm{HFK}}(L;R)$ is a free R–module supported in only one δ–grading. If in addition the δ–grading equals $-\frac{1}{2}\sigma(L)$, where $\sigma(L)$ is the signature of the link, then we say that L is *Floer homologically σ–thin*.

Manolescu and Ozsváth [10] showed that a class of links which is a natural generalization of alternating knots has the homologically σ–thin property over $\mathbb{F}_2$. Precisely, we recall the following definition from [17]:

Definition 6.2 The set of *quasialternating links* $\mathcal{Q}$ is the smallest set of links satisfying the following properties:

(1) The unknot is in $\mathcal{Q}$.

(2) If L_∞ is a link that admits a projection with a crossing such that
  (a) both resolutions L_0 and L_1 at that crossing are in $\mathcal{Q}$, and
  (b) $\det(L_\infty) = \det(L_0) + \det(L_1)$,
  then L_∞ is in $\mathcal{Q}$.

The result of Manolescu and Ozsváth [10] is then the following statement.

Theorem 6.3 (Manolescu and Ozsváth) *Quasialternating links are Floer homologically σ–thin over $\mathbb{F}_2$.*

The proof of the theorem is essentially an application of Manolescu's unoriented skein exact triangle; the main work is in tracking the changes in the δ–grading of the maps involved. The same idea works in the current context as well: with grid diagrams, we may similarly track the changes in the δ–grading, defined by the formula above. Of course, in our case, we will be proving the result for $\widehat{\mathrm{GH}}$ of a link instead of $\widehat{\mathrm{HFK}}$, as in Theorem 1.5. From this, we will obtain Theorem 1.6, a strengthened version of Theorem 6.3.

To begin, fix a crossing c_0 in the planar diagram of a link L_∞. Let L_+ be the link with a positive crossing at c_0, and L_- the link with a negative crossing; then either $L_\infty = L_+$ or $L_\infty = L_-$. Let L_h and L_v be the unoriented and oriented resolutions of L_∞ at c_0 respectively, and choose an arbitrary orientation for L_h. This is illustrated in Figure 12. Comparing with Figure 1, if $L_\infty = L_+$, then $L_0 = L_h$ and $L_1 = L_v$; if instead $L_\infty = L_-$, then $L_0 = L_v$ and $L_1 = L_h$.

Denote by D_+, D_-, D_v, D_h the planar diagrams of L_+, L_-, L_v, L_h, differing from each other only at c_0. The following lemma is used in [10]; for L_+, the first equality is

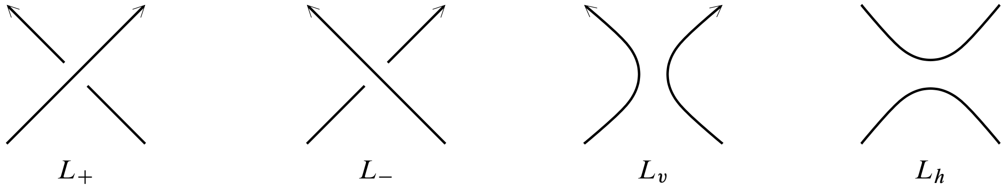

Figure 12: L_+, L_-, L_h and L_v near a point

proven by Murasugi [13], while the second is also inspired by a result of Murasugi [14]. The equalities for L_- are similarly obtained.

Lemma 6.4 *Suppose that* $\det(L_v), \det(L_h) > 0$. *Let* e *denote the difference between the number of negative crossings in* D_h *and the number of such crossings in* D_+. *If* $\det(L_+) = \det(L_v) + \det(L_h)$, *then*

(1) $\sigma(L_v) - \sigma(L_+) = 1$,

(2) $\sigma(L_h) - \sigma(L_+) = e$.

If $\det(L_-) = \det(L_v) + \det(L_h)$, *then*

(1) $\sigma(L_v) - \sigma(L_-) = -1$,

(2) $\sigma(L_h) - \sigma(L_-) = e$.

Now we investigate the changes in the δ–grading in the maps f_k defined in Section 3 and Section 4.

Given two grid diagrams $\mathbb{G}, \mathbb{G}'$ of the same size n, we can think of them as being on the same grid (ie consisting of the same horizontal and vertical circles) but having different Os and Xs; therefore, we can write $\mathbb{G} = (n, \mathbb{O}, \mathbb{X})$ and $\mathbb{G}' = (n, \mathbb{O}', \mathbb{X}')$. This allows us to identify $\boldsymbol{S}(\mathbb{G})$ and $\boldsymbol{S}(\mathbb{G}')$, by viewing each generator $\boldsymbol{x} \in \boldsymbol{S}(\mathbb{G})$ as the permutation $\sigma_{\boldsymbol{x}}$ referred to in Remark 4.3. In this point of view, we will write $\delta_{\mathbb{G}}(\boldsymbol{x})$ and $\delta_{\mathbb{G}'}(\boldsymbol{x})$ to denote the gradings defined by applying the δ–grading formula to $(\mathbb{O}, \mathbb{X})$ and $(\mathbb{O}', \mathbb{X}')$, respectively.

In particular, for the rest of this section, we will view $\mathbb{G}_\infty, \mathbb{G}_0, \mathbb{G}_1$ in this manner.

Lemma 6.5 *Let* $\boldsymbol{x} \in \boldsymbol{S}(\mathbb{G})$ *be a fixed generator in a grid diagram* $\mathbb{G} = (n, \mathbb{O}, \mathbb{X})$. *Then:*

(1) *If there is an empty rectangle* r *from* $\boldsymbol{x}$ *to* $\boldsymbol{y}$, *possibly containing* Os *and* Xs, *then* $\delta_{\mathbb{G}}(\boldsymbol{y}) = \delta_{\mathbb{G}}(\boldsymbol{x}) + 1 - \#(\mathrm{Int}(r) \cap (\mathbb{O} \cup \mathbb{X}))$.

(2) *Suppose* $\mathbb{G}' = (n, \mathbb{O}', \mathbb{X})$ *is a grid diagram identical to* $\mathbb{G}$ *except in two adjacent columns, where the horizontal positions of a pair of* O *markers are interchanged, as in Figure 13. Then*

  (a) $\delta_{\mathbb{G}'}(\boldsymbol{x}) - \delta_{\mathbb{G}}(\boldsymbol{x}) = -\frac{1}{2}$ *if, in* $\mathbb{G}$, *the component of* $\boldsymbol{x}$ *on the vertical circle between the markers lies to the northeast of one marker and to the southwest of the other (the component of* $\boldsymbol{x}$ *indicated in the diagrams by a solid point);*

  (b) $\delta_{\mathbb{G}'}(\boldsymbol{x}) - \delta_{\mathbb{G}}(\boldsymbol{x}) = \frac{1}{2}$ *otherwise (the component of* $\boldsymbol{x}$ *indicated by a hollow point).*

  The same statement holds if, instead, the horizontal positions of a pair of X *markers are interchanged, with* $\mathbb{G}' = (n, \mathbb{O}, \mathbb{X}')$.

(3) *If the diagram $\mathbb{G}'$ is obtained from $\mathbb{G}$ by reversing the orientation of a component K of L, then $\delta_{\mathbb{G}'}(\boldsymbol{x}) - \delta_{\mathbb{G}}(\boldsymbol{x}) = -\frac{\epsilon}{2}$, where ϵ is the difference between the number of negative crossings in $\mathbb{G}'$ and the number of such crossings in $\mathbb{G}$.*

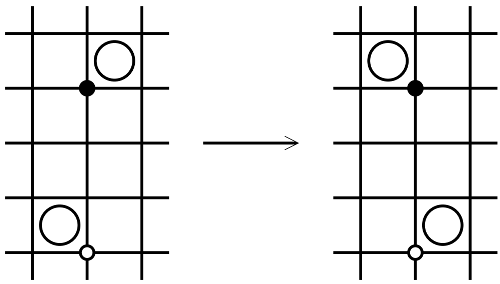

Figure 13: Moving two markers across a vertical circle. Observe that before the move, the solid point lies to the northeast of one of the Os and to the southwest of the other, thus forming two southwest–northeast pairs with the markers that are destroyed in the move. The hollow point forms the same number of pairs with the markers before and after the move.

Proof Recall that

$$\delta(\boldsymbol{x}) = -\mathcal{J}(\boldsymbol{x},\boldsymbol{x}) + \mathcal{J}(\boldsymbol{x},\mathbb{X}) + \mathcal{J}(\boldsymbol{x},\mathbb{O}) - \tfrac{1}{2}\mathcal{J}(\mathbb{X},\mathbb{X}) - \tfrac{1}{2}\mathcal{J}(\mathbb{O},\mathbb{O}) - \tfrac{1}{2}(n+1).$$

Observe that $\mathcal{J}(A, B)$ is the number of southwest–northeast pairs between A and B, divided by two. We make the following observations.

(1) Observe that the last three terms in $\delta(\boldsymbol{x})$ and $\delta(\boldsymbol{y})$ are identical. Switching from $\boldsymbol{x}$ to $\boldsymbol{y}$ destroys exactly one southwest–northeast pair, which is counted twice and contributes $+1$ to $\mathcal{J}(\boldsymbol{x},\boldsymbol{x})$. For every O or X inside $\mathrm{Int}(r)$, switching from $\boldsymbol{x}$ to $\boldsymbol{y}$ destroys two southwest–northeast pairs, each counted once, that contribute $+1$ to $\mathcal{J}(\boldsymbol{x},\mathbb{X}) + \mathcal{J}(\boldsymbol{x},\mathbb{O})$.

(2) Consider the first case, where $\mathbb{G} = (n,\mathbb{O},\mathbb{X})$ and $\mathbb{G}' = (n,\mathbb{O}',\mathbb{X})$, as in Figure 13. We first see that $\mathcal{J}(\mathbb{O}',\mathbb{O}') - \mathcal{J}(\mathbb{O},\mathbb{O}) = -1$. If the component of $\boldsymbol{x}$ on the vertical circle lies to the northeast of one marker and to the southwest of the other, then $\mathcal{J}(\boldsymbol{x},\mathbb{O}') - \mathcal{J}(\boldsymbol{x},\mathbb{O}) = -1$; otherwise, this quantity is 0. All other terms in $\delta_{\mathbb{G}}(\boldsymbol{x})$ and $\delta_{\mathbb{G}'}(\boldsymbol{x})$ are identical. The second case is entirely analogous.

(3) This is simply a restatement of [12, Proposition 5.4].

This completes the proof of the lemma. □

The following is an easy consequence of Lemma 6.5(1)–(2).

Lemma 6.6 *Suppose $\mathbb{G}_k$ and $\mathbb{G}_{k+1}$ are grid presentations of two links with compatible orientations (so $\mathbb{G}_k$ and $\mathbb{G}_{k+1}$ have compatible sets of O s), and $f_k\colon \mathbb{G}_k \to \mathbb{G}_{k+1}$ is the map defined in Sections 3 and 4. If $\boldsymbol{x} \in \boldsymbol{S}(\mathbb{G}_k)$, and $\boldsymbol{y} \in \boldsymbol{S}(\mathbb{G}_{k+1})$ appears in $f_k(\boldsymbol{x})$, then $\delta_{\mathbb{G}_{k+1}}(\boldsymbol{y}) - \delta_{\mathbb{G}_k}(\boldsymbol{x}) = +\frac{1}{2}$.*

Proof Note first that the grid diagrams $\mathbb{G}_k = (n, \mathbb{O}, \mathbb{X})$ and $\mathbb{G}_{k+1} = (n, \mathbb{O}', \mathbb{X}')$ are related as in Lemma 6.5(2). Refer to Figure 14. Without loss of generality, assume that the two relevant markers are O s. (The case where they are Xs is completely analogous.) Label the two markers in $\mathbb{O}$ by O_1 and O_2, so that O_1 lies to the northeast of O_2; similarly, label the markers in $\mathbb{O}'$ by O_1' and O_2', so that O_1' lies to the northwest of O_2'. In the current framework, β_∞, β_0 and β_1 are all represented by the vertical circle β in the middle. Let x be the β–component of $\boldsymbol{x}$.

Suppose $\boldsymbol{y} \in \boldsymbol{S}(\mathbb{G}_{k+1})$ appears in $f_k(\boldsymbol{x})$. Then

$$\delta_{\mathbb{G}_{k+1}}(\boldsymbol{y}) - \delta_{\mathbb{G}_k}(\boldsymbol{x}) = (\delta_{\mathbb{G}_{k+1}}(\boldsymbol{x}) - \delta_{\mathbb{G}_k}(\boldsymbol{x})) + (\delta_{\mathbb{G}_{k+1}}(\boldsymbol{y}) - \delta_{\mathbb{G}_{k+1}}(\boldsymbol{x})).$$

There are two cases, as follows.

Suppose there is an empty pentagon p from $\boldsymbol{x}$ to $\boldsymbol{y}$; in the framework described in this section, it can in fact be viewed as an empty rectangle r from $\boldsymbol{x}$ to $\boldsymbol{y}$ in $\mathbb{G}_{k+1}$. Observe that the horizontal circles divide $\mathbb{T}$ into a number of components; let A be the unique component containing both O_1 in $\mathbb{G}_k$ and O_1' in $\mathbb{G}_{k+1}$. Since the boundary of p contains u_k (cf Figure 3), which lies in A, the boundary of r must contain $\beta \cap A$. Therefore, one of two scenarios must be true:

(1) The point x lies to the northeast of O_2 and to the southwest of O_1. Then $\mathrm{Int}(r) \cap (\mathbb{O} \cup \mathbb{X}) = \varnothing$. By Lemma 6.5(2)(a), $\delta_{\mathbb{G}_{k+1}}(\boldsymbol{x}) - \delta_{\mathbb{G}_k}(\boldsymbol{x}) = -\frac{1}{2}$, and by Lemma 6.5(1), $\delta_{\mathbb{G}_{k+1}}(\boldsymbol{y}) - \delta_{\mathbb{G}_{k+1}}(\boldsymbol{x}) = 1$. This is illustrated by the top row of Figure 14.

(2) The point x lies elsewhere. Then $\mathrm{Int}(r) \cap (\mathbb{O} \cup \mathbb{X})$ is either $\{O_1'\}$ or $\{O_2'\}$. By Lemma 6.5(2)(b), $\delta_{\mathbb{G}_{k+1}}(\boldsymbol{x}) - \delta_{\mathbb{G}_k}(\boldsymbol{x}) = +\frac{1}{2}$, and by Lemma 6.5(1), we have $\delta_{\mathbb{G}_{k+1}}(\boldsymbol{y}) - \delta_{\mathbb{G}_{k+1}}(\boldsymbol{x}) = 1 - 1 = 0$. This is illustrated by the middle row of Figure 14.

Either way, we get that $\delta_{\mathbb{G}_{k+1}}(\boldsymbol{y}) - \delta_{\mathbb{G}_k}(\boldsymbol{x}) = +\frac{1}{2}$.

Suppose there is a triangle p from $\boldsymbol{x}$ to $\boldsymbol{y}$; then, in the framework of this section, $\boldsymbol{x} = \boldsymbol{y}$, and so $\delta_{\mathbb{G}_{k+1}}(\boldsymbol{y}) - \delta_{\mathbb{G}_{k+1}}(\boldsymbol{x}) = 0$. Since the boundary of p contains v_k (cf Figure 3), we see that $\mathrm{Int}(p) \subset \mathrm{Int}(t_k) \cup \mathrm{Int}(t_{k+1})$. This implies that x cannot lie both to the northeast of O_2 and to the southwest of O_1. Thus, by Lemma 6.5(2)(b), $\delta_{\mathbb{G}_{k+1}}(\boldsymbol{x}) - \delta_{\mathbb{G}_k}(\boldsymbol{x}) = +\frac{1}{2}$. This case is illustrated by the bottom row of Figure 14. □

We can now prove a graded version of Theorem 1.3.

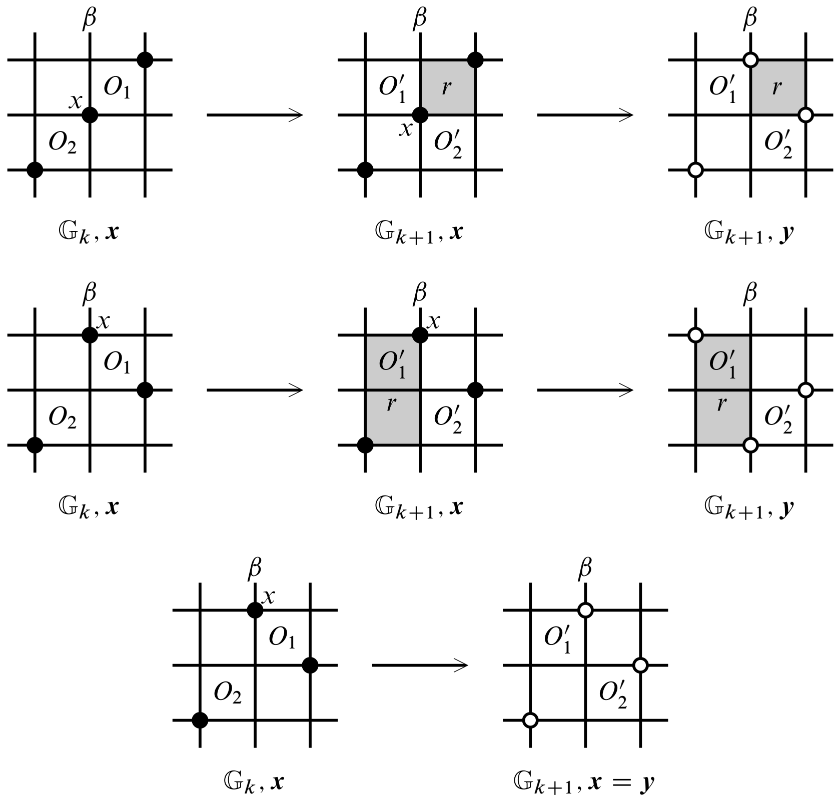

Figure 14: Computing the δ–grading change under f_k when $\mathbb{G}_k$ and $\mathbb{G}_{k+1}$ have compatible orientations. The top and middle rows show the case when the polygon p being counted is a pentagon, while the bottom row shows the case when p is a triangle. In both cases, $\boldsymbol{y}$ appears in $f_k(\boldsymbol{x})$. Note that there may be multiple horizontal circles between the O s, which are omitted for the sake of simplicity, in each of the figures above.

Proposition 6.7 *With respect to the δ–grading the skein exact sequence in Theorem 1.3 can be written as*

$$\cdots \to \widehat{\mathrm{GH}}_{*-\frac{1}{2}}(L_v; R) \otimes V^{n-\ell_v} \to \widehat{\mathrm{GH}}_*(L_+; R) \otimes V^{n-\ell_+} \to \widehat{\mathrm{GH}}_{*-\frac{e}{2}}(L_h; R) \otimes V^{n-\ell_h} \to \widehat{\mathrm{GH}}_{*-\frac{1}{2}+1}(L_v; R) \otimes V^{n-\ell_v} \to \cdots$$

and

$$\cdots \to \widehat{\mathrm{GH}}_{*-\frac{e}{2}}(L_h; R) \otimes V^{n-\ell_h} \to \widehat{\mathrm{GH}}_*(L_-; R) \otimes V^{n-\ell_-} \to \widehat{\mathrm{GH}}_{*+\frac{1}{2}}(L_v; R) \otimes V^{n-\ell_v} \to \widehat{\mathrm{GH}}_{*-\frac{e}{2}+1}(L_h; R) \otimes V^{n-\ell_h} \to \cdots ,$$

where $R = \mathbb{F}_2$ or $\mathbb{Z}$, V is a free module of rank 2 over R with grading zero, and e is as in the statement of Lemma 6.4.

Proof Suppose L_∞ has a positive crossing at c_0, so that $L_+ = L_\infty$, $L_v = L_1$ and $L_h = L_0$. Let $\mathbb{G}_\infty, \mathbb{G}_0, \mathbb{G}_1$ be grid diagrams for L_∞, L_0, L_1 respectively, differing with each other only at c_0 as in Figure 2. Note that in Figure 2 only Xs are used as markers, whereas in the present context both Xs and Os are used. Then we are to prove that in the exact sequence

$$\cdots \to \widetilde{\mathrm{GH}}(\mathbb{G}_1) \xrightarrow{(f_1)_*} \widetilde{\mathrm{GH}}(\mathbb{G}_\infty) \xrightarrow{(f_\infty)_*} \widetilde{\mathrm{GH}}(\mathbb{G}_0) \xrightarrow{(f_0)_*} \widetilde{\mathrm{GH}}(\mathbb{G}_1) \to \cdots,$$

the map f_1 shifts the δ–grading by $+\frac{1}{2}$, f_∞ by $-\frac{e}{2}$, and f_0 by $+\frac{1}{2}(e+1)$.

Refer to Figure 15. Since L_1 and L_∞ have compatible orientations, Lemma 6.6 shows that the map f_1 shifts the δ–grading by $+\frac{1}{2}$.

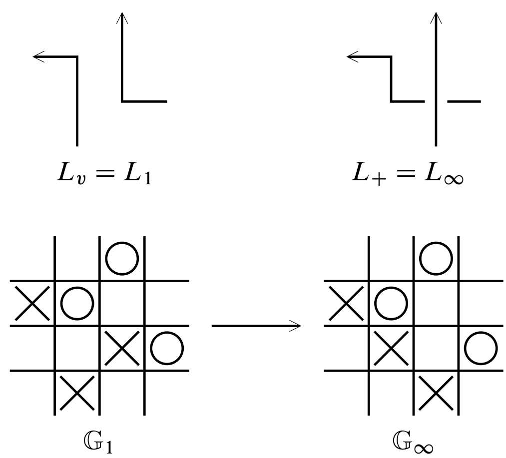

Figure 15: A straightforward application of Lemma 6.6 gives the δ–grading shift of f_1

Let us now focus on f_∞. Let $\boldsymbol{x} \in \boldsymbol{S}(\mathbb{G}_\infty)$ be a generator, and suppose $\boldsymbol{y} \in \boldsymbol{S}(\mathbb{G}_0)$ appears in $f_\infty(\boldsymbol{x})$. We proceed according to whether the two strands of the link L_+ meeting at the crossing c_0 belong to different components, or to the same component, of L_+, as follows.

Suppose the strands belong to different components; see Figure 16. Starting from $\mathbb{G}_\infty$, we can reverse the orientation of one of the two components to obtain a new diagram $\mathbb{G}'_\infty$; then we can write

$$\delta_{\mathbb{G}_0}(\boldsymbol{y}) - \delta_{\mathbb{G}_\infty}(\boldsymbol{x}) = (\delta_{\mathbb{G}'_\infty}(\boldsymbol{x}) - \delta_{\mathbb{G}_\infty}(\boldsymbol{x})) + (\delta_{\mathbb{G}_0}(\boldsymbol{y}) - \delta_{\mathbb{G}'_\infty}(\boldsymbol{x})).$$

By Lemma 6.5(3), we have $\delta_{\mathbb{G}'_\infty}(\boldsymbol{x}) - \delta_{\mathbb{G}_\infty}(\boldsymbol{x}) = -\frac{\epsilon}{2}$, where ϵ is the difference between the number of negative crossings in $\mathbb{G}'_\infty$ and the number of such crossings in $\mathbb{G}_\infty$ (which represents L_+). Now observe that $\mathbb{G}'_\infty$ and $\mathbb{G}_0$ have compatible orientations;

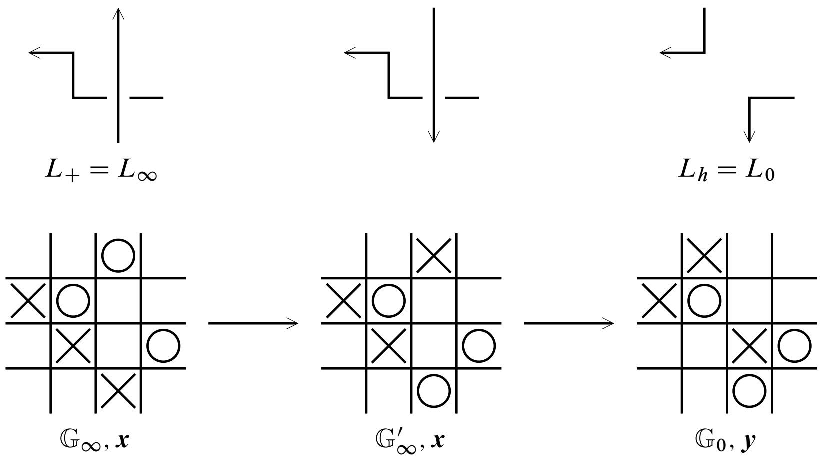

Figure 16: Computation of the δ–grading shift of f_∞, when the two strands meeting at c_0 belong to different components. The generators $\boldsymbol{x}$ and $\boldsymbol{y}$ are not shown, as their positions do not affect the argument.

also, $\mathbb{G}'_\infty$ has one more negative crossing, near c_0, than does $\mathbb{G}_0$ (which represents L_h). Therefore, $\epsilon = e + 1$. Furthermore, the fact that $\mathbb{G}'_\infty$ and $\mathbb{G}_0$ have compatible orientations implies that Lemma 6.6 can be applied, so $\delta_{\mathbb{G}_0}(\boldsymbol{y}) - \delta_{\mathbb{G}'_\infty}(\boldsymbol{x}) = +\frac{1}{2}$. Thus,

$$\delta_{\mathbb{G}_0}(\boldsymbol{y}) - \delta_{\mathbb{G}_\infty}(\boldsymbol{x}) = -\tfrac{1}{2}(e+1) + \tfrac{1}{2} = -\tfrac{e}{2}.$$

Suppose now the strands belong to the same component. As in the proof of Lemma 6.6, we write

$$\delta_{\mathbb{G}_0}(\boldsymbol{y}) - \delta_{\mathbb{G}_\infty}(\boldsymbol{x}) = (\delta_{\mathbb{G}_0}(\boldsymbol{x}) - \delta_{\mathbb{G}_\infty}(\boldsymbol{x})) + (\delta_{\mathbb{G}_0}(\boldsymbol{y}) - \delta_{\mathbb{G}_0}(\boldsymbol{x})).$$

Note that this is different from the equation we use when the strands belong to different components. We first compute $\delta_{\mathbb{G}_0}(\boldsymbol{x}) - \delta_{\mathbb{G}_\infty}(\boldsymbol{x})$; see Figure 17. Since the two strands in L_+ belong to the same component, the two strands in L_v must belong to different components. Therefore, we can reverse the orientation of one of these components in $\mathbb{G}_1$ (which represents L_v) to obtain a new diagram $\mathbb{G}'_1$; then

$$\delta_{\mathbb{G}_0}(\boldsymbol{x}) - \delta_{\mathbb{G}_\infty}(\boldsymbol{x}) = (\delta_{\mathbb{G}_1}(\boldsymbol{x}) - \delta_{\mathbb{G}_\infty}(\boldsymbol{x})) + (\delta_{\mathbb{G}'_1}(\boldsymbol{x}) - \delta_{\mathbb{G}_1}(\boldsymbol{x})) + (\delta_{\mathbb{G}_0}(\boldsymbol{x}) - \delta_{\mathbb{G}'_1}(\boldsymbol{x})).$$

Let x be the β–component of $\boldsymbol{x}$. We can now compute, by Lemma 6.5(2),

$$\delta_{\mathbb{G}_1}(\boldsymbol{x}) - \delta_{\mathbb{G}_\infty}(\boldsymbol{x}) = \begin{cases} +\frac{1}{2} & \text{if } x \text{ lies at the hollow square in Figure 17,} \\ -\frac{1}{2} & \text{otherwise.} \end{cases}$$

Now by Lemma 6.5(3), $\delta_{\mathbb{G}'_1}(\boldsymbol{x}) - \delta_{\mathbb{G}_1}(\boldsymbol{x}) = -\frac{\epsilon}{2}$, where ϵ is the difference between the number of negative crossings in $\mathbb{G}'_1$ and the number of such crossings in $\mathbb{G}_1$. Observe that $\mathbb{G}_1$ and $\mathbb{G}_\infty$ (which represents L_+) have compatible orientations and hence the same number of negative crossings. Similarly, $\mathbb{G}_1'$ and $\mathbb{G}_0$ have compatible orientations and hence the same number of negative crossings. Therefore, we see that $\epsilon = e$. Next, again by Lemma 6.5(2), we have

$$\delta_{\mathbb{G}_0}(\boldsymbol{x}) - \delta_{\mathbb{G}_1'}(\boldsymbol{x}) = \begin{cases} -\frac{1}{2} & \text{if } x \text{ lies at the solid point, hollow point} \\ & \text{or hollow square in Figure 17;} \\ +\frac{1}{2} & \text{otherwise.} \end{cases}$$

Combining these calculations, we conclude that

$$\delta_{\mathbb{G}_0}(\boldsymbol{x}) - \delta_{\mathbb{G}_\infty}(\boldsymbol{x}) = \begin{cases} -\frac{1}{2}(e+2) & \text{if } x \text{ lies at the solid point or hollow point} \\ & \text{in Figure 17;} \\ -\frac{e}{2} & \text{otherwise.} \end{cases}$$

We now return to computing $\delta_{\mathbb{G}_0}(\boldsymbol{y}) - \delta_{\mathbb{G}_\infty}(\boldsymbol{x})$.

Suppose there is an empty pentagon p from $\boldsymbol{x}$ to $\boldsymbol{y}$; in the framework described in this section, it can in fact be viewed as an empty rectangle r from $\boldsymbol{x}$ to $\boldsymbol{y}$ in $\mathbb{G}_0$. By an argument similar to that in the proof of Lemma 6.6, one of two scenarios must hold:

(1) If the point x lies at the solid point or at the hollow point in Figure 17, then $\mathrm{Int}(r) \cap (\mathbb{O} \cup \mathbb{X}) = \varnothing$. By Lemma 6.5(1), $\delta_{\mathbb{G}_0}(\boldsymbol{y}) - \delta_{\mathbb{G}_0}(\boldsymbol{x}) = 1$.

(2) If the point x lies elsewhere, then $\mathrm{Int}(r) \cap (\mathbb{O} \cup \mathbb{X})$ contains exactly one point. By Lemma 6.5(1), $\delta_{\mathbb{G}_0}(\boldsymbol{y}) - \delta_{\mathbb{G}_0}(\boldsymbol{x}) = 1 - 1 = 0$.

Combining with our earlier calculation, we get that $\delta_{\mathbb{G}_0}(\boldsymbol{y}) - \delta_{\mathbb{G}_\infty}(\boldsymbol{x}) = -\frac{e}{2}$.

Suppose there is a triangle p from $\boldsymbol{x}$ to $\boldsymbol{y}$; then, in the framework of this section, $\boldsymbol{x} = \boldsymbol{y}$, and so $\delta_{\mathbb{G}_0}(\boldsymbol{y}) - \delta_{\mathbb{G}_0}(\boldsymbol{x}) = 0$. Since the boundary of p contains v_∞ (cf Figure 3), we see that $\mathrm{Int}(p) \subset \mathrm{Int}(t_\infty) \cup \mathrm{Int}(t_0)$. This implies that x cannot lie at the solid point or at the hollow point in Figure 17. Therefore, by our earlier calculation, again we have $\delta_{\mathbb{G}_0}(\boldsymbol{y}) - \delta_{\mathbb{G}_\infty}(\boldsymbol{x}) = -\frac{e}{2}$.

Finally, a calculation analogous to that for f_∞ can be done for f_0, and we obtain, in this case, that the shift in the δ–grading is $+\frac{1}{2}(e+1)$.

The case when $L_\infty = L_-$ has a negative crossing at c_0 is similar. □

Remark 6.8 The proof above only shows that the maps f_k are graded chain maps in the exact sequence. To show that $\widetilde{\mathrm{GC}}(\mathbb{G}_k)$ is *graded* quasi-isomorphic to the mapping cone of f_{k+1} (cf Section 5), we would have to show analogous statements for the chain homotopies φ_k also. The proof of such statements are omitted here, but are completely analogous to the proof of Proposition 6.7.

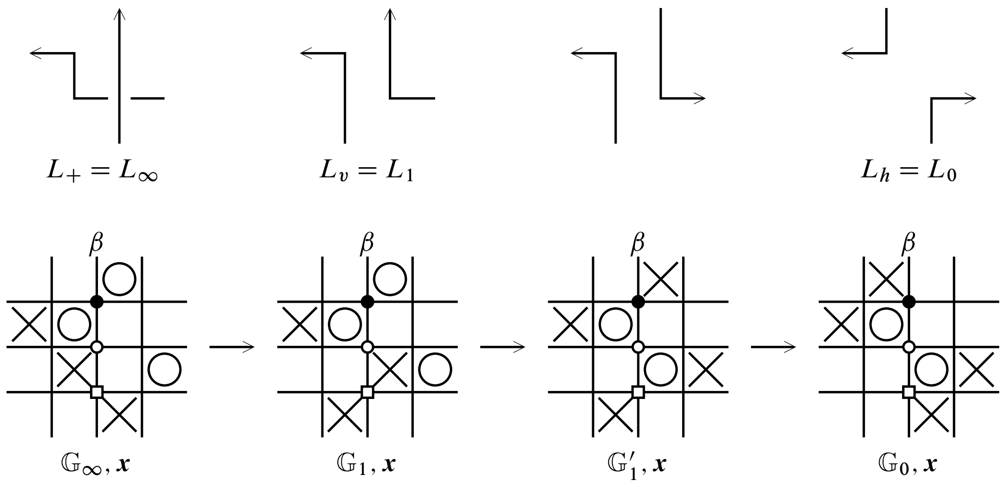

Figure 17: Computation of the δ–grading shift of f_∞, when the two strands meeting at c_0 belong to the same component. The computation depends on the position of x, the β–component of $\boldsymbol{x}$; the special cases are indicated by the solid point, the hollow point, and the hollow square.

Proof of Theorem 1.5 This is just a restatement of Proposition 6.7, taking into account the result of Lemma 6.4. □

Proof of Theorem 1.6 By definition, every quasialternating link has nonzero determinant. It can easily be checked that the unknot is homologically σ–thin. If L_0 and L_1 are resolutions of L_∞ that are quasialternating, then by induction, L_0 and L_1 are homologically σ–thin. The exact sequence in Theorem 1.5 collapses into short exact sequences, all but one of which are zero. Thus L_∞ is also homologically σ–thin. □

References

[1] **J A Baldwin**, *On the spectral sequence from Khovanov homology to Heegaard Floer homology*, Int. Math. Res. Not. 2011 (2011) 3426–3470 MR

[2] **J A Baldwin**, **A S Levine**, *A combinatorial spanning tree model for knot Floer homology*, Adv. Math. 231 (2012) 1886–1939 MR

[3] **D Bar-Natan**, *On Khovanov's categorification of the Jones polynomial*, Algebr. Geom. Topol. 2 (2002) 337–370 MR

[4] **P R Cromwell**, *Embedding knots and links in an open book, I: Basic properties*, Topology Appl. 64 (1995) 37–58 MR

[5] **I A Dynnikov**, *Arc-presentations of links: monotonic simplification*, Fund. Math. 190 (2006) 29–76 MR

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

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

[8] **R Lipshitz**, **P S Ozsváth**, **D P Thurston**, *Bordered Floer homology and the spectral sequence of a branched double cover, I*, J. Topol. 7 (2014) 1155–1199 MR

[9] **C Manolescu**, *An unoriented skein exact triangle for knot Floer homology*, Math. Res. Lett. 14 (2007) 839–852 MR

[10] **C Manolescu**, **P Ozsváth**, *On the Khovanov and knot Floer homologies of quasi-alternating links*, from "Proceedings of Gökova Geometry–Topology Conference 2007" (S Akbulut, T Önder, R J Stern, editors), GGT, Gökova (2008) 60–81 MR

[11] **C Manolescu**, **P Ozsváth**, **S Sarkar**, *A combinatorial description of knot Floer homology*, Ann. of Math. 169 (2009) 633–660 MR

[12] **C Manolescu**, **P Ozsváth**, **Z Szabó**, **D Thurston**, *On combinatorial link Floer homology*, Geom. Topol. 11 (2007) 2339–2412 MR

[13] **K Murasugi**, *On a certain numerical invariant of link types*, Trans. Amer. Math. Soc. 117 (1965) 387–422 MR

[14] **K Murasugi**, *On the signature of links*, Topology 9 (1970) 283–298 MR

[15] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and knot invariants*, Adv. Math. 186 (2004) 58–116 MR

[16] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and topological invariants for closed three-manifolds*, Ann. of Math. 159 (2004) 1027–1158 MR

[17] **P Ozsváth**, **Z Szabó**, *On the Heegaard Floer homology of branched double-covers*, Adv. Math. 194 (2005) 1–33 MR

[18] **P Ozsváth**, **Z Szabó**, *Holomorphic disks, link invariants and the multi-variable Alexander polynomial*, Algebr. Geom. Topol. 8 (2008) 615–692 MR

[19] **J Rasmussen**, *Floer homology and knot complements*, PhD thesis, Harvard University (2003) arXiv

[20] **J Rasmussen**, *Knot polynomials and knot homologies*, from "Geometry and topology of manifolds" (H U Boden, I Hambleton, A J Nicas, B D Park, editors), Fields Inst. Commun. 47, Amer. Math. Soc., Providence, RI (2005) 261–280 MR

[21] **S Sarkar**, *A note on sign conventions in link Floer homology*, Quantum Topol. 2 (2011) 217–239 MR

Department of Mathematics, Columbia University
New York, NY 10027, United States

`cmmwong@math.columbia.edu`

`http://www.math.columbia.edu/~cmmwong`

Received: 26 May 2013 Revised: 9 September 2016

Algebraic & Geometric Topology 17 (2017) 1323–1373

Pattern-equivariant homology

JAMES J WALTON

Pattern-equivariant (PE) cohomology is a well-established tool with which to interpret the Čech cohomology groups of a tiling space in a highly geometric way. We consider homology groups of PE infinite chains and establish Poincaré duality between the PE cohomology and PE homology. The Penrose kite and dart tilings are taken as our central running example; we show how through this formalism one may give highly approachable geometric descriptions of the generators of the Čech cohomology of their tiling space. These invariants are also considered in the context of rotational symmetry. Poincaré duality fails over integer coefficients for the "ePE homology groups" based upon chains which are PE with respect to orientation-preserving Euclidean motions between patches. As a result we construct a new invariant, which is of relevance to the cohomology of rotational tiling spaces. We present an efficient method of computation of the PE and ePE (co)homology groups for hierarchical tilings.

52C23; 37B50, 52C22, 55N05

Introduction

In the past few decades a rich class of highly ordered patterns has emerged whose central examples, despite lacking global translational symmetries, exhibit intricate internal structure, imbuing these patterns with properties akin to those enjoyed by periodically repeating patterns. The field of *aperiodic order* aims to study such patterns, and to establish connections between their properties, and their constructions, to other fields of mathematics and the natural sciences. To name a few, aperiodic order has interactions with areas of mathematics such as mathematical logic — see Lafitte and Weiss [30] — as established by Berger's proof [7] of the undecidability of the domino problem; Diophantine approximation (see Arnoux, Berthé, Ei and Ito [2], Berthé and Siegel [8], Haynes, Kelly and Weiss [21] or Haynes, Koivusalo and Walton [22]); the structure of attractors by Clark and Hunton [12]; and symbolic dynamics (see Schmidt [40]). Outside of pure mathematics, aperiodic order's most notable impetus comes from solid state physics, in the wake of the discovery of quasicrystals; Shechtman, Blech, Gratias and Cahn [41].

A full understanding of a periodic tiling, modulo locally defined reversible redecorations, amounts to an understanding of its symmetry group. In the aperiodic setting, the

Published: 17 July 2017 DOI: 10.2140/agt.2017.17.1323

complexity and incredible diversity of examples demands a multifaceted approach. Techniques from the theory of groupoids (see Bellissard, Julien and Savinien [6]), semigroups (see Kellendonk and Lawson [26]), C^*–algebras (see Anderson and Putnam [1]), dynamical systems (see Clark and Sadun [13] and Kellendonk [25]), ergodic theory (see Radin [35]) and shape theory (see Clark and Hunton [12]) find natural rôles in the field, and of course these tools have tightly knit connections to each other; see Kellendonk and Putnam [27]. One approach to studying a given aperiodic tiling $\mathfrak{T}$ is to associate to it a moduli space Ω, sometimes called the *tiling space*, of locally indistinguishable tilings imbued with a natural topology; see Sadun's book [38] for an accessible introduction to the theory. A central goal is then to formulate methods of computing topological invariants of Ω, and to describe what these invariants actually tell us about the original tiling $\mathfrak{T}$. An important perspective, particularly for the latter half of this objective, is provided by Kellendonk and Putnam's theory of pattern-equivariant (PE) cohomology; see Kellendonk [24] and Kellendonk and Putnam [28]. PE cohomology allows for an intuitive description of the Čech cohomology $\check{H}^\bullet(\Omega)$ of tiling spaces. Over $\mathbb{R}$ coefficients the PE cochain groups may be defined using PE differential forms [24], and over general abelian coefficients, when the tiling has a cellular structure, with PE cellular cochains; see Sadun [37]. Rather than just providing a reflection of topological invariants of tiling spaces, on the contrary, these PE invariants are of principal relevance to aperiodic structures and their connections with other fields in their own right; see, for example, Kelly and Sadun's use of them [29] in a topological proof of theorems of Kesten and Oren regarding the discrepancy of irrational rotations. It is perhaps more appropriate to view the isomorphism between $\check{H}^\bullet(\Omega)$ and the PE cohomology as an elegant interpretation of the *PE cohomology*, rather than vice versa, of theoretical and computational importance.

In this paper we introduce the *pattern-equivariant homology groups* of a tiling. These homology groups are based on infinite, or noncompactly supported cellular chains, sometimes known as "Borel–Moore chains". We say that such a chain is *pattern-equivariant* if there exists some $r > 0$ for which the coefficient of a cell only depends on the translation class of that cell and its surrounding patch of tiles to radius r. We show in Theorem 2.2, via a classical "cell, dual-cell" argument, that for a tiling of finite local complexity (see Section 1.1) we have PE Poincaré duality:

Theorem 2.2 *For a polytopal tiling $\mathfrak{T}$ of $\mathbb{R}^d$ of finite local complexity, we have PE Poincaré duality $H^\bullet(\mathfrak{T}^1) \cong H_{d-\bullet}(\mathfrak{T}^1)$ between the PE cohomology and PE homology of $\mathfrak{T}$.*

The upshot of this is that one may give quite beautiful, and informative, geometric depictions of the elements of the Čech cohomology groups $\check{H}^\bullet(\Omega)$ of tiling spaces. For

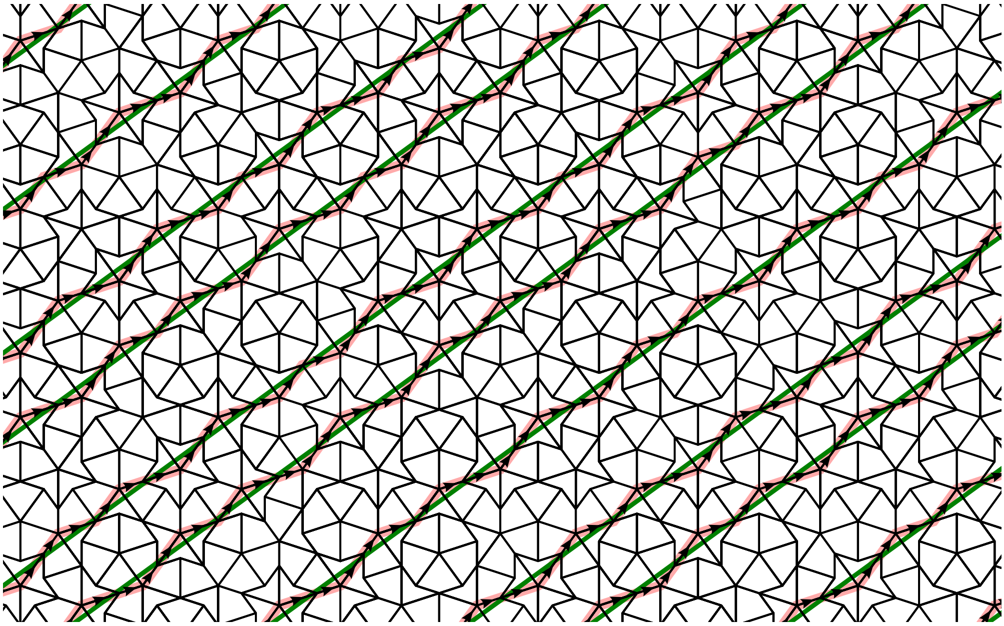

Figure 1: The PE 1–cycle ν_0, with green Ammann bars of the supertiling.

example, the cohomology of the translational hull $\Omega^1_{\mathfrak{T}}$ of a Penrose kite and dart tiling in degree one is $\check{H}^1(\Omega^1_{\mathfrak{T}}) \cong H^1(\mathfrak{T}^1) \cong H_1(\mathfrak{T}^1) \cong \mathbb{Z}^5$. The generators of this group have down-to-Earth interpretations in terms of important geometric features of the Penrose tilings. For example, one such generator, depicted in Figure 1, is closely linked to Ammann bars of the Penrose tilings (of which, see the discussion of Grünbaum and Shephard [20, Section 10.6]). Another simple geometric feature of the Penrose tilings is that the dart tiles arrange as loops, leading to the cycle depicted in Figure 2. As described in Example 2.3 these two chains, and close analogues of them, give a nearly complete description of $H_1(\mathfrak{T}^1)$.

In Section 3 we consider these PE invariants in the context of rotational symmetry. Whilst for a tiling of finite local complexity the action of rotation on the PE homology and cohomology agree via Poincaré duality (Proposition 3.11), the actions at the (co)chain level behave differently. We consider *ePE* chains and cochains, which are required to have the same coefficients at any two cells whenever those cells agree on patches of sufficiently large radius up to orientation-preserving Euclidean motion (rather than just translations as in the case of the PE homology groups). We show in Theorem 3.3 that over divisible coefficients G we still have Poincaré duality $H^\bullet(\mathfrak{T}^0; G) \cong H_{d-\bullet}(\mathfrak{T}^0; G)$ between the ePE cohomology and ePE homology, but over $\mathbb{Z}$ coefficients this typically fails. For example, for the Penrose kite and dart tilings the degree zero ePE homology group has a copy of an order five cyclic subgroup not present in the corresponding ePE cohomology group in degree two. A degree zero ePE torsion element is depicted in Figure 3.

So whilst the PE homology gives a curious alternative way of visualising PE invariants, the ePE homology provides a new invariant to the ePE cohomology, or the Čech cohomology of the associated space $\Omega^0_{\mathfrak{T}}$ (defined by Barge, Diamond, Hunton and Sadun [5], or see Section 3.1). We show in [42] how this new invariant may naturally be incorporated into a spectral sequence converging to the Čech cohomology of the "Euclidean hull" $\Omega^{\mathrm{rot}}_{\mathfrak{T}}$ (see Section 3.1) of a two-dimensional tiling. The only potentially nontrivial map of this spectral sequence has a very simple description in terms of the local combinatorics of the tiling. This procedure dovetails conveniently with the methods that we shall introduce in Section 4 to efficiently compute the Čech cohomology of Euclidean hulls of hierarchical tilings, leading to some new computations on the cohomologies of these spaces.

We show how the ePE homology, ePE cohomology and rotationally invariant part of the PE cohomology are related for a two-dimensional tiling in Theorem 3.13 and give the corresponding calculations for the Penrose kite and dart tilings. In general, over rational coefficients all three are canonically isomorphic, but over integral coefficients the canonical map from the ePE cohomology to the rotationally invariant part of the PE cohomology is rarely an isomorphism. It turns out that this map naturally factorises through the ePE homology.

The techniques that we present are not limited to tilings of Euclidean space. In Section 3.5 we introduce the notion of a *system of internal symmetries*, which neatly encodes the necessary data required to define PE cohomology and various other related constructions. This allows us, for example, to apply the same techniques to non-Euclidean tilings, such as the combinatorial pentagonal tilings of Bowers and Stephenson [11].

In Section 4 we change tack by considering the problem of how to actually compute the PE homology for certain examples. The PE homology formalism naturally leads to a simple and efficient method of computation for invariants of a hierarchical tiling which is closely related to that of Barge, Diamond, Hunton and Sadun [5]. The descriptions of the PE and ePE homology groups that appear in this paper for the Penrose kite and dart tilings are made possible through this method of calculation. The method is directly applicable to a broad range of tilings, including "mixed substitution tilings" (see Gähler and Maloney [18]) but also non-Euclidean examples, such as the pentagonal tilings of Bowers and Stephenson mentioned above. The "approximant homology groups" of the computation and the "connecting maps" between them have a direct description in terms of the combinatorics of the star patches, making it highly amenable to computer implementation. Gonçalves [19] used the duals of these approximant chain complexes for a computation of the K–theory of the C^*–algebra of the stable equivalence relation of a substitution tiling. Our method of computation of the PE homology groups seems to confirm the observation there of a certain duality between these K–groups and the K–theory of the tiling space.

Organisation of paper

In Section 1 we shall recall how one may associate to a Euclidean tiling $\mathfrak{T}$ its translational hull $\Omega^1_{\mathfrak{T}}$. When $\mathfrak{T}$ has FLC, we also describe the presentation of $\Omega^1_{\mathfrak{T}}$ as an inverse limit of approximants. In Section 2 we recall the PE cohomology of an FLC tiling $\mathfrak{T}$, and how it may be identified with the Čech cohomology $\check{H}^\bullet(\Omega^1_{\mathfrak{T}})$ of the tiling space. We then introduce the PE homology of an FLC tiling and establish PE Poincaré duality between the PE cohomology and PE homology.

In Section 3 we consider PE homology in the context of rotational symmetry. The ePE (co)homology groups are defined in Section 3.2, where we show, in Theorem 3.3, that the ePE cohomology and ePE homology are Poincaré dual when taken over suitably divisible coefficients. In Section 3.3 we show how Poincaré duality for the ePE homology of a two-dimensional tiling is restored for $\mathbb{Z}$ coefficients by a simple modification of the ePE homology. The action of rotation on the PE cohomology of an FLC tiling, and its interaction with the ePE homology, is considered in Section 3.4. In Section 3.5 we demonstrate how the techniques of Section 3 may be naturally extended to a more general framework.

In Section 4 we develop a method of computation of the PE homology for polytopal substitution tilings, close in spirit to the BDHS approach [5]. In Section 4.4 we explain how the method is modified to compute the ePE homology, and how it may be applied to more general settings, such as mixed substitution systems or to non-Euclidean examples.

Acknowledgements I thank John Hunton, Alex Clark, Lorenzo Sadun and Dan Rust for numerous helpful discussions. I particularly thank the enormous efforts of the anonymous referee, whose suggestions have greatly improved the final version of this article. This research was supported by EPSRC.

1 Tilings and tiling spaces

1.1 Cellular, polytopal and dual tilings

Recall that a CW complex is called *regular* if the attaching maps of its cells may be taken to be homeomorphisms. A *cellular tiling* of $\mathbb{R}^d$ shall be defined to be a pair

$\mathfrak{T} = (\mathcal{T}, l)$ of a regular CW decomposition $\mathcal{T}$ of $\mathbb{R}^d$ along with a *labelling* l of $\mathcal{T}$, by which we mean a map from the cells of $\mathcal{T}$ to some set of "labels" L. We shall take *cell* to mean a closed cell. If the cells are convex polytopes then we call $\mathfrak{T}$ a *polytopal tiling*. For brevity, we will often refer to a cellular tiling as simply a *tiling*, and a d–cell of a tiling as a *tile*. A *patch* of $\mathfrak{T}$ is a finite subcomplex $\mathcal{P}$ of $\mathcal{T}$ together with the labelling restricted to $\mathcal{P}$. For a bounded set $U \subseteq \mathbb{R}^d$, we let $\mathfrak{T}[U]$ be the patch supported on the set of tiles t for which $t \cap U \neq \varnothing$.

Homeomorphisms of $\mathbb{R}^d$ act on tilings and patches in the obvious way. Two patches are called *translation equivalent* if one is a translate of the other. The *diameter* of a patch is defined to be the diameter of the support of its tiles. A tiling or, more generally, a collection of tilings, is said to have (*translational*) *finite local complexity* (*FLC*) if, for any $r > 0$, there are only finitely many patches of diameter at most r up to translation equivalence. It is not difficult to see that a cellular tiling has FLC if and only if there are only finitely many translation classes of cells and the labelling function takes on only finitely many distinct values.

One may wish to consider other forms of decoration of $\mathbb{R}^d$, such as Delone sets, or tilings with overlapping or fractal tiles, and many of the concepts that we describe here have obvious analogues for them. However, when such a pattern has FLC it is always essentially equivalent to a polytopal tiling. In more detail, we say that $\mathfrak{T}'$ *is locally derivable from* $\mathfrak{T}$ if there exists some $r > 0$ for which $(S_1\mathfrak{T}')[B_0] = (S_2\mathfrak{T}')[B_0]$ whenever $(S_1\mathfrak{T})[B_r] = (S_2\mathfrak{T})[B_r]$, where the S_i are translations $x \mapsto x + t_i$ and B_r is the closed ball of radius r centred at the origin. The tilings $\mathfrak{T}$ and $\mathfrak{T}'$ are said to be *mutually locally derivable* (*MLD*) if each is locally derivable from the other. Loosely, this means that $\mathfrak{T}$ and $\mathfrak{T}'$ only differ in a very cosmetic sense, via locally defined redecoration rules. This concept was introduced in [4], along with the finer relation of *S-MLD* equivalence, which takes into account general Euclidean isometries rather than just translations by replacing the translations S_i in the definition of a local derivation above with Euclidean motions. FLC patterns (or even eFLC ones; see Section 3.1) are always S-MLD to polytopal tilings via a Voronoi construction.

In the following sections we shall usually take our tilings to be polytopal. Since the properties of tilings of interest to us are preserved under S-MLD equivalence, this is not a harsh restriction. The major motivation for this choice is that some useful constructions may be defined for a polytopal complex, namely the barycentric subdivision and dual complex. In fact, it is sufficient for these constructions to use regular CW complexes as our starting point, but the most efficient way of dealing with this more general case is to pass to a combinatorial setting, which we do not cover in full detail here although shall outline in Section 3.5.

For a polytopal tiling $\mathfrak{T} = (\mathcal{T}, l)$ we may construct the barycentric subdivision $\mathcal{T}_\Delta$ of its underlying CW decomposition geometrically, as follows. For each cell $c \in \mathcal{T}$, define $b(c)$, its *barycentre*, to be the centre of mass of c in its supporting hyperplane. We write $c_1 \preceq c_2$ for closed cells c_1, c_2 of $\mathcal{T}$ to mean that $c_1 \subseteq c_2$, and $c_1 \prec c_2$ if this inclusion is strict. The k–skeleton $\mathcal{T}_\Delta^k$ for $k = 0, \dots, d$ is defined by taking as k–cells those simplices which are the convex hulls of the vertices $\{b(c_0), b(c_1), \dots, b(c_k)\}$ for a chain of cells $c_0 \prec c_1 \prec \cdots \prec c_k$ of $\mathcal{T}$ of length $k+1$. Such a cell may be labelled by the sequence of labels $(l(c_0), l(c_1), \dots, l(c_k))$; we define the tiling $\mathfrak{T}_\Delta = (\mathcal{T}_\Delta, l_\Delta)$, where l_Δ is the labelling of the cells of $\mathcal{T}_\Delta$ defined in this way. Assuming (without loss of generality, up to S-MLD equivalence) that cells of different dimension have different labels, it is easily verified that $\mathfrak{T}$ and $\mathfrak{T}_\Delta$ are S-MLD.

We may reconstruct $\mathcal{T}$ from its barycentric subdivision $\mathcal{T}_\Delta$ by identifying an open k–cell c of $\mathcal{T}$ with the conglomeration of open simplicial cells corresponding to chains $c_0 \prec \cdots \prec c_j \prec c$ terminating in c. Flipping this process on its head, we obtain the *dual complex* $\widehat{\mathcal{T}}$. That is, we define an open k–cell of $\widehat{\mathcal{T}}$ as the union of open simplicial cells corresponding to chains $c \prec c_0 \prec \cdots \prec c_j$ emanating from a $(d-k)$–cell c. Similarly to $\mathcal{T}_\Delta$, we may easily label $\widehat{\mathcal{T}}$ so as to define a *dual tiling* of $\mathfrak{T}$ which is S-MLD to $\mathfrak{T}_\Delta$, and hence also S-MLD to the original tiling $\mathfrak{T}$. The dual tiling $\widehat{\mathfrak{T}}$ typically won't have convex polytopal cells, but it is cellular, owing to the piecewise linearity of the polytopal decomposition $\mathcal{T}$. The k–cells of $\mathfrak{T}$ are naturally in bijection with the $(d-k)$–cells of $\widehat{\mathfrak{T}}$, and we have that $a \preceq b$ for cells of $\mathfrak{T}$ if and only if $\widehat{a} \succeq \widehat{b}$ for the corresponding dual cells of $\widehat{\mathfrak{T}}$. A similar process would have worked for $\mathcal{T}$ only regular cellular. However, the decomposition of $\mathbb{R}^d$ defined by $\widehat{\mathcal{T}}$ need not be cellular even for nonpiecewise linear *simplicial* complexes $\mathcal{T}$. Even so, the resulting dual decomposition $\widehat{\mathcal{T}}$ still retains the analogous homological properties to a CW complex needed to define cellular homology (see [31, Section 8.64]) and so the constructions and arguments to follow can, with little extra effort, be extended to this case.

1.2 Tiling spaces

To a tiling $\mathfrak{T}$ one may associate a moduli space $\Omega^1_{\mathfrak{T}}$ which, as a set, consists of tilings "locally indistinguishable" from $\mathfrak{T}$. Let S be a set of tilings of $\mathbb{R}^d$. We wish to endow S with a geometry which expresses the intuitive idea that two tilings are close if, up to a small perturbation, the two agree to a large radius about the origin.

An approach which neatly applies to a large class of tilings, and captures this idea most directly, proceeds as follows. Let $H(\mathbb{R}^d)$ be the space of homeomorphisms of $\mathbb{R}^d$; these shall serve as our perturbations. Equipped with the compact–open topology, we may consider a neighbourhood $V \subseteq H(\mathbb{R}^d)$ of the identity as "small" if its elements only perturb points within a large distance from the origin of $\mathbb{R}^d$ a small amount. In this case, for $f \in V$ it is intuitive to think of $f\mathfrak{T}$ as a small perturbation of $\mathfrak{T}$. In fact, $\mathfrak{T}$ should still be "close" to any other tiling $\mathfrak{T}'$ so long as $\mathfrak{T}[K]$ and $(f\mathfrak{T}')[K]$ agree for some $K \subseteq \mathbb{R}^d$ containing a large neighbourhood of the origin. For a neighbourhood $V \subseteq H(\mathbb{R}^d)$ of $\mathrm{id}_{\mathbb{R}^d}$ and bounded $K \subseteq \mathbb{R}^d$, we define

$$U(K,V) := \{(\mathfrak{T}_1,\mathfrak{T}_2) \in S \times S \mid \mathfrak{T}_1[K] = (f\mathfrak{T}_2)[K]\}.$$

It is not difficult to verify that the resulting collection of sets $U(K,V)$ is a base for a uniformity on S. If the reader is unfamiliar with uniformities, the only important point here is that we have a uniform notion of tilings being "close": $\mathfrak{T}_1$ *is considered close to* $\mathfrak{T}_2$*, as judged by* K *and* V*, whenever* $(\mathfrak{T}_1,\mathfrak{T}_2) \in U(K,V)$. With K a large neighbourhood of the origin and V a set of homeomorphisms moving points only a very small amount in the vicinity of K, we recover our intuitive notion of $\mathfrak{T}_1$ and $\mathfrak{T}_2$ being close, when they agree to a large radius up to a small perturbation. The above construction easily generalises to other decorations of $\mathbb{R}^d$, such as Delone sets, and also tilings with infinite label sets which are equipped with a metric (see [34]).

For a tiling $\mathfrak{T}$ we define the *translational hull* or *tiling space* as

$$\Omega^1_{\mathfrak{T}} := \overline{\{\mathfrak{T} + x \mid x \in \mathbb{R}^d\}},$$

where the completion is taken with respect to the uniformity defined above. In the case that $\mathfrak{T}$ has FLC, two patches agree up to a small perturbation when they agree up to a small translation. So in this case the sets

$$U(K,\epsilon) := \{(\mathfrak{T}_1,\mathfrak{T}_2) \mid \mathfrak{T}_1[K] = (\mathfrak{T}_2 + x)[K] \text{ for } \|x\| \leq \epsilon\}$$

serve as a base for our uniformity, where the $K \subseteq \mathbb{R}^d$ are bounded and $\epsilon > 0$. Loosely, two tilings are "close" if and only if their central patches agree to a large radius (parametrised by K) up to a small translation (parametrised by ϵ). It is not difficult to show that the tiling space $\Omega^1_{\mathfrak{T}}$ is a compact space whose points may be identified with those tilings whose patches are translates of patches of $\mathfrak{T}$. So one may take as basic open neighbourhoods of a tiling $\mathfrak{T}_1$ in $\Omega^1_{\mathfrak{T}}$ the *cylinder sets*

$$C(R,\epsilon,\mathfrak{T}_1) := \{\mathfrak{T}_2 \in \Omega^1_{\mathfrak{T}} \mid \mathfrak{T}_1[B_R] = (\mathfrak{T}_2 + x)[B_R] \text{ for } \|x\| \leq \epsilon\}$$

of tilings of the hull, which, up to a translation of at most ϵ, agree with $\mathfrak{T}_1$ to radius R.

1.3 Inverse limit presentations

Another simplification granted to us by finite local complexity is that the tiling space $\Omega^1_{\mathfrak{T}}$ may be presented as an inverse limit of CW complexes Γ^1_i, following Gähler's (unpublished) construction; see [38] for details, and also the alternative approach of Barge, Diamond, Hunton and Sadun [5]. Inductively define the *i–corona* of a tile as follows: the 0–corona of a tile t is the patch whose single tile is t; for $i \in \mathbb{N}$, the i–corona is the patch of tiles which have nonempty intersection with the $(i-1)$–corona. That is, one constructs the i–corona of t by taking t and then iteratively appending neighbouring tiles i times. For $x, y \in \mathbb{R}^d$, write $x \sim_i y$ to mean that there are two tiles t_x and t_y of $\mathfrak{T}$ containing x and y, respectively, for which the i–corona of t_x is equal to the i–corona of t_y, up to a translation taking x to y. This is typically not an equivalence relation, so we define the *approximant* Γ_i^1 to be the quotient of $\mathbb{R}^d$ by the transitive closure of the relation $\sim_i$. More intuitively, we form Γ_i^1 by taking a copy of the central tile from each translation class of i–corona, glueing them along their boundaries according to how they can meet in the tiling. We define the *i–corona* of a lower-dimensional cell to be the intersection of i–coronas of the tiles which it is contained in. An alternative way of defining approximants, which avoids taking a transitive closure (although identifies more points of $\mathbb{R}^d$ at each level), is to identify cells of the tiling which share the same i–coronas, up to a translation. Each approximant naturally inherits a cellular decomposition from that of the tiling.

For $i \leq j$, cells of $\mathfrak{T}$ identified in Γ_j^1 are also identified in Γ_i^1, so we have "forgetful" cellular quotient maps $\pi_{i,j}\colon \Gamma_j^1 \to \Gamma_i^1$. The inverse limit of this projective system

$$\varprojlim(\Gamma_i^1, \pi_{i,j}) := \left\{ (x_i)_{i\in\mathbb{N}_0} \in \prod_{i=0}^{\infty} \Gamma_i^1 \;\middle|\; \pi_{i,j}(x_j) = x_i \right\}$$

is homeomorphic to the tiling space $\Omega^1_{\mathfrak{T}}$. The central idea here is that a point of Γ_i^1 describes how to tile a neighbourhood of the origin, where the sizes of these neighbourhoods increase with i. An element of the inverse limit space then corresponds to a consistent sequence of choices of larger and larger patches about the origin, so it defines a tiling. Any two tilings which are "close" correspond to points of the inverse limit which are "close" on an approximant Γ_i^1 for large i, and vice versa.

2 Translational pattern-equivariance

2.1 Identifying Čech with PE cohomology

Locally, the tiling space of an FLC tiling has a product structure of cylinder sets $U \times C$, where U is an open subset of $\mathbb{R}^d$, corresponding to small translations, and C is a totally disconnected space, corresponding to different ways of completing a finite patch to a full tiling. Globally, $\Omega^1_{\mathfrak{T}}$ is a torus bundle with totally disconnected fibre [39]. Many classical invariants — homotopy groups and singular (co)homology groups, for example — are ill-suited to studying $\Omega^1_{\mathfrak{T}}$ when $\mathfrak{T}$ is nonperiodic, in which case this space is not locally connected. A commonly employed topological invariant with which to study tiling spaces is Čech cohomology $\check{H}^\bullet(-)$. We shall not cover its definition here (see [10, Chapter 2.10]), although we recall two important features of it:

(1) Čech cohomology is naturally isomorphic to singular cohomology on the category of spaces homotopy equivalent to CW complexes and continuous maps.

(2) For a projective system $(\Gamma_i, \pi_{i,j})$ of compact, Hausdorff spaces Γ_i, we have an isomorphism $\check{H}^\bullet(\varprojlim(\Gamma_i, \pi_{i,j})) \cong \varinjlim(\check{H}^\bullet(\Gamma_i), \pi^*_{i,j})$.

Pattern-equivariant cohomology is a tool designed to give intuitive descriptions of the Čech cohomology of tiling spaces. It was first defined by Kellendonk and Putnam [28] (see also [24]), where they showed that it is isomorphic to the Čech cohomology of the tiling space taken over $\mathbb{R}$ coefficients. It is constructed by restricting the de Rham cochain complex of $\mathbb{R}^d$ of smooth forms to a subcochain complex of forms which, loosely, are determined pointwise by the local decoration of the underlying tiling to some bounded radius.

A second approach, introduced by Sadun [37], is to use cellular cochains, and has the advantage of generalising to arbitrary abelian coefficients. Let $\mathfrak{T} = (\mathcal{T}, l)$ be a cellular tiling (recall that $\mathcal{T}$ is the underlying cell complex of $\mathfrak{T}$). Denote by $C^\bullet(\mathcal{T})$ the cellular cochain complex of $\mathcal{T}$,

$$C^\bullet(\mathcal{T}) := 0 \to C^0(\mathcal{T}) \xrightarrow{\delta^0} C^1(\mathcal{T}) \xrightarrow{\delta^1} \cdots \xrightarrow{\delta^{d-1}} C^d(\mathcal{T}) \to 0,$$

where each $C^k(\mathcal{T})$ is the group of cellular k–cochains and δ^k is the degree k cellular coboundary map. A cellular k–cochain ψ is a function which assigns to each orientation ω_c of k–cell c an integer, satisfying $\psi(\omega_c^+) = -\psi(\omega_c^-)$ for opposite orientations ω_c^+ and ω_c^- of a cell c. Of course, choosing an orientation for each k–cell induces an isomorphism $C^k(\mathcal{T}) \cong \prod_{k\text{–cells}} \mathbb{Z}$. Choose orientations for the k–cells so that $\omega_c + x = \omega_{c+x}$ whenever c and $c+x$ are cells of $\mathcal{T}$, where $\omega_c + x$ is the orientation on $c+x$ induced from ω_c by translation. Write $\psi(c) := \psi(\omega_c)$, where ω_c is the chosen orientation of c. A cochain ψ is called *pattern-equivariant* (*PE*) if there exists some $i \in \mathbb{N}_0$ for which $\psi(c_1) = \psi(c_2)$ whenever c_1 and c_2 have identical i–coronas, up to a translation taking c_1 to c_2.

It is easy to check that the coboundary of a PE cochain is PE. Define $C^\bullet(\mathfrak{T}^1)$ to be the subcochain complex of $C^\bullet(\mathcal{T})$ consisting of PE cochains. Its cohomology $H^\bullet(\mathfrak{T}^1)$ is called the *pattern-equivariant cohomology of* $\mathfrak{T}$.

A cellular cochain $\psi \in C^k(\mathcal{T})$ is PE if and only if it is a pullback cochain from some approximant, that is, if $\psi = \pi_i^*(\widetilde{\psi})$, where $\widetilde{\psi} \in C^k(\Gamma_i^1)$ is a cellular cochain on an approximant and π_i is the (cellular) quotient map $\pi_i\colon \mathbb{R}^d \to \Gamma_i^1$ defining Γ_i^1. This fact, in combination with the description of the tiling space $\Omega_{\mathfrak{T}}^1$ as an inverse limit of Gähler complexes and the two features of Čech cohomology given above, leads to the proof of the following:

Theorem 2.1 [37] *The PE cohomology $H^\bullet(\mathfrak{T}^1)$ of an FLC tiling $\mathfrak{T}$ is isomorphic to the Čech cohomology $\check{H}^\bullet(\Omega_{\mathfrak{T}}^1)$ of its tiling space.*

2.2 PE homology and Poincaré duality

We shall now define the PE *homology* groups of a cellular tiling $\mathfrak{T}$. The construction runs almost identically to the construction of the cellular PE cohomology groups above, but where we took cellular coboundary maps before we shall take instead cellular *boundary* maps. In more detail, let $C_\bullet^{\mathrm{BM}}(\mathcal{T})$ denote the *cellular Borel–Moore chain complex*,

$$C_\bullet^{\mathrm{BM}}(\mathcal{T}) := 0 \leftarrow C_0^{\mathrm{BM}}(\mathcal{T}) \xleftarrow{\partial_1} C_1^{\mathrm{BM}}(\mathcal{T}) \xleftarrow{\partial_2} \cdots \xleftarrow{\partial_d} C_d^{\mathrm{BM}}(\mathcal{T}) \leftarrow 0.$$

The chain groups $C_k^{\mathrm{BM}}(\mathcal{T})$ are canonically isomorphic to the cochain groups $C^k(\mathcal{T})$. That is, up to a choice of orientations for the k–cells, a cellular Borel–Moore chain $\sigma \in C_k^{\mathrm{BM}}(\mathcal{T})$ is given by a choice of integer for each k–cell. But we think of its elements as possibly infinite, or noncompactly supported cellular *chains*. The boundary maps ∂_k are the linear extension to these chain groups of the cellular boundary maps of the standard cellular chain complex of $\mathcal{T}$.

Pattern-equivariance of a chain $\sigma \in C_k^{\mathrm{BM}}(\mathcal{T})$ is defined identically to that of a cochain. That is, σ is PE if there exists some $i \in \mathbb{N}_0$ for which, for any two k–cells c_1, c_2 of $\mathcal{T}$ with identical i–coronas in $\mathfrak{T}$ up to a translation, c_1 and c_2 have the same coefficient in σ. It is easy to see that if σ is PE then $\partial(\sigma)$ is also PE. Restricting to PE cellular Borel–Moore chains, we obtain a subchain complex $C_\bullet(\mathfrak{T}^1)$ of $C_\bullet^{\mathrm{BM}}(\mathcal{T})$ whose homology $H_\bullet(\mathfrak{T}^1)$ we shall call the *pattern-equivariant homology of* $\mathfrak{T}$. So the elements of the PE homology are represented by, typically, noncompactly supported cellular *cycles* (chains with trivial boundary), where two PE cycles σ_1 and σ_2 are homologous if $\sigma_1 = \sigma_2 + \partial(\tau)$ for some PE chain τ.

These homology groups certainly have a highly geometric definition, but what do they measure? Through a Poincaré duality argument, we may in fact identify them with the (reindexed) PE cohomology groups and thus, in light of Theorem 2.1, with the Čech cohomology groups of the tiling space:

Theorem 2.2 *For a polytopal tiling $\mathfrak{T}$ of $\mathbb{R}^d$ of finite local complexity, we have PE Poincaré duality $H^\bullet(\mathfrak{T}^1) \cong H_{d-\bullet}(\mathfrak{T}^1)$.*

Proof Classical Poincaré duality provides an isomorphism of complexes

$$-\cap\Gamma\colon C^\bullet(\mathcal{T}) \xrightarrow{\cong} C^{\mathrm{BM}}_{d-\bullet}(\widehat{\mathcal{T}}),$$

via the cap product with a cellular Borel–Moore fundamental class Γ, a d–cycle for which each oriented d–cell has coefficient either $+1$ or -1, pairing orientations of k–cells with orientations of their dual $(d-k)$–cells. Here, $\mathcal{T}$ is the underlying cell complex of $\mathfrak{T}$ and $\widehat{\mathcal{T}}$ is its dual complex. By definition, a (co)chain is PE whenever it assigns coefficients to cells in a way which only depends locally on the tiling to some bounded neighbourhood of that cell. The fundamental class Γ is also PE. Since the cap product (and here, its inverse) is defined in a local manner, and $\mathfrak{T}$ and $\widehat{\mathfrak{T}}$ are MLD, a k–cochain ψ of $\mathfrak{T}$ is PE if and only if its dual $(d-k)$-chain $\psi\cap\Gamma$ of $\widehat{\mathfrak{T}}$ is PE. So $-\cap\Gamma$ restricts to an isomorphism between PE complexes,

$$-\cap\Gamma\colon C^\bullet(\mathfrak{T}^1) \xrightarrow{\cong} C_{d-\bullet}(\widehat{\mathfrak{T}}^1).$$

The barycentric tiling $\mathfrak{T}_\Delta$ refines both $\mathfrak{T}$ and $\widehat{\mathfrak{T}}$, and is MLD to both. As one may expect, taking such a refinement does not effect PE (co)homology. This shall be shown, in a more general setting, in Lemma 3.2. Precisely, we have quasi-isomorphisms $\iota\colon C_\bullet(\mathfrak{T}^1) \to C_\bullet(\mathfrak{T}^1_\Delta)$ and $\hat\iota\colon C_\bullet(\widehat{\mathfrak{T}}^1) \to C_\bullet(\mathfrak{T}^1_\Delta)$; recall that a quasi-isomorphism is a (co)chain map which induces an isomorphism on (co)homology. In summation we have the diagram of quasi-isomorphisms

$$C^\bullet(\mathfrak{T}^1) \xrightarrow{-\cap\Gamma} C_{d-\bullet}(\widehat{\mathfrak{T}}^1) \xrightarrow{\hat\iota} C_{d-\bullet}(\mathfrak{T}^1_\Delta) \xleftarrow{\iota} C_{d-\bullet}(\mathfrak{T}^1),$$

from which PE Poincaré duality $H^\bullet(\mathfrak{T}^1) \cong H_{d-\bullet}(\mathfrak{T}^1)$ follows. □

We note, in passing, that PE homology is named according to its geometric construction: via PE *chains*. However, we shall not attempt to make it functorial with respect to any sort of class of structure preserving maps between tilings. In fact, we do not expect for there to be a natural way of achieving this: with respect to particular kinds of continuous maps between tiling spaces, pattern-equivariance is a property which pulls back rather than pushes forward.

Example 2.3 Let $\mathfrak{T}$ be a Penrose kite and dart tiling. The Čech cohomology of the translational hull of the Penrose tilings was first calculated in [1] (although see also the earlier closely related K–theoretic calculations of Kellendonk in the groupoid setting [23]). In Section 4 we provide a different way of computing these groups which, as a direct by-product, provides us with explicit descriptions of the generators in terms of PE chains. Consistently with previous calculations, we find that

$$H_{2-k}(\mathfrak{T}^1) \cong H^k(\mathfrak{T}^1) \cong \check{H}^k(\Omega^1_{\mathfrak{T}}) \cong \begin{cases} \mathbb{Z} & \text{for } k=0, \\ \mathbb{Z}^5 & \text{for } k=1, \\ \mathbb{Z}^8 & \text{for } k=2. \end{cases}$$

Let P_c be a pair of a patch P from $\mathfrak{T}$ along with a choice of oriented k–cell c from P, taken up to translation equivalence. We have an associated PE *indicator k–chain* $\mathbb{1}(P_c) \in C_k(\mathfrak{T}^1)$ for which each k–cell of $\mathfrak{T}$ has coefficient one when it is contained in an ambient patch for which the pair agrees with P_c up to translation, and all other cells have coefficient zero. The degree zero PE homology group $H_0(\mathfrak{T}^1)$ for a Penrose kite and dart tiling may be freely generated by indicator 0–chains $\mathbb{1}(P_v) \in C_0(\mathfrak{T}^1)$, where each P_v is one of the vertex-stars of $\mathfrak{T}$, paired with its central vertex. The full list of possible translation classes of such patches, up to rotation by some $\frac{2\pi k}{10}$, are given (and named, according to Conway's notation) in Figure 4. As an example of a choice of elements freely generating $H_0(\mathfrak{T}^1)$, we may choose two "queen" vertices, one oriented as in Figure 4 and the other a $\frac{2\pi}{10}$ rotate of it, and six "king" vertices, each a $\frac{2\pi k}{10}$ rotate of that of Figure 4 with $k = 0, \dots, 5$.

We shall go into more detail on generators for $H_0(\mathfrak{T}^1)$ in Example 3.15. In degree one, there are two particularly beautiful cycles that we wish to discuss here. There is a PE 1–cycle ρ' given by running along the bottoms of the dart tilings, with 1–cells oriented to point to, say, the right when the darts are oriented to point upwards. The resulting cycle is illustrated in red in Figure 2, where we have removed cell orientations and the 1–skeleton of the tiling to decrease clutter. The extra embellishments of the figure shall be discussed in Section 4; there is a green 1–cycle for the analogous chain of the supertiling of $\mathfrak{T}$, along with a blue PE 2–chain whose boundary relates the two. As one can see, ρ' forms a disjoint union of clockwise and anticlockwise running loops. Interestingly, deducing which of these two options is the case at some cell of a loop requires consideration of arbitrarily large patches; in fact, for specific kite and dart tilings there exists a single infinite fractal-like path along dart tiles. But ρ' is not a generator, there exists another PE 1–cycle ρ for which $[\rho'] = 2[\rho]$ in $H_1(\mathfrak{T}^1)$. The loops of Figure 2 come in two types: ones where the darts along the loops are $\frac{2\pi k}{10}$ rotates of an upwards pointing dart tile with k odd, and ones where the darts are even rotates. The 1–cycle ρ is given by restricting ρ' to those loops in one of these two parities; both choices are homologous and equal to $\frac{1}{2}[\rho']$ in $H_1(\mathfrak{T}^1)$.

A second generator ν_0 for $H_1(\mathfrak{T}^1)$ is depicted in Figure 1. The cycle arranges as a union of infinite paths along the 1–skeleton which closely approximate the *Ammann bars* [20, Section 10.6] of the supertiling of $\mathfrak{T}$, illustrated in the figure in green. There are ten further chains ν_k defined by $\frac{2\pi k}{10}$ rotates of ν_0 (see Section 3.4). We calculate that $H_1(\mathfrak{T}^1)$ is freely generated by the homology classes of ρ, ν_0, ν_1, ν_2 and ν_3; every other PE 1–cycle is equal, up to a PE 2–boundary, to a linear combination of

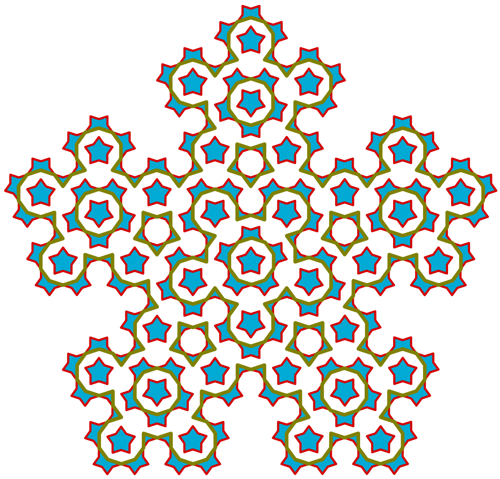

Figure 2: The PE 1–cycle ρ' of a Penrose kite and dart tiling (red), with the analogous chain of the supertiling (green) and a PE 2–chain (blue) whose boundary relates the two.

these cycles. It turns out that $\nu_4 \simeq -\nu_0 + \nu_1 - \nu_2 + \nu_3$ in homology. This formula is unsurprising following the observation that one may associate each ν_k with a direction given by a tenth root of unity, and we have the identity $\sum_{k=0}^{4}(-1)^k \exp\bigl(\frac{2\pi i \cdot k}{10}\bigr) = 0$.

3 PE homology and rotations

In the previous section we showed that topological invariants of tiling spaces may be described in a highly geometric way, using infinite cellular chains on the tiling. However, PE homology is essentially just offering a different perspective on the generators of the PE cohomology here. As we shall see in Section 4, PE homology does provide a valuable alternative insight into the calculation of these invariants for hierarchical tilings. In this section, we shall show that PE homology provides a new invariant to the PE cohomology when one considers these invariants in the context of rotational symmetries.

3.1 Rotational tiling spaces

Let $\mathrm{SE}(d) \cong \mathbb{R}^d \rtimes \mathrm{SO}(d)$ denote the transformation group of orientation-preserving isometries of $\mathbb{R}^d$; elements of $\mathrm{SE}(d)$ shall be called *rigid motions*. There are two topological spaces naturally associated to a tiling $\mathfrak{T}$ of $\mathbb{R}^d$ which incorporate the action of $\mathrm{SE}(d)$ on $\mathfrak{T}$. The first, defined analogously to the translational hull $\Omega^1_{\mathfrak{T}}$, is the *Euclidean hull*

$$\Omega^{\mathrm{rot}}_{\mathfrak{T}} := \overline{\{f\mathfrak{T} \mid f \in \mathrm{SE}(d)\}}.$$

It follows directly from the definitions that the special Euclidean group $\mathrm{SE}(d)$ acts uniformly on the Euclidean orbit of $\mathfrak{T}$, and so this action extends to the entire Euclidean hull. In particular, the subgroup $\mathrm{SO}(d) \leq \mathrm{SE}(d)$ of rotations at the origin acts on $\Omega^{\mathrm{rot}}_{\mathfrak{T}}$. The second space, the one which we shall concentrate on in this section, is the quotient

$$\Omega^0_{\mathfrak{T}} := \Omega^{\mathrm{rot}}_{\mathfrak{T}} / \mathrm{SO}(d).$$

We shall say that $\mathfrak{T}$ has *Euclidean finite local complexity* (*eFLC*, for short) if, for every $r > 0$, there exist only finitely many patches of diameter at most r up to rigid motion. Interesting examples of tilings which have eFLC, but not translational FLC, are the Conway–Radin pinwheel tilings of $\mathbb{R}^2$, whose tiles are all rigid motions of a $(1, 2, \sqrt{5})$ triangle, or its reflection, but are found in the tiling pointing in infinitely many directions. Much of what can be said for FLC tilings and their translational hulls has an analogue for eFLC tilings and their Euclidean hulls. In particular, for an eFLC tiling $\mathfrak{T}$, its Euclidean hull $\Omega^{\mathrm{rot}}_{\mathfrak{T}}$ is a compact space whose points may be identified with those tilings whose patches are rigid motions of the patches of $\mathfrak{T}$. The space $\Omega^0_{\mathfrak{T}}$ is then the quotient of $\Omega^{\mathrm{rot}}_{\mathfrak{T}}$ given by identifying tilings which differ by a rotation at the origin. One may define inverse limit presentations of these spaces in a similar way to the construction of the Gähler complexes, which is tantamount to being able to define pattern-equivariant cohomology.

To explain this further, we now focus on the space $\Omega^0_{\mathfrak{T}}$. For $i \in \mathbb{N}_0$ we define CW complexes Γ^0_i analogously to the complexes of the translational setting, replacing translations with rigid motions. For example, we may define the complexes Γ^0_i by identifying cells c_1, c_2 of $\mathfrak{T}$ via rigid motions which take c_1 to c_2, and the i–corona of c_1 to the i–corona of c_2. The CW complexes Γ^0_i, along with the "forgetful maps" between them, define a projective system whose inverse limit is homeomorphic to $\Omega^0_{\mathfrak{T}}$.

It may be the case that cells of $\mathfrak{T}$ have *nontrivial isotropy*, that is, there may be cells c whose i–coronas are preserved by rigid motions mapping c to itself nontrivially, which will cause points of c to be identified in the quotient spaces Γ^0_i. Given a cell $c \in \mathcal{T}$, the rigid motions mapping c to itself and preserving its i–corona is a group, which we call the *isotropy*, and denote by $\mathrm{Iso}(c,i)$. Write $\widetilde{\mathrm{Iso}}(c,i)$, the *cell isotropy*, for the group of transformations of $\mathrm{Iso}(c,i)$ restricted to c.

The cell isotropy groups of the barycentric subdivision $\mathfrak{T}_\Delta$ are always trivial. Indeed, a barycentric cell is determined by its vertices, which are determined by a chain of incidences $c_0 \prec \cdots \prec c_k$ in $\mathcal{T}$, and a rigid motion taking a barycentric cell to itself is determined by the map restricted to these vertices. A nontrivial map on such a vertex set would have to correspond to a rigid motion taking $b(c_i)$ to some $b(c_j)$ with $i \neq j$, which cannot be the case since $c_i \neq c_j$ have distinct dimensions, and thus, by assumption, distinct labels.

3.2 Euclidean pattern-equivariance

A cellular cochain $\psi \in C^k(\mathcal{T})$ shall be called *Euclidean pattern-equivariant* (*ePE*) if there exists some $i \in \mathbb{N}_0$ for which $\psi(\omega_c) = \psi(f_*(\omega_c))$ whenever f is a rigid motion taking the i–corona of a k–cell c to the i–corona of some other k–cell; here, ω_c is an orientation on c and $f_*(\omega_c)$ is the push-forward of this orientation to the cell $f(c)$. In the case that the cells of $\mathfrak{T}$ have trivial cell isotropy, one may consistently orient the cells of $\mathfrak{T}$, and this definition then just says that there exists some i for which ψ is constant on cells which have identical i–coronas up to rigid motion. The coboundary of an ePE cochain is ePE, so we have a cochain complex $C^\bullet(\mathfrak{T}^0)$ defined by restricting $C^\bullet(\mathcal{T})$ to ePE cochains. Taking the cohomology of this cochain complex, we define the *ePE cohomology* $H^\bullet(\mathfrak{T}^0)$.

One may follow the proof from [37] of Theorem 2.1 almost word-for-word, replacing the Gähler complexes Γ_i^1 by the complexes Γ_i^0, to obtain the following:

Theorem 3.1 *Let $\mathfrak{T}$ be an eFLC tiling whose cell isotropy groups $\widetilde{\mathrm{Iso}}(c,i)$ are trivial for some $i \in \mathbb{N}_0$. Then the ePE cohomology $H^\bullet(\mathfrak{T}^0)$ is isomorphic to the Čech cohomology $\check{H}^\bullet(\Omega^0_{\mathfrak{T}})$.*

We define Euclidean pattern-equivariance for cellular Borel–Moore chains identically as for cochains. Restricting $C^{\mathrm{BM}}_\bullet(\mathcal{T})$ to ePE chains we thus define the *ePE chain complex* $C_\bullet(\mathfrak{T}^0)$ and its homology, the *ePE homology* $H_\bullet(\mathfrak{T}^0)$.

The proof of PE Poincaré duality in Theorem 2.2 essentially relied on two simple observations:

(1) The classical Poincaré duality isomorphism $-\cap\Gamma$, given by taking the cap product with a Borel–Moore fundamental class, restricts to a cochain isomorphism $-\cap\Gamma\colon C^\bullet(\mathfrak{T}^1) \to C_{d-\bullet}(\widehat{\mathfrak{T}}^1)$ between the PE cohomology of $\mathfrak{T}$ and the PE homology of its dual tiling.

(2) The refinement of a tiling to its barycentric subdivision does not effect PE homology.

Step (1) will still hold for the ePE complexes: we have a Poincaré duality isomorphism $-\cap\Gamma\colon C^\bullet(\mathfrak{T}^0) \to C_{d-\bullet}(\widehat{\mathfrak{T}}^0)$ between ePE cochains of a tiling and ePE chains of its dual tiling. However, step (2) will not generally hold when restricting to ePE (co)chains if our tiling has nontrivial cell isotropy. One would like to refuse taking the ePE (co)homology of a tiling which has cells of nontrivial isotropy by replacing it with the barycentric subdivision. Unfortunately, our hand is forced, since, in the presence of nontrivial *patch* isotropy, one of $\mathfrak{T}$ or $\widehat{\mathfrak{T}}$ will have nontrivial *cell* isotropy.

The next lemma determines to what extent one may expect refinement to preserve ePE homology and cohomology. Thus far, our invariants have been taken over $\mathbb{Z}$ coefficients. For a unital ring G we may consider the cochain complex $C^\bullet(\mathcal{T};G)$ of cellular cochains which assign to oriented cells elements of G, and similarly for $C_\bullet^{\mathrm{BM}}(\mathcal{T};G)$. We may restrict these complexes to PE and ePE (co)chains, and denote the corresponding (co)homology by $H^\bullet(\mathfrak{T}^1;G)$, etc. We say that G has *division by n* if $n \cdot 1_G$ has a multiplicative inverse in G, where 1_G is the multiplicative identity in G and

$$n \cdot g := \underbrace{g + g + \cdots + g}_{n \text{ times}}.$$

Lemma 3.2 *Let $\mathfrak{T}$ be a polytopal tiling with eFLC and G be a unital ring. If for some $K \in \mathbb{N}_0$ the coefficient ring G has division by $\#\widetilde{\mathrm{Iso}}(c,K)$ for every cell c of $\mathfrak{T}$, then there exist quasi-isomorphisms*

$$\iota\colon C_\bullet(\mathfrak{T}^0;G) \to C_\bullet(\mathfrak{T}^0_\Delta;G),$$
$$\kappa\colon C^\bullet(\mathfrak{T}^0_\Delta;G) \to C^\bullet(\mathfrak{T}^0;G).$$

The analogous statement holds for the refinement of the dual tiling $\widehat{\mathfrak{T}}$ to $\mathfrak{T}_\Delta$.

Proof We shall prove the existence of the homology quasi-isomorphism, the proof for cohomology is similar. An *elementary chain* is a chain which assigns coefficient 1 to some oriented cell and 0 to all other cells. We have a chain map

$$\iota\colon C_\bullet^{\mathrm{BM}}(\mathcal{T};G) \to C_\bullet^{\mathrm{BM}}(\mathcal{T}_\Delta;G),$$

which is defined by sending an elementary chain to the corresponding barycentric chain with coefficient 1 on barycentric k–cells contained in c, suitably oriented with respect to c, and 0 on all other cells. It is easy to see that ι restricts to ePE chains and we claim that it is a quasi-isomorphism.

To show that ι is surjective on homology, let $\sigma \in C_k(\mathfrak{T}^0_\Delta)$ be an ePE cycle of the barycentric subdivision; σ is in the image of ι if and only if it is supported on the k–skeleton of $\mathfrak{T}$. If $k = d$, then σ is already supported on the d–skeleton, so suppose that $k < d$.

Whilst σ need not be in the image of ι at the chain level, there exists some τ for which $\sigma + \partial(\tau)$ is. To construct τ, we firstly find an ePE chain $\tau(d)$ for which $\sigma + \partial(\tau(d))$ is supported on the $(d-1)$–skeleton. Having the same i–corona up to rigid motion is an equivalence relation on the cells of $\mathfrak{T}$ for every $i \in \mathbb{N}_0$. By Euclidean pattern-equivariance of σ, there exists some i for which, if two d–cells c_1, c_2 of $\mathfrak{T}$ have identical i–coronas up to a rigid motion f, then f sends σ restricted at c_1 to its restriction at c_2. We may choose $i \geq K$; note that, since $\widetilde{\mathrm{Iso}}(c,i)$ is a subgroup of $\widetilde{\mathrm{Iso}}(c,K)$, we have that $\#\widetilde{\mathrm{Iso}}(c,i)$ divides $\#\widetilde{\mathrm{Iso}}(c,K)$, so the coefficient group G has division by $\#\widetilde{\mathrm{Iso}}(c,i)$ for every cell c of $\mathfrak{T}$.

For each equivalence class of d–cell, choose a representative c and a barycentric $(k+1)$–chain τ_c supported on c for which $\sigma + \partial(\tau_c)$ is supported outside of the interior of c; by the homological properties of cells of a CW decomposition, we may find such a chain. Define the $(k+1)$–chain τ_c' by copying τ_c to every cell equivalent to c, via every rigid motion which preserves the i–corona of c. We define $\tau(d) := \sum \tau_c'/(\#\widetilde{\mathrm{Iso}}(c,i))$, where the sum is taken over every equivalence class of d–cell.

The chain $\tau(d)$ is ePE by construction, and we claim that $\sigma + \partial(\tau(d))$ is supported on the $(d-1)$–skeleton. Indeed, let c be a chosen representative of d–cell; we have that $\partial(\tau_c) = -\sigma_c$, where σ_c is the restriction of σ to the interior barycentric cells of c. By our assumption on σ being ePE, for any $f \in \widetilde{\mathrm{Iso}}(c,i)$ we have that $f_*(\partial(\tau_c)) = f_*(-\sigma_c) = -\sigma_c$. Hence, the restriction of $\partial(\tau(d))$ to the interior of c is given by

$$\sum_{f \in \widetilde{\mathrm{Iso}}(c,i)} f_*(\partial(\tau_c))/(\#\widetilde{\mathrm{Iso}}(c,i)) = \sum_{f \in \widetilde{\mathrm{Iso}}(c,i)} -\sigma_c/(\#\widetilde{\mathrm{Iso}}(c,i)) = -\sigma_c.$$

By construction of $\tau(d)$, the same is true at every other d–cell equivalent to c, and by our assumption on σ being ePE it follows that $\sigma + \partial(\tau(d))$ is supported on the $(d-1)$–skeleton.

We may continue this procedure down the skeleta. That is, we may construct in an analogous way ePE chains $\tau(d), \tau(d-1), \ldots, \tau(k+1)$ for which $\sigma + \partial\bigl(\sum_{m=n}^{d} \tau(m)\bigr)$ is supported on the n–skeleton of $\mathfrak{T}$. It follows that $\sigma + \partial\bigl(\sum_{m=k+1}^{d} \tau(m)\bigr)$ is in the image of ι, so ι_* is surjective on homology.

Showing injectivity of ι_* is analogous (indeed, the above is really just a relative homology argument applied to the filtration of the skeleta). Suppose that $\iota(\sigma) = \partial(\tau)$ for $\sigma \in C_k(\mathfrak{T}^0)$ and $\tau \in C_{k+1}(\mathfrak{T}^0_\Delta)$. Then τ is ePE and has boundary in the k–skeleton of $\mathfrak{T}$. We may construct ePE $(k+2)$–chains $\nu(d), \nu(d-1), \ldots, \nu(k+2)$, analogously to above, for which $\tau + \partial\bigl(\sum_{m=k+2}^{d} \nu(m)\bigr)$ is contained in the $(k+1)$–skeleton. So there is an ePE chain τ' with $\iota(\tau') = \tau + \partial\bigl(\sum_{m=k+2}^{d} \nu(m)\bigr)$. It follows from $\partial(\iota(\tau')) = \partial(\tau) = \iota(\sigma)$ that $\sigma = \partial(\tau')$ represents zero in homology, as desired. □

By the above lemma, the ePE (co)homology of a tiling is stable under barycentric subdivision after one application, by the fact that the cell isotropy groups $\widetilde{\mathrm{Iso}}(c,i)$ are trivial in the barycentric subdivision. Invariance under barycentric refinement allows us to deduce ePE Poincaré duality, so long as our coefficient group is suitably divisible:

Theorem 3.3 *Let $\mathfrak{T}$ be an eFLC polytopal tiling. Suppose that, for some $K \in \mathbb{N}$, the coefficient ring G has division by the orders of isotropy groups $\#\,\mathrm{Iso}(c,K)$ for every cell c of $\mathfrak{T}$. Then we have ePE Poincaré duality $H^\bullet(\mathfrak{T}^0;G) \cong H_{d-\bullet}(\mathfrak{T}^0;G)$.*

Proof The proof is essentially identical to the proof of translational PE Poincaré duality of Theorem 2.2. All that needs to be checked is that we have invariance under refinement to the barycentric subdivision for the tiling and dual tiling, that is, that we have quasi-isomorphisms $\iota\colon C_\bullet(\mathfrak{T}^0;G) \to C_\bullet(\mathfrak{T}^0_\Delta;G)$ and $\hat{\iota}\colon C_\bullet(\widehat{\mathfrak{T}}^0;G) \to C_\bullet(\mathfrak{T}^0_\Delta;G)$.

The cell isotropy groups $\widetilde{\mathrm{Iso}}(c,K)$ of the tiling are quotient groups of the isotropy groups of K–coronas $\mathrm{Iso}(c,K)$ by the subgroups of those transformations leaving c fixed, and similarly for the dual tiling. Furthermore, any rigid motion preserving the $(K+1)$–corona of a dual cell $\hat{c}$, sending $\hat{c}$ to itself, also preserves the K–corona of the cell c in the original tiling. It follows that the cell isotropy groups (at level K for $\mathfrak{T}$ and $K+1$ for the dual tiling) have orders which divide those of the groups $\mathrm{Iso}(c,K)$. A unital ring which has division by n also has division by any divisor of n, and so by Lemma 3.2 we have the required refinement quasi-isomorphisms ι and $\hat{\iota}$. $\square$

Example 3.4 Let $\mathfrak{T}$ be the periodic cellular tiling of $\mathbb{R}^2$ of unit squares whose vertices lie on the integer lattice, with the standard cellular decomposition. The cells of $\mathfrak{T}$ have nontrivial isotropy: $\widetilde{\mathrm{Iso}}(f,i) \cong \mathbb{Z}/4$ for a face f and $\widetilde{\mathrm{Iso}}(e,i) \cong \mathbb{Z}/2$ for an edge e. So the ePE (co)homology groups are not necessarily invariant under barycentric subdivision unless taken over coefficients G with division by 4. Since there is only one 0–cell and one 2–cell up to rigid motion, the ePE complexes over $\mathbb{Q}$ coefficients read

$$0 \to \mathbb{Q} \to 0 \to \mathbb{Q} \to 0.$$

There is no generator in degree one, since an indicator (co)chain at an edge e is not invariant under the rigid motion at e reversing its orientation. So the ePE invariants are $H^k(\mathfrak{T}^0;\mathbb{Q}) \cong H_{2-k}(\mathfrak{T}^0;\mathbb{Q}) \cong \mathbb{Q}$ for $k=0,\,2$ and are trivial otherwise.

To calculate over $\mathbb{Z}$ coefficients, we pass to the barycentric subdivision $\mathfrak{T}_\Delta$ so that the cells have trivial isotropy. In this case we have that $H^k(\mathfrak{T}^0_\Delta) \cong \mathbb{Z}$ for $k=0,2$ and are trivial otherwise. This agrees with the observation that $\Omega^0_{\mathfrak{T}}$ is homeomorphic to the 2–sphere, which by Theorem 3.1 has isomorphic cohomology. For the ePE homology we have that $H_k(\mathfrak{T}^0_\Delta) \cong \mathbb{Z} \oplus (\mathbb{Z}/2) \oplus (\mathbb{Z}/4)$, 0, $\mathbb{Z}$ for $k=0,\,1,\,2$, respectively.

For this example ePE Poincaré duality $H^k(\mathfrak{T}^0_\Delta) \cong H_{d-k}(\mathfrak{T}^0_\Delta)$ fails over $\mathbb{Z}$ coefficients. Theorem 3.3 does not apply since, whilst the *cells* of $\mathfrak{T}_\Delta$ have trivial isotropy, the isotropy groups $\mathrm{Iso}(v,i)$ of rigid motions preserving *patches* are nontrivial. The ePE homology in degree zero for a periodic tiling of equilateral triangles is $\mathbb{Z} \oplus (\mathbb{Z}/2) \oplus (\mathbb{Z}/3)$.

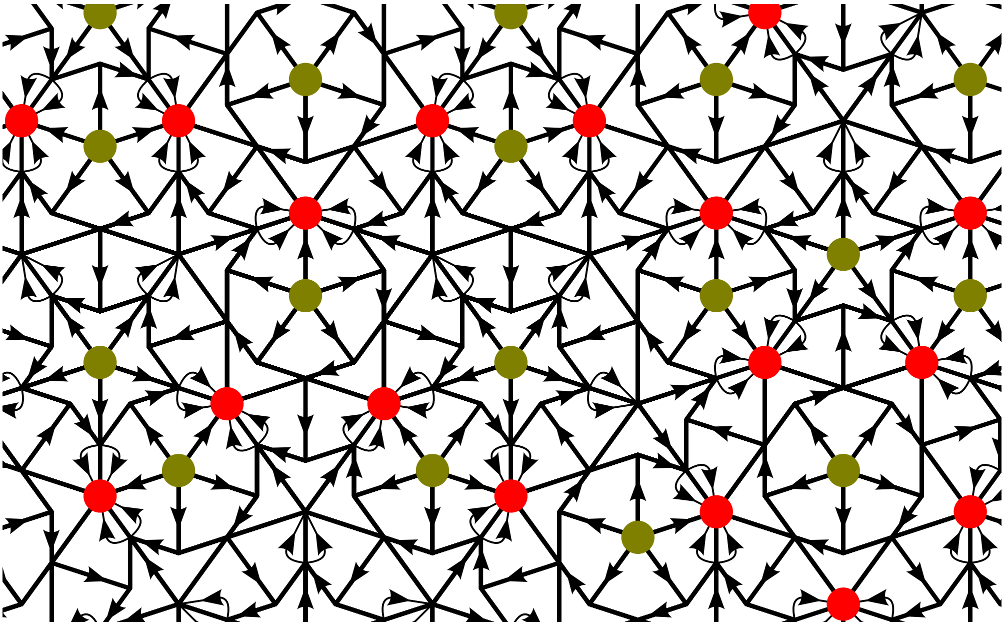

Figure 3: Torsion element t with $5t + \partial_1(-\mathbb{1}(\mathrm{E1}) + \mathbb{1}(\mathrm{E2}) - \mathbb{1}(\mathrm{E4}) - 2 \cdot \mathbb{1}(\mathrm{E7})) = 0$

However, its associated moduli space $\Omega^0_{\mathfrak{T}}$ is still the 2–sphere, so we see that the ePE homology is not a topological invariant of $\Omega^0_{\mathfrak{T}}$ but of a finer structure.

Example 3.5 Let $\mathfrak{T}$ be a Penrose kite and dart tiling. Its ePE cohomology is

$$H^k(\mathfrak{T}^0) \cong \check{H}^k(\Omega^0_{\mathfrak{T}}) \cong \begin{cases} \mathbb{Z} & \text{for } k = 0, \\ \mathbb{Z} & \text{for } k = 1, \\ \mathbb{Z}^2 & \text{for } k = 2. \end{cases}$$

In degrees $k = 1,\ 2$ we have ePE Poincaré duality $H_k(\mathfrak{T}^0) \cong H^{2-k}(\mathfrak{T}^0)$. But in degree zero, as we shall calculate in Section 4, we have that $H_0(\mathfrak{T}^0) \cong \mathbb{Z}^2 \oplus (\mathbb{Z}/5)$. A 0–chain t representing a 5–torsion homology class is depicted in Figure 3, along with an ePE 1–chain whose boundary is $-5t$. Specifically, the torsion element $t = \mathbb{1}(\text{sun}) + \mathbb{1}(\text{star}) - \mathbb{1}(\text{queen})$ is a linear combination of indicator 0–chains of certain rigid equivalence classes of star patches of 0–cells (the seven equivalence classes of such star patches are given in Figure 4). This torsion element turns out to be relevant to calculation of the Čech cohomology $\check{H}^\bullet(\Omega^{\mathrm{rot}}_{\mathfrak{T}})$ of the Euclidean hull [42].

Generators for the free part of $H_0(\mathfrak{T}^0)$ may be taken as $\mathbb{1}(\text{sun})$ and $\mathbb{1}(\text{star})$. A generator of $H_1(\mathfrak{T}^0) \cong \mathbb{Z}$ is illustrated in red in Figure 2. It is the cycle running along the bottoms of the dart tiles, named ρ' in the discussion of Example 2.3.

3.3 Restoring Poincaré duality

The above examples show that ePE Poincaré duality can fail in the presence of nontrivial rotational symmetry. We consider the discrepancy between the ePE homology and ePE cohomology to be a feature of interest, which is relevant to the cohomology of Euclidean hulls [42]. However, to relate the ePE homology back to more familiar invariants, we shall describe here how one may modify the definition of ePE homology so as to restore duality with the ePE cohomology.

We restrict to the case that $\mathfrak{T}$ is a tiling of $\mathbb{R}^2$. The higher-dimensional situation is much more complicated; see the comments at the end of this subsection. We assume that $\mathfrak{T}$ has eFLC and that it has been suitably subdivided so that any points of local rotational symmetry are contained in the vertex set of $\mathfrak{T}$; this may be achieved for any eFLC polytopal tiling by a single barycentric subdivision.

Definition 3.6 Define the subchain complex

$$C_\bullet^\dagger(\mathfrak{T}^0) := 0 \leftarrow C_0^\dagger(\mathfrak{T}^0) \xleftarrow{\partial_1} C_1^\dagger(\mathfrak{T}^0) \xleftarrow{\partial_2} C_2^\dagger(\mathfrak{T}^0) \leftarrow 0$$

of the ePE complex $C_\bullet(\mathfrak{T}^0)$ of $\mathfrak{T}$ as follows. We let $C_k^\dagger(\mathfrak{T}^0) := C_k(\mathfrak{T}^0)$ for $k = 1, 2$. In degree zero we let $C_0^\dagger(\mathfrak{T}^0)$ consist of those ePE chains σ for which there exists some $i \in \mathbb{N}$ such that, whenever the i–corona of a vertex v has rotational symmetry of order n, then the coefficient of v in σ is divisible by n. Denote the homology of this chain complex by $H_\bullet^\dagger(\mathfrak{T}^0)$.

To see that $C_\bullet^\dagger(\mathfrak{T}^0)$ is well-defined, firstly note that the boundary of an ePE chain is ePE, so it suffices to check that, given an ePE 1–chain σ, there exists some i for which $\partial(\sigma)$ assigns values multiples of n to vertices whose i–coronas have n–fold symmetry. Since σ is ePE, there exists some j for which σ assigns the same (oriented) coefficients to any two edges whose j–coronas are equivalent up to a rigid motion. Suppose that the $(j+1)$–corona of a vertex v has n–fold rotational symmetry; these symmetries induce rigid equivalences between j–coronas of the edges incident with v. Since no edge is fixed by any nontrivial rotation, the rotations partition these edges into orbits of n elements, each having equivalent j–coronas. It follows that the coefficient of v in $\partial(\sigma)$ is some multiple of n, as desired.

With only minor modifications to the proof of Theorem 3.3 we obtain the following:

Theorem 3.7 *Let $\mathfrak{T}$ be a polytopal tiling of $\mathbb{R}^2$ with eFLC and with points of local rotational symmetry contained in the vertex set of $\mathfrak{T}$. Then we have Poincaré duality*

$$H^\bullet(\mathfrak{T}^0) \cong H_{2-\bullet}^\dagger(\mathfrak{T}^0).$$

Proof From the classical Poincaré duality pairing, we have an isomorphism of complexes $C^\bullet(\mathfrak{T}^0) \cong C_{d-\bullet}(\widehat{\mathfrak{T}}^0)$ between the ePE cohomology and the ePE homology of the dual tiling. The issue with ePE Poincaré duality is that we do not necessarily have an isomorphism $H_\bullet(\widehat{\mathfrak{T}}^0) \cong H_\bullet(\mathfrak{T}^0)$. In particular, we may not have a refinement quasi-isomorphism $\hat{\iota}\colon C_\bullet(\widehat{\mathfrak{T}}^0) \to C_\bullet(\mathfrak{T}^0_\Delta)$; the conditions of Lemma 3.2 are not satisfied since vertices with local rotational symmetry in $\mathfrak{T}$ lead to dual tiles of $\widehat{\mathfrak{T}}$ with nontrivial cell isotropy.

Following the proof of Lemma 3.2, we see that $\hat{\iota}$ can be made a quasi-isomorphism by replacing its range with $C^\dagger_\bullet(\mathfrak{T}^0_\Delta)$. Let

$$\hat{\iota}^\dagger\colon C_\bullet(\widehat{\mathfrak{T}}^0) \to C^\dagger_\bullet(\mathfrak{T}^0_\Delta)$$

be the canonical inclusion of chain complexes. The map $\hat{\iota}$ may fail to be a quasi-isomorphism in degree zero. It may not be the case that an ePE 0–chain σ of $\mathfrak{T}_\Delta$ is homologous to a chain supported on the 0–skeleton of $\widehat{\mathfrak{T}}$, since we are forced to "remove" 0–chains of $C_0(\mathfrak{T}^0_\Delta)$ from barycentres of rotationally invariant dual cells in multiples of the local symmetry at the corresponding vertices of $\mathfrak{T}$. This issue is alleviated by passing to $C^\dagger_\bullet(\mathfrak{T}^0_\Delta)$, since now the barycentres of such dual cells may only be assigned coefficients which are multiples of the orders of symmetries of the corresponding i–corona in the dual tiling for some sufficiently large i. The rest of the proof follows similarly to the proof of Theorem 3.3; we end up with the diagram of quasi-isomorphisms

$$C^\bullet(\mathfrak{T}^0) \xrightarrow{-\cap\Gamma} C_{d-\bullet}(\widehat{\mathfrak{T}}^0) \xrightarrow{\hat{\iota}^\dagger} C^\dagger_{d-\bullet}(\mathfrak{T}^0_\Delta) \xleftarrow{\iota^\dagger} C^\dagger_{d-\bullet}(\mathfrak{T}^0). \qquad \square$$

This modification to the ePE homology is perhaps not too surprising when we think of the approximant spaces to Ω^0 as branched *orbifolds* rather than just quotient spaces. Points of rotational symmetry in the tiling correspond to cone points on these orbifolds, and the modification above is essentially to count such points with fractional multiplicity.

The above theorem shows that we may express the ePE cohomology of a two-dimensional tiling, and hence the Čech cohomology of the associated space $\Omega^0_{\mathfrak{T}}$, in terms of the ePE homology of $\mathfrak{T}$ but with certain restricted coefficients in degree zero. One may ask on the relationship between the ePE homology before and after the modification to restore Poincaré duality in degree zero. The only nontrivial chain group of the relative chain complex of the pair $C_\bullet(\mathfrak{T}^0) \le C^\dagger_\bullet(\mathfrak{T}^0)$ is a torsion group in degree zero. It is not difficult to show that it is isomorphic to $\prod_{\mathfrak{T}_i} \mathbb{Z}/n_i\mathbb{Z}$, where the product is taken over all rotation classes of tilings $\mathfrak{T}_i$ of $\Omega^0_{\mathfrak{T}}$ with n_i–fold rotational symmetry at the origin, at least in the case that there are only finitely many such tilings (and a similar statement holds still with infinitely many such tilings). It follows that the ePE homology and the Čech cohomology of $\Omega^0_{\mathfrak{T}}$ are isomorphic in cohomological degrees one and two, and in degree zero we have that $H_0(\mathfrak{T}^0)$ is an extension of $\check{H}^2(\Omega^0_{\mathfrak{T}})$ by a torsion group determined by the rotational symmetries of tilings in the hull of $\mathfrak{T}$.

In higher dimensions the situation is far more complicated, and we delay exposition of it to future work. The precise relationship between the ePE homology and ePE cohomology is then best expressed via a more complicated gadget, a spectral sequence analogous to the one of Zeeman [43].

Example 3.8 Consider again the periodic tiling $\mathfrak{T}$ of unit squares. Its barycentric subdivision $\mathfrak{T}_\Delta$ has trivial cell isotropy, but has rotational symmetry at the vertices of $\mathfrak{T}_\Delta$ (ie at the barycentres of the cells of $\mathfrak{T}$). In particular, the vertices have rotational symmetry of orders 4, 2 and 4 at the vertices of $\mathfrak{T}_\Delta$ corresponding, respectively, to the vertices, edges and faces of $\mathfrak{T}$. So we replace the degree zero ePE chain group

$$C_0(\mathfrak{T}^0_\Delta) \cong \mathbb{Z}\langle v\rangle \oplus \mathbb{Z}\langle e\rangle \oplus \mathbb{Z}\langle f\rangle.$$

by its modified version

$$C_0^\dagger(\mathfrak{T}^0_\Delta) \cong 4\mathbb{Z}\langle v\rangle \oplus 2\mathbb{Z}\langle e\rangle \oplus 4\mathbb{Z}\langle f\rangle.$$

One easily computes the resulting homology group in degree zero to be $H_0^\dagger(\mathfrak{T}^0_\Delta) \cong \mathbb{Z}$, restoring Poincaré duality:

$$H_0^\dagger(\mathfrak{T}^0_\Delta) \cong H^2(\mathfrak{T}^0_\Delta) \cong \check{H}^2(\Omega^0_{\mathfrak{T}_\Delta}) \cong H^2(S^2) \cong \mathbb{Z}.$$

Example 3.9 We saw in Example 3.5 that ePE Poincaré duality $H_k(\mathfrak{T}^0) \cong H^{2-k}(\mathfrak{T}^0)$ fails for the Penrose kite and dart tilings in homological degree $k=0$; we have extra 5–torsion in the ePE homology, a generator is depicted in Figure 3. Our method of calculation for the ePE homology of substitution tilings in Section 4 may be modified to compute instead $H_0^\dagger(\mathfrak{T}^0)$. We calculate that, indeed, Poincaré duality is restored:

$$H_0^\dagger(\mathfrak{T}^0) \cong H^2(\mathfrak{T}^0) \cong \check{H}^2(\Omega^0_{\mathfrak{T}}) \cong \mathbb{Z}^2.$$

The modified degree zero ePE homology group is freely generated by, for example, the indicator 0–chains of the "queen" and "king" vertex types (see Figure 4).

3.4 Rotation actions on translational PE cohomology

In the case that our tiling $\mathfrak{T}$ is FLC (and not just eFLC) there is an alternative way of integrating the action of rotations with the PE invariants of $\mathfrak{T}$. Firstly, we assume that some rotation group acts nicely on $\mathfrak{T}$:

Definition 3.10 We say that a finite subgroup $\Theta \leq \mathrm{SO}(d)$ *acts on $\mathfrak{T}$ by rotations* if, for every patch P of $\mathfrak{T}$ and $g \in \Theta$, we have that gP is also a patch of $\mathfrak{T}$, up to translation. If, additionally, we have that patches P and Q agree up to a rigid motion only when P and gQ agree up to translation for some $g \in \Theta$, then we say that $\mathfrak{T}$ *has rotation group* Θ.

A PE k–cochain $\psi \in C^k(\mathfrak{T}^1)$ may be identified with a sum of indicator cochains of i–coronas of k–cells of $\mathfrak{T}$ for some sufficiently large i. So if Θ acts on $\mathfrak{T}$ by rotations, then Θ naturally acts on $C^\bullet(\mathfrak{T}^1)$. On an indicator cochain $\mathbb{1}(P_c)$, the group Θ acts by $g \cdot \mathbb{1}(P_c) := \mathbb{1}(g^{-1} \cdot P_c)$. Replacing cellular cochains with cellular Borel–Moore chains, the same is true of the PE *homology*. The action of Θ commutes with the (co)boundary maps, and so we have an action of Θ on the PE (co)homology. Furthermore, Θ naturally acts as a group of homeomorphisms on the tiling space $\Omega^1_{\mathfrak{T}}$; the points of $\Omega^1_{\mathfrak{T}}$ may be identified with tilings, and Θ acts by rotation at the origin. So Θ acts on the Čech cohomology $\check{H}^\bullet(\Omega^1_{\mathfrak{T}})$. These actions are compatible:

Proposition 3.11 *Suppose that $\mathfrak{T}$ is FLC and that Θ acts on $\mathfrak{T}$ by rotations. Then the isomorphisms $\check{H}^\bullet(\Omega^1_{\mathfrak{T}}) \cong H^\bullet(\mathfrak{T}^1) \cong H_{d-\bullet}(\mathfrak{T}^1)$ of Theorems 2.1 and 2.2 commute with the group actions of Θ.*

Proof The action of rotation on $\Omega^1_{\mathfrak{T}}$ is canonically induced at the level of the approximants Γ^1_i. The isomorphism between the Čech cohomology of an inverse limit space and the direct limit of the cohomologies of its approximants is natural with respect to maps like this, so the isomorphisms

$$\check{H}^\bullet(\Omega^1_{\mathfrak{T}}) \cong \check{H}^\bullet(\varprojlim(\Gamma^1_i, \pi_{i,j})) \cong \varinjlim(H^\bullet(\Gamma^1_i), \pi^*_{i,j}) \cong H^\bullet(\mathfrak{T}^1)$$

each commute with the action of rotation. The Poincaré duality isomorphism $H^\bullet(\mathfrak{T}^1) \cong H_{d-\bullet}(\mathfrak{T}^1)$ of Theorem 2.2 was induced by the classical pairing of a cochain with its dual chain, along with the induced maps (and their inverses) associated to barycentric refinement. Each of these maps, at the chain level, are easily seen to commute with the action of rotation. $\square$

Suppose that $\mathfrak{T}$ has rotation group Θ. One may ask to what extent the Čech cohomology $\check{H}^\bullet(\Omega^0_{\mathfrak{T}})$ naturally corresponds to the subgroup of $\check{H}^\bullet(\Omega^1_{\mathfrak{T}})$ of elements of which are invariant under the action of Θ. More concretely, we have the quotient map

$$q\colon \Omega^1_{\mathfrak{T}} \to \Omega^0_{\mathfrak{T}} = \Omega^1_{\mathfrak{T}}/\Theta$$

given by identifying tilings which agree up to a rotation at the origin. Since $q = q \circ g$ for all $g \in \Theta$, the induced map

$$q^*\colon \check{H}^\bullet(\Omega^0_{\mathfrak{T}}) \to \check{H}^\bullet(\Omega^1_{\mathfrak{T}})$$

has image contained in the *rotationally invariant part of* $\check{H}^\bullet(\Omega^1_{\mathfrak{T}})$, denoted by

$$\check{H}^\bullet_\Theta(\Omega^1_{\mathfrak{T}}) := \{[\psi] \in \check{H}^\bullet(\Omega^1_{\mathfrak{T}}) \mid [\psi] = g \cdot [\psi] \text{ for all } g \in \Theta\}.$$

Let q^*_Θ be the corestriction of q^* to the range $\check{H}^\bullet_\Theta(\Omega^1_{\mathfrak{T}})$. It is not difficult to show that q^*_Θ is an isomorphism when taking cohomology over divisible coefficients:

Proposition 3.12 *Let $\mathfrak{T}$ be FLC with rotation group Θ. If G is a unital ring with division by $\#\Theta$ then $q^*_\Theta\colon \check{H}^\bullet(\Omega^0_{\mathfrak{T}};G) \to \check{H}^\bullet_\Theta(\Omega^1_{\mathfrak{T}};G)$ is an isomorphism.*

Proof We suppose that the cell isotropy groups of $\mathfrak{T}$ are trivial (without loss of generality, since otherwise we may simply pass to the barycentric subdivision). By the natural identification of the Čech cohomology with the PE cohomology, the map q^* corresponds to the induced map of the inclusion of cochain complexes

$$\iota\colon C^\bullet(\mathfrak{T}^0) \hookrightarrow C^\bullet(\mathfrak{T}^1);$$

note that $C^\bullet(\mathfrak{T}^0)$ is the subcochain complex of $C^\bullet(\mathfrak{T}^1)$ of cochains which are invariant under the action of Θ. There is a self-cochain map

$$r\colon C^\bullet(\mathfrak{T}^1) \to C^\bullet(\mathfrak{T}^1)$$

defined by $r(\psi) := \sum_{g\in\Theta} g\cdot\psi$. Clearly $g\cdot r(\psi) = r(\psi)$ for any $\psi \in C^\bullet(\mathfrak{T}^1)$, so r in fact defines a map into the ePE cochain complex. We have that $r\circ\iota = \iota\circ r$ is the times $\#\Theta$ map upon restriction to the rotationally invariant part of the PE cohomology. By our assumption on the divisibility of the coefficient group G, this map is an isomorphism when taking cohomology over G coefficients, and so q^*_Θ, is an isomorphism. □

When working over nondivisible coefficients, q^*_Θ is typically not an isomorphism. For a two-dimensional tiling, we may factor q^*_Θ through the ePE homology of $\mathfrak{T}$:

Theorem 3.13 *Suppose that $\mathfrak{T}$ is two-dimensional, FLC, has points of local rotational symmetry contained in the vertex set of $\mathfrak{T}$ and has rotation group Θ. Then we have the following commutative triangle, with i injective:*

$$\begin{tikzcd} \check{H}^\bullet(\Omega^0_{\mathfrak{T}}) \arrow[rr, "q^*_\Theta"] \arrow[dr, hook, "i"'] & & \check{H}^\bullet_\Theta(\Omega^1_{\mathfrak{T}}) \\ & H_{2-\bullet}(\mathfrak{T}^0) \arrow[ur, "f"'] & \end{tikzcd}$$

Proof The inclusions of chain complexes $C^\dagger_\bullet(\mathfrak{T}^0) \le C_\bullet(\mathfrak{T}^0) \le C_\bullet(\mathfrak{T}^1)$ induce a triangle much like the one above. The first inclusion induces an inclusion on homology, since the corresponding relative homology group is concentrated in degree zero. The rest of the proof follows from establishing that, up to isomorphism, the map q^* corresponds to the induced map of the inclusion $C_\bullet^\dagger(\mathfrak{T}^0) \leq C_\bullet(\mathfrak{T}^1)$. This is a straightforward check, applying the ideas of the proof of Proposition 3.11. □

One might hope for the homomorphism f in the theorem above to always be surjective. In that case, we may interpret the theorem as follows: the ePE homology extends the ePE cohomology by adding "missing" elements corresponding to PE cochains whose cohomology classes are Θ–invariant, but are not actually represented by Θ–invariant (ePE) cochains. Whilst f usually does have a larger image than q_Θ^*, the example of the Penrose tiling below shows that, in fact, f need not be surjective in general.

Example 3.14 Let $\mathfrak{T}_\Delta$ be the barycentric subdivision of the periodic square tiling. In degree two we have that $\check{H}^2(\Omega_\mathfrak{T}^1) \cong \mathbb{Z}$ which, in terms of PE cohomology, is freely generated by an indicator cochain for the square tiles, by a choice of some 2–simplex of the barycentric subdivision of the unit square. Its cohomology class (but not the cochain itself) is invariant under rotation, so $H_\Theta^2(\Omega_\mathfrak{T}^1) = H^2(\Omega_\mathfrak{T}^1)$.

The ePE cohomology is freely generated by the 2–cochain which indicates each 2–simplex of a chosen handedness, the map $q_\Theta^*\colon \mathbb{Z} \to \mathbb{Z}$ is the times 4 map in degree two. So there are classes of the PE cohomology which are invariant under rotation, but are not represented by rotationally invariant cochains. These "missing" elements are represented in the ePE homology, though. We have that $H_0(\mathfrak{T}_\Delta) \cong \mathbb{Z} \oplus (\mathbb{Z}/2) \oplus (\mathbb{Z}/4)$ has free part generated by the 0–chain indicating the centres of squares. In degree zero q_Θ^* factorises as $q_\Theta^* = f \circ i$ where i is the $\times 4$ map onto the free component of $H_0^\dagger(\mathfrak{T}^0)$, and the map f is given by the projection $f(x,[y]_2,[z]_4) = x$.

Example 3.15 Let $\mathfrak{T}$ be a Penrose kite and dart tiling. The cohomology $\check{H}^2(\Omega_\mathfrak{T}^1)$, along with the action of rotation by $\mathbb{Z}/10$ on it, was first analysed in [32]. By Proposition 3.11, we may essentially mimic such calculations using instead PE homology. We compute (according to the method to be outlined in the next section), consistently with previous calculations, that over rational coefficients the action of rotation on $\check{H}^2(\Omega_\mathfrak{T}^1;\mathbb{Q}) \cong H_0(\mathfrak{T};\mathbb{Q}) \cong \mathbb{Q}^8$ splits into the following irreducible subrepresentations. We have two one-dimensional irreducibles corresponding to the trivial representation. There are two one-dimensional irreducibles corresponding to the representation sending the generator $[1]_{10} \in \mathbb{Z}/10$ to the map $x \mapsto -x$. And we have a four-dimensional irreducible, the "vector representation" $\mathbb{Q}[r]/(r^4 - r^3 + r^2 - r + 1)$, which sends $[1]_{10}$ to the map $(w,x,y,z) \mapsto (-z, w+z, x-z, y+z)$.

However, over integral coefficients the representation does not decompose into irreducibles. We find elements $\sigma_1,\dots,\sigma_4$ of $H_0(\mathfrak{T}^1) \cong \mathbb{Z}^8$ upon which rotation acts trivially on σ_1 and σ_2, and sends $\sigma_3 \mapsto -\sigma_3$ and $\sigma_4 \mapsto -\sigma_4$. One may extend either of the pairs (σ_1,σ_2) and (σ_3,σ_4) to integer bases for $\mathbb{Z}^8$, but the integer span of $(\sigma_1,\sigma_2,\sigma_3,\sigma_4)$ only extends to an index 4 subgroup of $H_0(\mathfrak{T}^1)$.

In degree one, over integral coefficients, the action of rotation on $\check{H}^1(\Omega^1_{\mathfrak{T}}) \cong H_1(\mathfrak{T}^1) \cong \mathbb{Z}^5$ decomposes to irreducibles. In terms of PE chains, it splits as the direct sum of the one-dimensional trivial representation generated by the cycle ρ (see Example 2.3) and the four-dimensional vector representation generated by ν_0 (see Example 2.3 and Figure 1) and its first three rotates.

We may now explain how the ePE (co)homology and the invariant part of the PE cohomology are tied together. Recall from Examples 3.5 and 3.9 that

$$\begin{aligned}
\check{H}^1(\Omega^0_{\mathfrak{T}}) &\cong H_1^{\dagger}(\mathfrak{T}^0) \cong H_1(\mathfrak{T}^0) \cong \mathbb{Z}\langle \rho' \rangle, \\
\check{H}^2(\Omega^0_{\mathfrak{T}}) &\cong H_0^{\dagger}(\mathfrak{T}^0) \cong \mathbb{Z}\langle \mathbb{1}(\text{queen}) \rangle \oplus \mathbb{Z}\langle \mathbb{1}(\text{king}) \rangle, \\
H_0(\mathfrak{T}^0) &\cong \mathbb{Z}\langle \mathbb{1}(\text{sun}) \rangle \oplus \mathbb{Z}\langle \mathbb{1}(\text{star}) \rangle \oplus (\mathbb{Z}/5)\langle \mathbb{1}(\text{sun}) + \mathbb{1}(\text{star}) - \mathbb{1}(\text{queen}) \rangle.
\end{aligned}$$

The rotationally invariant parts of the cohomology are

$$\check{H}^1_{\Theta}(\Omega^1_{\mathfrak{T}}) \cong \mathbb{Z}, \quad \check{H}^2_{\Theta}(\Omega^1_{\mathfrak{T}}) \cong \mathbb{Z}^2.$$

Repeating this calculation using PE homology, we find that, indeed, the rotationally invariant part of $H_0(\mathfrak{T}^1)$ is isomorphic to $\mathbb{Z}^2$, and of $H_1(\mathfrak{T}^1)$ is isomorphic to $\mathbb{Z}$. Furthermore, we may calculate explicit generators for these subgroups in PE homology. We find that the rotationally invariant part in degree zero is generated by PE 0–chains σ_1 and σ_2 for which $5\sigma_1 \simeq \mathbb{1}(\text{queen})$ and $5\sigma_2 \simeq \mathbb{1}(\text{king})$. The rotationally invariant part of the PE homology in degree one is freely generated by the chain ρ, discussed in Example 2.3, given by restricting the 1–chain of Figure 2 to loops of a chosen rotational parity.

With respect to the basis elements discussed above, we may summarise with the following commutative diagrams in cohomological degrees one and two:

$$\begin{tikzcd}
\check{H}^1(\Omega^0_{\mathfrak{T}}) \cong \mathbb{Z} \arrow[rr, "q^1_{\Theta} = \times 2"] \arrow[dr, "i^1 = \mathrm{id}"'] & & \check{H}^1_{\Theta}(\Omega^1_{\mathfrak{T}}) \cong \mathbb{Z} \\
& H_1(\mathfrak{T}^0) \cong \mathbb{Z} \arrow[ur, "f^1 = \times 2"'] &
\end{tikzcd}$$

$$\begin{tikzcd}
\check{H}^2(\Omega^0_{\mathfrak{T}}) \cong \mathbb{Z}^2 \arrow[rr, "q^2_{\Theta}(x,y) = (5x,5y)"] \arrow[dr, "i^2"'] & & \check{H}^2_{\Theta}(\Omega^1_{\mathfrak{T}}) \cong \mathbb{Z}^2 \\
& H_0(\mathfrak{T}^0) \cong \mathbb{Z}^2 \oplus (\mathbb{Z}/5) \arrow[ur, "f^2"'] &
\end{tikzcd}$$

The maps i^2 and f^2 are given by $i^2(x,y) = (x+2y, x-3y, [4x+3y]_5)$ and $f^2(x,y,[z]_5) = (3x+2y, x-y)$. In degree one, the ePE (co)homology only corresponds to an index 2 subgroup of the rotationally invariant part of the PE cohomology. In top cohomological degree, we have that the ePE cohomology $H^2(\mathfrak{T}^0)$ (which corresponds to $\check{H}^2(\Omega^0_{\mathfrak{T}})$ or $H^\dagger_0(\mathfrak{T}^0)$) maps to an index 25 subgroup of the rotationally invariant part $\check{H}^2_\Theta(\Omega^1_{\mathfrak{T}})$ of the PE cohomology. More, but not all, is added by considering instead the ePE *homology*: $H^\dagger_0(\mathfrak{T}^0) \cong H^2(\mathfrak{T}^0)$ is an index 5 subgroup of the ePE homology $H_0(\mathfrak{T}^0)$, and the image of $H_0(\mathfrak{T}^0)$ under f is an index 5 subgroup of $\check{H}^2_\Theta(\Omega^1_{\mathfrak{T}})$.

3.5 Generalising the PE framework

Many of the constructions and results of this section did not rely on having a polytopal tiling of Euclidean space, so much as simply having a cell complex (the underlying complex of the tiling) along with a notion of when cells of that complex are equivalent to a certain radius (that is, when those cells have identical i–coronas, up to an agreed type of transformation). There are interesting examples of combinatorial tilings, such as the pentagonal tilings of Bowers and Stephenson [11], which are most naturally viewed as tilings of spaces which are non-Euclidean. We outline below a unified setting which allows one to deal with tilings such as this, as well as more general structures; see Example 3.18.

Recall that a CW complex is called *regular* if the attaching maps of its cells may be taken to be homeomorphisms. Regular CW complexes are a sensible starting point for us here, since they allow for the construction of barycentric subdivisions and dual complexes (which, in analogy to simplicial complexes, is owing to them being essentially determined combinatorially by their face posets; see [9]). Let $\mathcal{T}$ be a regular CW complex, it will play the rôle of the underlying cell complex of our tiling of interest.

To define the analogue of pattern-equivariance of a (co)chain, we need a notion of two (oriented) cells being equivalent in the tiling to a certain magnitude. This "magnitude" could be parametrised by, say, $\mathbb{R}_{>0}$ if we want to express agreement between local patches to a certain radius, or perhaps by $\mathbb{N}$ for a combinatorial notion of patch size, such as agreement between i–coronas. Ultimately, there is no gain in preferring one, or indeed either of these choices. Recall that a partially ordered set $(\Lambda, \leq)$ is called *directed* if for any two elements $\lambda_1, \lambda_2 \in \Lambda$ there is a third satisfying $\lambda \geq \lambda_1, \lambda_2$. We shall let some such directed set parametrise magnitude of agreement between cells of our tiling.

It is not quite enough to know which cells of our tiling are equivalent to a certain magnitude, one also needs to know *how* they are equivalent. For a cell c of $\mathcal{T}$, its (*closed*) *star* $\mathrm{St}_{\mathcal{T}}(c)$ is defined to be the subcomplex of $\mathcal{T}$ whose support is the set of cells containing c. The way in which a cell c_1 may be considered as "equivalent" to a cell c_2 to some magnitude λ will be recorded by a finite set of cellular homeomorphisms $\Phi\colon \mathrm{St}_{\mathcal{T}}(c_1) \to \mathrm{St}_{\mathcal{T}}(c_2)$. The star of a cell defines a neighbourhood of that cell, so we may consider such morphisms as defining germs of maps by which two cells are equivalent.

With these interpretations of the ingredients, we may give the following definition, which provides a structure upon which one may define PE (co)homology and various other related constructions. The axioms will be further motivated below.

Definition 3.16 A *system of internal symmetries* (or *SIS*, for short) $\mathfrak{T}$ consists of the following data:

- A finite-dimensional and locally finite regular CW complex $\mathcal{T}$.
- A directed set $(\Lambda, \leq)$ called the *magnitude poset*.
- For each $\lambda \in \Lambda$ and each pair of cells a, $b \in \mathcal{T}$, a set $\mathfrak{T}^{\lambda}_{a,b}$ of cellular homeomorphisms $\Phi\colon \mathrm{St}_{\mathcal{T}}(a) \to \mathrm{St}_{\mathcal{T}}(b)$ sending a to b. We denote the collection of all such morphisms by $\mathfrak{T}^{\lambda}$.

This data is required to satisfy the following:

(G1) For all $\lambda \in \Lambda$ and $a \in \mathcal{T}$ we have that $\mathrm{id}_{\mathrm{St}(a)} \in \mathfrak{T}^{\lambda}_{a,a}$.

(G2) For all $\lambda \in \Lambda$ and $\Phi \in \mathfrak{T}^{\lambda}_{a,b}$ we have that $\Phi^{-1} \in \mathfrak{T}^{\lambda}_{b,a}$.

(G3) For all $\lambda \in \Lambda$, $\Phi_1 \in \mathfrak{T}^{\lambda}_{a,b}$ and $\Phi_2 \in \mathfrak{T}^{\lambda}_{b,c}$, we have that $\Phi_2 \circ \Phi_1 \in \mathfrak{T}^{\lambda}_{a,c}$.

(Inc) For all $\lambda_1 \leq \lambda_2$ we have that $\mathfrak{T}^{\lambda_1} \supseteq \mathfrak{T}^{\lambda_2}$.

(Res) For all $\lambda \in \Lambda$ there exists some $\lambda_{\mathrm{res}} \geq \lambda$ satisfying the following. Given any $b \in \mathcal{T}$ and face $a \preceq b$, every morphism of $\mathfrak{T}^{\lambda_{\mathrm{res}}}_{a,-}$ restricts to a morphism of $\mathfrak{T}^{\lambda}_{b,-}$.

($\widehat{\text{Res}}$) Dually, for all $\lambda \in \Lambda$ there exists some $\lambda_{\widehat{\mathrm{res}}} \geq \lambda$ satisfying the following. Given any $a \in \mathcal{T}$ and coface $b \succeq a$, every morphism of $\mathfrak{T}^{\lambda_{\widehat{\mathrm{res}}}}_{b,-}$ is a restriction of some morphism of $\mathfrak{T}^{\lambda}_{a,-}$.

As explained above, morphisms $\Phi \in \mathfrak{T}^{\lambda}_{a,b}$ should be interpreted as recording that *cell a is equivalent to cell b to magnitude λ via Φ*. The groupoid axioms (G1)–(G3) state that such morphisms should include the identity morphisms, be invertible and that compatible morphisms are composable. The inclusion axiom (Inc) simply states that if two cells are equivalent to magnitude λ_2, via morphism Φ, then they are still equivalent via Φ to any smaller magnitude $\lambda_1 \leq \lambda_2$. In short, for $\lambda_1 \leq \lambda_2$ we have an inclusion of groupoids $\mathfrak{T}^{\lambda_2} \hookrightarrow \mathfrak{T}^{\lambda_1}$. The final two axioms (Res) and ($\widehat{\text{Res}}$) of restriction and corestriction establish a coherence between the cellular structure of $\mathcal{T}$ and the restrictions between the various morphisms of $\mathfrak{T}$.

The results of this section may be generalised to systems of internal symmetries with only minor modifications to the definitions and proofs, although these proofs are most efficiently given in a combinatorial setting. We may derive direct analogues of the following for SISs:

(1) PE or ePE (co)homology.

(2) The tiling space inverse limit presentation $\Omega_{\mathfrak{T}} = \varprojlim_{\lambda \in \Lambda}(\Gamma_\lambda, \pi_{\lambda,\mu})$ where the approximants Γ_λ are CW complexes given by identifying cells of $\mathcal{T}$ which are equivalent to magnitude λ.

(3) Theorem 3.1, that the Čech cohomology of $\Omega_{\mathfrak{T}}$ agrees with the PE cohomology of $\mathfrak{T}$ when the cells of $\mathfrak{T}$ (for sufficiently large magnitudes) have trivial isotropy.

(4) Lemma 3.2, that we have invariance of the PE (co)homology of $\mathfrak{T}$ over G coefficients under barycentric refinement whenever G (for sufficiently large magnitudes) has divisibility by the order of isotropy of cells in $\mathfrak{T}$.

(5) Theorem 3.3, ePE Poincaré duality $H^\bullet(\mathfrak{T}; G) \cong H_{d-\bullet}(\mathfrak{T}; G)$. For this to hold, we need firstly that the ambient space of $\mathfrak{T}$ is a G–oriented d–manifold (or even just homology G–manifold) with pattern-equivariant fundamental class Γ. We also require the analogous condition on the divisibility of the coefficient ring; that is, there exists some $\lambda \in \Lambda$ for which G has division by the order of isotropy groups $\mathfrak{T}^\lambda_{a,a}$ at every cell a of $\mathfrak{T}$.

Example 3.17 The initial insight in Penrose's discovery of his famous tilings was that "a regular pentagon can be subdivided into six smaller ones, leaving only five slim triangular gaps" [33]. Bowers and Stephenson [11] took a similar subdivision but chose, instead of methodically filling in the slim triangular gaps, to simply remove them by identifying edges of the pentagons. Of course, this cannot be achieved in Euclidean space with regular pentagons; the result is a *combinatorial* substitution. One may produce, in an analogous way to in the Euclidean setting of [1; 16], limiting combinatorial tilings. Declaring that each 2–cell of such a combinatorial tiling should metrically correspond to a regular pentagon, the resulting tilings are of spaces which are homeomorphic, but not isometric to Euclidean 2–space.

There is no notion of translation on the ambient spaces of these tilings, but there is of orientation. Let $\mathcal{T}$ be a Bowers–Stephenson pentagonal tiling, which we consider here simply as a regular cell complex with a choice of identification of each 2–cell with the regular pentagon. We may define a corresponding SIS $\mathfrak{T}^0$ as follows. Given cells a, $b \in \mathcal{T}$ and $i \in \mathbb{N}$, consider the collection of maps taking the i–corona of a to the i–corona of b, preserving orientation and distances on each pentagonal tile (such a map is, of course, determined combinatorially by how it acts on cells). We let $(\mathfrak{T}^0)^i_{a,b}$ be the set of such maps restricted to the stars of a and b.

We may construct from this data an inverse limit of approximants and associated inverse limit space $\Omega^0_{\mathfrak{T}}$, analogously to the Euclidean setting. The points of $\Omega^0_{\mathfrak{T}}$ may be identified with pointed Bowers–Stephenson pentagonal tilings, two being "close" if they agree via an orientation-preserving isometry on large patches about their origins, up to a small perturbation of their origins. We have analogues of the ePE (co)homology groups, and of Theorems 3.3 and 3.7. We may not identify the (integer) ePE cohomology with the Čech cohomology of $\Omega^0_{\mathfrak{T}}$, since the cells have nontrivial isotropy, but we may after replacing the tiling with its barycentric subdivision. These cohomology groups are Poincaré dual to a modification of the ePE homology of $\mathfrak{T}^0_\Delta$, defined in an analogous fashion to the modified chain complexes $C^\dagger_\bullet(\mathfrak{T}^0_\Delta)$ of Definition 3.6.

Our method of computation of the ePE homology groups in Section 4 easily generalises to examples such as this, we find that

$$\begin{aligned}\check{H}^1(\Omega^0_{\mathfrak{T}}) &\cong H^\dagger_1(\mathfrak{T}^0_\Delta) \cong H_1(\mathfrak{T}^0_\Delta) \cong 0,\\ \check{H}^2(\Omega^0_{\mathfrak{T}}) &\cong H^\dagger_0(\mathfrak{T}^0_\Delta) \cong \mathbb{Z} \oplus \mathbb{Z}\left[\tfrac{1}{6}\right],\\ H_0(\mathfrak{T}^0_\Delta) &\cong \mathbb{Z} \oplus \mathbb{Z}\left[\tfrac{1}{6}\right].\end{aligned}$$

Example 3.18 In this example we shall see how more general objects are also naturally captured in this framework. The magnitude poset will be $\mathbb{N}$, but endowed with the partial ordering of $m \le n$ if m divides n; note that $(\mathbb{N}, |)$ is a directed set. Let $\mathcal{T}$ be the standard cellular decomposition of $\mathbb{R}^d$ associated to the tiling of unit cubes with vertices at the lattice points $\mathbb{Z}^d$. If cells $a, b \in \mathcal{T}$ are equal up to a translation in $n\mathbb{Z}^d$, then we let $(\mathfrak{T}^1)^n_{a,b}$ consist of the single map given by the restriction of this translation between the stars of a and b. Otherwise, we set $(\mathfrak{T}^1)^n_{a,b} = \varnothing$.

It is easy to check that $\mathfrak{T}^1$ thus defined satisfies the conditions of an SIS. A (co)chain of $\mathcal{T}$ is PE with respect to $\mathfrak{T}^1$ if and only if it is invariant under translation by some full rank sublattice of $\mathbb{Z}^d$. We have trivial isotropy (everything is generated by translations), and the analogous theorems and constructions of the previous sections apply. For example, in dimension $d = 1$ we may calculate that

$$\begin{aligned}\check{H}^0(\Omega^1_{\mathfrak{T}}) &\cong H^0(\mathfrak{T}^1) \cong H_1(\mathfrak{T}^1) \cong \mathbb{Z},\\ \check{H}^1(\Omega^1_{\mathfrak{T}}) &\cong H^1(\mathfrak{T}^1) \cong H_0(\mathfrak{T}^1) \cong \mathbb{Q}.\end{aligned}$$

The first isomorphisms are given by the analogue of Theorem 2.1 and the second by the analogue of PE Poincaré duality of Theorem 2.2. The tiling space $\Omega^1_{\mathfrak{T}}$ is homeomorphic — in a fashion entirely analogous to the Gähler construction — to the inverse limit

$$\Omega^1_{\mathfrak{T}} = \varprojlim(S^1, \pi_{m,n}),$$

where, for $m \mid n$, the map $\pi_{m,n}$ is the standard degree n/m covering map of S^1. The degree one PE homology group $H_1(\mathfrak{T}^1)$ is generated by a Borel–Moore fundamental class for $\mathbb{R}^1$ and the homology class $p/q \in H_0(\mathfrak{T}^1) \cong \mathbb{Q}$ is represented by, for example, the Borel–Moore 0–chain which assigns value p to each of the vertices of $q\mathbb{Z}$. Note that the sequence $n_i := i!$ is linearly ordered and cofinal in $(\mathbb{N}, \mid)$, so the tiling space could instead be expressed as

$$\Omega^1_{\mathfrak{T}} = \varprojlim(S^1 \xleftarrow{\times 2} S^1 \xleftarrow{\times 3} S^1 \xleftarrow{\times 4} S^1 \xleftarrow{\times 5} \cdots).$$

We may restrict the construction of $\mathfrak{T}^1$ above to the linearly ordered subset of magnitudes $\{2^n \mid n \in \mathbb{N}_0\}$. In this case, a (co)chain is PE if and only if it invariant under translation by $2^n\mathbb{Z}^d$ for some $n \in \mathbb{N}_0$. For $d = 1$, the corresponding tiling space is the dyadic solenoid

$$\Omega^1_{\mathfrak{T}} = \varprojlim(S^1 \xleftarrow{\times 2} S^1 \xleftarrow{\times 2} S^1 \xleftarrow{\times 2} S^1 \xleftarrow{\times 2} \cdots).$$

Whilst these examples are not given by tilings of finite local complexity, they are close in spirit. Indeed, one may think of the dyadic example above as a hierarchical tiling. The tiles (unit cubes) may be grouped into supertiles (cubes of side-length 2), which may be grouped into level n supertiles (cubes of side-length 2^n). However, the groupings of tiles cannot be determined using local geometric information in the underlying tiling; one says that the substitution corresponding to this example is *nonrecognisable*. The resulting system of internal symmetries is what one would get if the tiling were capable of deducing such an imposed hierarchy from local geometric information. For an alternative derivation of the dyadic solenoid as the tiling space of an infinite local complexity tiling, see [34].

4 PE homology of hierarchical tilings

4.1 Substitution tilings and their hulls

The two main approaches to producing interesting examples of aperiodic tilings, such as the Penrose tilings, are through the cut-and-project method (see [15]) and through tiling substitutions (see [1]). A substitution rule consists of a finite collection of *prototiles* of $\mathbb{R}^d$, a rule for subdividing them and an expanding dilation which, when applied to the subdivided prototiles, defines patches of translates of the original prototiles. By iterating the substitution and inflating, one produces successively larger patches. A tiling is said to be *admitted by the substitution rule* if every finite patch of it is a subpatch of a translate of some iteratively substituted prototile. In analogy with symbolic dynamics, one may think of the substitution rule as generating the allowed language for a family of tilings.

Under certain conditions on the substitution rule, tilings admitted by it exist and, in addition, for each such tiling $\mathfrak{T}_0$ there is a *supertiling* $\mathfrak{T}_1$, based on inflated versions of the prototiles, which subdivides to $\mathfrak{T}_0$ and is itself a (dilation of an) admitted tiling. So $\mathfrak{T}_0$ has a hierarchical structure: there is an infinite list of substitution tilings $\mathfrak{T}_0, \mathfrak{T}_1, \mathfrak{T}_2, \ldots$ of progressively larger tiles for which the tiles of $\mathfrak{T}_n$ may be grouped to form the tiles of $\mathfrak{T}_{n+1}$, with the substitution decomposing $\mathfrak{T}_{n+1}$ to $\mathfrak{T}_n$. For fuller details on the definition of substitution rules and their tilings, we refer the reader to [1; 16; 38].

To compute the Čech cohomology of a substitution tiling space Ω^1, one typically constructs a finite CW complex Γ along with a self-map f of Γ for which

$$\Omega^1 \cong \varprojlim(\Gamma \xleftarrow{f} \Gamma \xleftarrow{f} \Gamma \xleftarrow{f} \cdots).$$

The CW complex Γ may be defined in terms of the short-range combinatorics of the patches of the substitution tilings, and the map f by the action of substitution. This makes the Čech cohomology of a substitution tiling space computable.

Anderson and Putnam showed that when the substitution rule has a property known as *forcing the border*, one may take Γ as, what is now known as, the *AP complex* [1], which is precisely the level zero Gähler complex Γ_0^1 (see Section 1.3). If the substitution fails to force the border, one may work with the *collared AP complex* Γ_1^1 instead. Whilst conceptually simple, passing to the collared complex can be computationally demanding; even for relatively simple substitution rules, the number of collared tiles can be unwieldy. A powerful alternative approach was developed by Barge, Diamond, Hunton and Sadun [5], which typically results in much smaller cochain complexes than for the collared AP complex. One constructs a CW complex K_ϵ by, instead of collaring tiles, collaring *points* of the ambient space of the tiling. A point of K_ϵ is a description of how to tile an ϵ–neighbourhood of the origin. The self map on the complex defined by the substitution is not cellular, but for small ϵ is homotopic to a cellular map in a canonical way, which is sufficient for cohomology computations. Another advantage of the BDHS approach is that the resulting inverse system possesses natural stratifications, which are useful in breaking down the calculations to something more tractable.

4.2 Overview of PE homology approach to calculation

We shall present below a method of calculation of the PE homology of a substitution tiling. There are various motivations for introducing it. Firstly, as we shall see, the "approximant complexes" and "connecting map" of the method are constructed from the combinatorial information of the substitution in a very direct way, which makes the approach highly amenable to computer implementation. An early implementation has been coded by the author, in collaboration with James Cranch, in the programming language Haskell, and at present is applicable to any polytopal substitution tiling of arbitrary dimension (although, aside from cubical substitutions, efficiently communicating the combinatorial information of a tiling substitution to the program is still problematic, an issue shared with any machine computation of the cohomology of substitution tilings). The approximant chain complexes of the method are much smaller than for the collared AP method; the combinatorial information required from the patches is the same as that for the BDHS approach.

Another reason for introducing this method is that it may be used to find the *ePE* homology of a substitution tiling (see Example 4.9), which, as we have seen, yields different information to the cohomology calculations. Furthermore, the method provides explicit generators in terms of pattern-equivariant chains. The result for the ePE homology of the Penrose tiling, along with precise descriptions of the generators of the ePE homology, is essential in [42].

In the translational setting, our approach can be seen to produce isomorphic direct limit diagrams to the approximant cohomologies of the BDHS method [5], at least after collaring points of the tiling for the BDHS approximants in a way compatible with the combinatorics of the tiling (although the method that we shall describe provides a more combinatorial way of determining this diagram). The argument proceeds via a stratification of the BDHS approximants (although one which is not preserved by the connecting maps) or by applying a certain homotopy to the projective system of BDHS approximants. However, the full details of this seem to be technical, at least in general dimensions, and we avoid providing them here. In any case, the approximant complexes and connecting maps between the approximant homologies that we shall define are most naturally described in the PE homology framework. The approximant complexes used here are precisely the duals of those used by Gonçalves [19] in his computation of the K–theory of the C^*–algebra associated to the stable relation of a one or two-dimensional substitution tiling. This K–theory appears to be dual in a certain sense to the K–theory of the hull (that is, of the unstable relation). The fact that our technique — which involves the duals of the approximant complexes of [19] — calculates the (regraded) Čech cohomology of the hull of a substitution tiling seems to confirm this duality. A complete confirmation of the relationship would, however, require consideration of the connecting maps of each method of calculation.

4.3 The method of computation

4.3.1 The approximant complex We shall assume throughout that $\mathfrak{T} = \mathfrak{T}_0$ is a cellular FLC tiling admitted by a primitive, recognisable, polytopal substitution rule ω

with inflation constant $\lambda > 1$. We refer the reader to [1] for the notion of a primitive polytopal substitution. We note that many of these assumptions may be relaxed substantially; instead of letting that detain us here, we shall discuss various generalisations in Section 4.4.

Being a *recognisable* substitution means that for any tiling $\mathfrak{T}_0$ admitted by ω there exists a unique FLC tiling $\mathfrak{T}_1$, based on prototiles which are λ–inflations of the original prototiles, for which

- the rescaled tiling $\lambda^{-1}(\mathfrak{T}_1)$ is also admitted by ω;
- $\omega(\mathfrak{T}_1) = \mathfrak{T}_0$;
- $\mathfrak{T}_1$ is locally derivable from $\mathfrak{T}_0$.

The first item simply states that the *supertiling* $\mathfrak{T}_1$ is itself an inflate of an admitted tiling. The second says that the substitution rule decomposes the supertiles of $\mathfrak{T}_1$ to the tiles of $\mathfrak{T}_0$. Thinking upside-down, one may group the tiles of $\mathfrak{T}_0$ so as to form the supertiling $\mathfrak{T}_1$. The third item states that this grouping may be performed using only local information. Since $\omega(\mathfrak{T}_1) = \mathfrak{T}_0$ implies that $\mathfrak{T}_0$ is locally derivable from $\mathfrak{T}_1$, the two are MLD. This process may be repeated, yielding a hierarchy of tilings $\{\mathfrak{T}_n\}_{n \in \mathbb{N}_0}$. Each $\mathfrak{T}_n$ carries a polytopal decomposition $\mathcal{T}_n$, with $\mathcal{T}_i$ refining $\mathcal{T}_j$ for $i \leq j$.

Given a k–cell c of $\mathcal{T}_0$, we name the pair of c along with the set of tiles properly containing c the *star* of c. Henceforth, the translation class of such a star (where translations preserve labels, if the cells are labelled) will be simply called a *star*, or a *k–star* if we wish to specify the dimension of the *central cell c*. The first step of the calculation is to enumerate the set of stars. This may be efficiently performed algorithmically as follows:

(1) Begin with the set of d–stars, consisting of the prototiles of $\mathfrak{T}$. Put these stars into sets S_0^{new} and S_0^{acc}.

(2) Suppose that $S_n^{\mathrm{new}} \neq \varnothing$ and S_n^{acc} have been constructed. Substitute each star of S_n^{new} and find all stars whose centres are contained in the substituted central (open) cell of the original star. All such stars which are not already elements of S_n^{acc} define the set S_{n+1}^{new}, and are added to S_n^{acc} to define S_{n+1}^{acc}.

(3) If S_n^{new} and S_n^{acc} have been constructed but $S_n^{\mathrm{new}} = \varnothing$, then the process is terminated and the full list of stars is $S := S_n^{\mathrm{acc}}$.

The stars, and the incidences between them, define our approximant chain complex:

Definition 4.1 We define the *approximant chain complex*

$$C_\bullet^{(0)}(\mathfrak{T}^1) := 0 \leftarrow C_0^{(0)}(\mathfrak{T}^1) \xleftarrow{\partial_1} C_1^{(0)}(\mathfrak{T}^1) \xleftarrow{\partial_2} \cdots \xleftarrow{\partial_d} C_d^{(0)}(\mathfrak{T}^1) \leftarrow 0$$

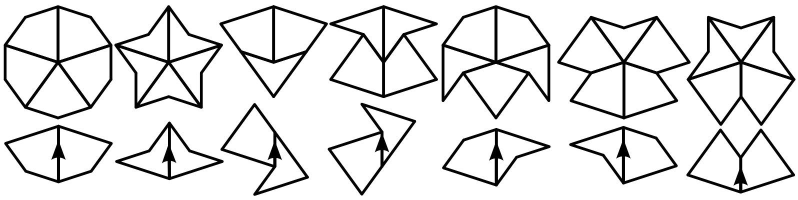

Figure 4: The seven rigid equivalence classes of 0–stars, named, in Conway's notation, "sun", "star", "ace", "deuce", "jack", "queen" and "king", in this order, followed by the seven 1–stars E1–E7

as follows. The degree k chain group $C_k^{(0)}(\mathfrak{T}^1)$ is freely generated by the k–stars; $C_k^{(0)}(\mathfrak{T}^1) \cong \mathbb{Z}^n$, where n is the number k–stars. The star of a $(k-1)$–cell c' determines the star of any k–cell c containing c'. Orienting the central cell of each star, the boundary maps ∂_k are induced from the standard cellular boundary maps by

$$\partial_k(s) = \sum_{(k-1)\text{–stars } s'} [s', s] \cdot s'$$

and extending linearly. Here, $[s', s]$ is the incidence number between the $(k-1)$–star s' and k–star s. It is defined, for fixed s' and s, to be the sum of incidence numbers $[c', c]$ where c' is the central $(k-1)$–cell of s' and c is a k–cell of s' whose star is s. The homology of the approximant chain complexes is the *approximant homology* $H_\bullet^{(0)}(\mathfrak{T}^1)$.

Note that since we are considering translation classes of stars, it may be that there are multiple occurrences of a k–star s in a $(k-1)$–star s'. An instructive perspective on the definition of the approximant chain complex is to identify a generator star s with the PE indicator chain $\mathbb{1}(s) \in C_k(\mathfrak{T}^1)$, the k–chain given by the (infinite) sum of k–cells in $\mathfrak{T} = \mathfrak{T}_0$ which are the centres of s in the tiling. With this identification, the boundary maps of the approximant complex correspond to the standard cellular boundary maps of $C_\bullet(\mathfrak{T}^1)$ defined in the ambient tiling. That is, $C_\bullet^{(0)}(\mathfrak{T}^1)$ is the subchain complex of the PE complex $C_\bullet(\mathfrak{T}^1)$ consisting of those chains which, at any given cell c, depend only on the local patch of tiles of $\mathfrak{T}$ properly containing c.

Example 4.2 In the one-dimensional case, we may identify our tiling with a bi-infinite sequence $s \in \mathcal{A}^{\mathbb{Z}}$ over a finite alphabet $\mathcal{A}$, and the substitution rule with a map $\omega\colon \mathcal{A} \to \mathcal{A}^*$ from the alphabet $\mathcal{A}$ to the set of nonempty words in $\mathcal{A}$. A 0–star then corresponds to a two-letter word from $\mathcal{A}^2$ which appears in s; let us denote the set of such *admissible* two-letter words by $\mathcal{A}_\omega^2$. A 1–star is simply a tile type, an element of $\mathcal{A}$. So the approximant complex is given by

$$0 \leftarrow \mathbb{Z}^m \xleftarrow{\partial_1} \mathbb{Z}^n \leftarrow 0,$$

where $m = \#\mathcal{A}^2_\omega$ and $n = \#\mathcal{A}$. The boundary map ∂_1 is defined on $a \in \mathcal{A}$ as the formal sum of admissible two-letter words whose first letter is a minus those whose second letter is a.

Example 4.3 One may easily verify that, up to rigid motion, there are seven distinct ways for a patch of Penrose kite and dart tiles to meet at a vertex, and seven ways for them to meet at an edge. These stars are given in Figure 4. Any rotate of such a patch by some $\frac{2\pi k}{10}$ appears in a Penrose kite and dart tiling, so there are 54 0–stars (the "sun" and "star" vertices are preserved by rotation by $\frac{2\pi}{5}$), there are 70 1–stars and there are 20 2–stars, corresponding to the rotates of the kite and dart tiles. So the approximant chain complex is of the form

$$C^{(0)}_\bullet(\mathfrak{T}^1) = 0 \leftarrow \mathbb{Z}^{54} \xleftarrow{\partial_1} \mathbb{Z}^{70} \xleftarrow{\partial_2} \mathbb{Z}^{20} \leftarrow 0.$$

The boundary maps have a simple description in terms of the standard cellular boundary maps. For example, the boundary of the indicator chain of the E1 edge is given by

$$\partial_1(\mathbb{1}(r^0\,\mathrm{E1})) = \mathbb{1}(r^1\,\mathrm{sun}) + \mathbb{1}(r^5\,\mathrm{jack}) - \mathbb{1}(r^0\,\mathrm{ace}) - \mathbb{1}(r^5\,\mathrm{queen});$$

the head of an E1 edge is always a "sun" or "jack" vertex, and the tail is always an "ace" or "queen". The notation r^k above indicates that the named patch has been rotated by $\frac{2\pi k}{10}$ relative to its depiction in Figure 4. We compute the approximant homology groups as

$$H^{(0)}_k(\mathfrak{T}^1) \cong \begin{cases} \mathbb{Z}^8 & \text{for } k = 0, \\ \mathbb{Z}^5 & \text{for } k = 1, \\ \mathbb{Z} & \text{for } k = 2. \end{cases}$$

Representative cycles for the generators of these approximant homology groups are as in Example 2.3.

4.3.2 The connecting map In the case of the Penrose kite and dart substitution, the inclusion of the approximant chain complex into the full PE chain complex is a quasi-isomorphism. This is not true in general. We shall now describe how one constructs a homomorphism

$$f\colon H^{(0)}_\bullet(\mathfrak{T}^1) \to H^{(0)}_\bullet(\mathfrak{T}^1)$$

from the approximant homology to itself, called the *connecting map*, for which the PE homology $H_\bullet(\mathfrak{T}^1)$ is isomorphic to the direct limit $\varinjlim(H^{(0)}_\bullet(\mathfrak{T}^1), f)$.

To construct f, we now need to consider the passage from the tiling $\mathfrak{T}_0$ to its supertiling $\mathfrak{T}_1$. By the explanation following Definition 4.1, we may identify $C^{(0)}_\bullet(\mathfrak{T}^1)$ with the subchain complex of PE chains which are determined at a cell c by the patch of tiles properly containing c. To define f, we firstly define an auxiliary chain complex $C_\bullet^{(1)}(\mathfrak{T}^1)$, still a subcomplex of $C_\bullet(\mathfrak{T}^1)$, which consists of those PE chains of $\mathfrak{T} = \mathfrak{T}_0$ which are determined at any given cell of $\mathcal{T}_0$ by only which *super*tiles of $\mathfrak{T}_1$ properly contain c. An alternative take on this is that one has a new set of generators which, in degree k, are given by k–cells of the substituted central cells of the original stars. The boundary maps of $C_\bullet^{(1)}(\mathfrak{T}^1)$ are induced from the standard cellular boundary maps, analogously to $C_\bullet^{(0)}(\mathfrak{T}^1)$. That is, for a translation class of oriented k–cell s from $\mathcal{T}_0$, labelled by the patch information of supertiles which contain it, we define $\partial_k(s)$ by identifying s with the PE indicator chain $\mathbb{1}(s) \in C_k(\mathfrak{T}^1)$, and then define $\partial_k(s)$ to be the element of $C_{k-1}^{(1)}(\mathfrak{T}^1)$ corresponding to $\partial(\mathbb{1}(s)) \in C_{k-1}(\mathfrak{T}^1)$.

There are two very natural maps from $C_\bullet^{(0)}(\mathfrak{T}^1)$ to $C_\bullet^{(1)}(\mathfrak{T}^1)$ which we shall use to construct f. Let s be some k–star, which we identify with the PE indicator chain $\mathbb{1}(s) \in C_k(\mathfrak{T}^1)$. Since $\mathbb{1}(s)$ is determined at any cell by the patch of supertiles properly containing that cell, $\mathbb{1}(s)$ is also an element of $C_k^{(1)}(\mathfrak{T}^1)$. We define the chain map ι as this inclusion of chain complexes.

The combinatorics, and hence stars, of $\mathfrak{T}_0$ and $\mathfrak{T}_1$ are identical. Since we may identify the stars of each, we may associate any chain $\sigma \in C_k^{(0)}(\mathfrak{T}^1)$ with a chain $\sigma' \in C_k^{\mathrm{BM}}(\mathcal{T}_1)$ which assigns coefficients to k–cells of $\mathcal{T}_1$ based on their neighbourhood stars in $\mathfrak{T}_1$ identically to how σ does in $\mathfrak{T}_0$. The chain map q is given by identifying σ' with its representation on the finer subcomplex $\mathcal{T}_0$, induced by identifying the elementary chain of a k–cell c of $\mathcal{T}_1$ with the sum of k–cells of $\mathcal{T}_0$ contained in c, suitably oriented with respect to c. It is easily seen that $q(\sigma) \in C_k^{(1)}(\mathfrak{T}^1)$. The chain map q is in some sense simply a refinement. So as one may expect, q is a quasi-isomorphism; that is, the induced map on homology

$$q_*\colon H_\bullet^{(0)}(\mathfrak{T}^1) \to H_\bullet^{(1)}(\mathfrak{T}^1)$$

is an isomorphism; see Lemma 4.14.

Definition 4.4 We define the *connecting map* as

$$f := (q_*)^{-1} \circ \iota_*\colon H_\bullet^{(0)}(\mathfrak{T}^1) \to H_\bullet^{(0)}(\mathfrak{T}^1),$$

where the chain maps ι and q are defined as above.

Theorem 4.5 *There is a canonical isomorphism* $\varinjlim(H_\bullet^{(0)}(\mathfrak{T}^1), f) \cong H_\bullet(\mathfrak{T}^1)$.

By *canonical* here, we mean that the isomorphism is induced by a natural association of cycles of the direct limit with PE cycles of $C_\bullet(\mathfrak{T}^1)$; a chain at the n^{th} level of the direct limit corresponds to a PE chain which only depends cellwise on its immediate surroundings in the level n supertiling $\mathfrak{T}_n$. We delay the details of the proof to the final subsection.

We remark that the connecting map f is *not* canonically induced by a chain map from $C_\bullet^{(0)}(\mathfrak{T}^1)$ to itself; it is defined at the level of homology rather than of chains. Nonetheless, we may provide a geometrically intuitive picture of the action of f on the approximant cycles. Suppose that $\sigma \in C_k^{(0)}(\mathfrak{T}^1)$ is a cycle, which we may identify with a PE k–cycle of $\mathfrak{T}_0$ that only depends at any k–cell c on the patch of tiles containing c. Considered as a chain living inside the supertiling $\mathfrak{T}_1$ (but still a chain of the finer complex $\mathcal{T}_0$), σ may no longer be supported on the k–skeleton of $\mathfrak{T}_1$, but it is still determined cellwise by the local patches in $\mathfrak{T}_1$. Due to the homological properties of the cells, we may find a chain $\tau \in C_{k+1}^{(1)}(\mathfrak{T}^1)$ for which $\sigma' := \sigma + \partial(\tau)$ is supported on the k–skeleton of $\mathfrak{T}_1$. The combinatorics of $\mathfrak{T}_1$ are identical to that of $\mathfrak{T}_0$, so we may identify σ' with a cycle of $C_k^{(0)}(\mathfrak{T}^1)$. The homology class of this cycle is precisely $f([\sigma])$, and does not depend on the representative of $[\sigma]$ or $\tau \in C_{k+1}^{(1)}(\mathfrak{T}^1)$ that we chose. So, to define $f([\sigma])$, we may "push" σ to the k–skeleton of $\mathfrak{T}_1$ in a way which is locally determined in $\mathfrak{T}_1$, and identify the result with the analogous homology class from $H_k^{(0)}(\mathfrak{T}^1)$.

Example 4.6 Recall from Example 4.2 that in the one-dimensional case we may identify the generators of the degree zero approximant group with the admissible two-letter words of $\mathcal{A}_\omega^2$, and in degree one with the letters of $\mathcal{A}$. In degree one, as is always the case in top degree, we have that $H_1(\mathfrak{T}^1) \cong \mathbb{Z}$ is generated by a fundamental class. To compute the connecting map in degree zero, let ab be an admissible two-letter word, which represents an indicator chain of $C_0^{(0)}(\mathfrak{T}^1)$ (which, abusing notation, we shall also name ab here). Considered as a chain of the supertiling, ab lifts to the element $\iota(ab)$ which marks each xy vertex of the supertiling for which the last letter of $\omega(x)$ is a and the first letter of $\omega(y)$ is b, as well as vertices of the original tiling interior to the supertiles, corresponding, for a supertile with label x, to occurrences of ab in $\omega(x)$. There exists a 1–chain τ of the original CW decomposition of the tiling which only depends on ambient supertiles and for which $ab + \partial(\tau)$ is supported on the 0–skeleton of the supertiling. For example, we may choose τ so as to shift all ab vertices of $\mathfrak{T}_0$ contained in the interiors of supertiles to the right endpoints of these supertiles. So $f([ab])$ is represented by the chain

$$\sigma := \sum_{xy \in \mathcal{A}_\omega^2} (\omega_{\text{left}}^{ab}(xy)) \cdot xy,$$

where $\omega_{\text{left}}^{ab}(xy)$ is the number of occurrences of the word ab in the substituted word $\omega(x) \cdot \omega(y)$ with first letter of that occurrence of ab lying to the left of the $\cdot$ placeholder. In the notation of the definition of the connecting map f, we have that $q(\sigma) = \iota(ab) + \partial(\tau)$, so $f([ab]) := q_*^{-1}(\iota_*([ab])) = [\sigma]$. Since those cycles associated to indicator cochains of admissible two-letter words generate $H_0^{(0)}(\mathfrak{T}^1)$, the above rule determines the connecting map f.

Example 4.7 The approximant homology groups for the Penrose kite and dart tilings are free abelian of ranks 8, 5 and 1 in degrees 0, 1 and 2, respectively. The connecting map turns out to be an isomorphism in each degree. There is a subtlety with the Penrose kite and dart substitution [33], in that the substituted tiles have larger support than the inflated prototiles. In particular, the cell complex $\mathcal{T}_0$ of the tiling does not refine the complex $\mathcal{T}_1$ of the supertiling. This is only a minor inconvenience; one may work over a finer complex, corresponding to a Robinson triangle tiling, which refines both. The general procedure described above remains essentially the same, and we shall subdue this point in our discussion.

To demonstrate a typical application of the connecting map, we shall consider how it acts on the cycle $\rho' \in C_1^{(0)}(\mathfrak{T}^1)$ of Example 2.3, illustrated in red in Figure 2, which trails the bottoms of the dart tiles. We firstly consider ρ' as a 1–cycle of the complex $\mathcal{T}_0$ which only depends at any given 1–cell by those *supertiles* of $\mathfrak{T}_1$ properly containing it; formally we consider the chain $\iota(\rho') \in C_1^{(1)}(\mathfrak{T}^1)$. Let $\tau \in C_2^{(1)}(\mathfrak{T}^1)$ be the indicator 2–chain of the dart tiles of $\mathfrak{T}_0$; it is the blue chain of Figure 2 (of course, we in fact have that τ is a member of the subcomplex $C_2^{(0)}(\mathfrak{T}^1)$, that is, τ only depends on ambient *tiles*, rather than *super*tiles in this case). Then $\rho' + \partial(\tau)$ is the 1–cycle, illustrated in green in Figure 2, which runs along the 1–cells at the bottoms of the superdart tiles, but with the opposite corresponding orientation to ρ'. That is to say, $\iota(\rho') + \partial(\tau) = q(-\rho')$, so $f([\rho']) := q_*^{-1}(\iota_*([\rho'])) = -[\rho']$. More informally, we "push" the 1–cycle ρ' to the 1–skeleton of the supertiling by adding to it the boundary of τ, which is defined at any 2–cell by only which supertiles contain it, and identify the result with the corresponding homology class of $H_1^{(0)}(\mathfrak{T}^1)$.

4.4 Generalisations

There are several ways in which the method discussed above may be generalised, and conditions of the substitution rule which may easily be relaxed. For example, the primitivity condition of the substitution and the compatibility of the substitution with the cellular decomposition may be weakened. More significantly, the method may be modified to apply to mixed substitution systems, to compute the ePE homology groups and applies naturally to non-Euclidean hierarchical tilings. Instead of providing the full details of each generalisation, which is not our main focus here, we shall mostly give brief outlines of the changes that need to be made in each case; the adaptations needed to the proofs of the analogues of Theorem 4.5 are relatively straightforward in each case.

4.4.1 Mixed substitutions A *mixed* or *multi*-substitution system [18] is a family of substitutions acting upon the same prototile set. Loosely, whereas the language for admissible tilings of a substitution rule ω is given by iteratively applying ω to the prototiles, in a mixed system one builds the language by applying the family of substitutions to the prototiles in some chosen sequence.

Passing to the setting of mixed substitutions adds far more generality. For example, the Sturmian words associated to some irrational number α may be defined using a mixed substitution system, which will be purely substitutive if and only if α is a quadratic irrational. In contrast to the purely substitutive case, the family of one-dimensional mixed substitution tilings exhibit an uncountable number of distinct isomorphism classes of degree one Čech cohomology groups [36]. In a mixed substitution tiling the tiles group together to the supertiles, those into higher order supertiles, and so on, just as in the purely substitutive case, but now the rules connecting the various levels of the hierarchy are not constant. A general framework which captures this idea is laid out in [17].

Passing to mixed substitution systems adds some complications, since the local combinatorics of the tilings $\mathfrak{T}_n$ and the passage between them vary in n. However, the method as described easily generalises to such examples. Now one needs to compute the list of stars for each $\mathfrak{T}_n$ to find the approximant homology groups, and the connecting maps may vary at each level.

Example 4.8 (Arnoux–Rauzy sequences) The Arnoux–Rauzy words were originally introduced in [3] as a generalisation of Sturmian words. Let $k \in \mathbb{N}_{\geq 2}$. The Arnoux–Rauzy substitutions are defined over the alphabet $\mathcal{A}_k = \{1, 2, \dots, k\}$ and the k different substitutions ρ_i are given by $\rho_i(j) = ji$ for $i \neq j$ and $\rho_i(i) = i$. Fix an infinite sequence $(n_i)_i = (n_0, n_1, \dots) \in \mathcal{A}_k^{\mathbb{N}_0}$ for which each element of $\mathcal{A}_k$ occurs infinitely often. Then there exist bi-infinite *Arnoux–Rauzy words* for which every finite subword is contained in some translate of a "supertile" $\rho_{n_0} \circ \rho_{n_1} \circ \cdots \circ \rho_{n_l}(i)$. We may consider such a word as defining a tiling of labelled unit intervals of $\mathbb{R}^1$. The system is recognisable, so for such a tiling $\mathfrak{T}_0$, one may uniquely group the tiles to a tiling $\mathfrak{T}_1$ of tiles of labelled intervals for which the substitution ρ_{n_0} decomposes $\mathfrak{T}_1$ to $\mathfrak{T}_0$. The process may be repeated, leading to an infinite hierarchy of tilings $\mathfrak{T}_n$ for which the substitution ρ_{n_i} subdivides $\mathfrak{T}_{i+1}$ to $\mathfrak{T}_i$; the supertiles of these tilings become arbitrarily large as one passes up the hierarchy.

The two-letter words of $\mathfrak{T}_i$ are the elements of $\mathcal{A}_k^2$ with at least one occurrence of $n_i \in \mathcal{A}_k$. So the degree zero approximant homology at level i, based upon stars of supertiles of $\mathfrak{T}_i$, is isomorphic to $\mathbb{Z}^k$, freely generated by the indicator 0–chains of vertices of the form $n_i \cdot j$, where $j \in \mathcal{A}_k$ is arbitrary. A simple calculation shows that, with this choice of basis, the connecting map between level i and level $i+1$ is the unimodular matrix M_i given by the identity matrix but with a column of 1's down the n_i^{th} column, which, incidentally, is the incidence matrix of the substitution ρ_{n_i}.

So the degree one Čech cohomology of the tiling space of the Arnoux–Rauzy words associated to any given sequence $(n_i)_i \in \mathcal{A}_k^{\mathbb{N}_0}$ is

$$\check{H}^1(\Omega^1_{\mathfrak{T}}) \cong H^1(\mathfrak{T}^1) \cong H_0(\mathfrak{T}^1) \cong \varinjlim(\mathbb{Z}^k \xrightarrow{M_0} \mathbb{Z}^k \xrightarrow{M_1} \mathbb{Z}^k \xrightarrow{M_2} \cdots) \cong \mathbb{Z}^k.$$

It is interesting to note that the matrices above are related to continued fraction algorithms. For the $k=2$ case, the Arnoux–Rauzy words are precisely the Sturmian words. To an irrational α, the sequence $(n_i)_i$ is chosen according to the continued fraction algorithm for α (see [14, Section 3.2]) and the sequence of matrices M_i of the above direct limit determine the partial quotients of α. Whilst the isomorphism classes of the first Čech cohomology groups do not distinguish these tiling spaces, their order structure [32] is a rich invariant. Although we shall not give full details here, features such as the order structure of the cohomology groups are preserved by the method calculation described above, via Poincaré duality.

4.4.2 Euclidean pattern-equivariance The method is easily adjusted to compute the ePE homology groups, based on chains which are determined cellwise by their i–coronas, for sufficiently large i, up to *rigid motion* rather than just up to translation. The method proceeds as before, but where one took translation classes of stars one simply now takes stars up to rigid equivalence.

Nontrivial isotropy may cause issues in this setting. If any of the stars have self-symmetries which act nontrivially on the central cells, then one may only compute over suitably divisible coefficients.

Computations of this sort are of particular interest since they provide different results to the ePE cohomology calculations. However, it is possible to modify the method to compute the ePE cohomology (and thus the Čech cohomology of the space $\Omega^0_{\mathfrak{T}}$) using this method for two-dimensional tilings. The ideas follow naturally from the discussions of Section 3.3. One computes $H^{\dagger}_{\bullet}(\mathfrak{T}^0)$ by using a similar method but restricting to indicator chains on 0–stars which assign coefficients which are divisible by n to 0–stars with n–fold rotational symmetry.

Example 4.9 Instead of providing more detail on the general method, we now demonstrate it on the Penrose kite and dart tilings, which should provide sufficient detail to the general method. To begin calculation, we must firstly enumerate the list of stars up to *rigid motion*, as we have already done in Figure 4. Since there are seven 0–stars, seven 1–stars and two 2–stars (the kites and darts) up to rigid motion, the approximant chain complex for the ePE homology is given by

$$C^{(0)}_{\bullet}(\mathfrak{T}^0) = 0 \leftarrow \mathbb{Z}^7 \xleftarrow{\partial_1} \mathbb{Z}^7 \xleftarrow{\partial_2} \mathbb{Z}^2 \leftarrow 0.$$

Again, the boundary maps are induced from the standard cellular boundary maps after identifying the generators of these chain groups with indicator chains of the tiling. For example,

$$\partial_1(\mathbb{1}(\mathrm{E1})) = \mathbb{1}(\mathrm{sun}) + \mathbb{1}(\mathrm{jack}) - \mathbb{1}(\mathrm{ace}) - \mathbb{1}(\mathrm{queen}).$$

We calculate the approximant homology groups as

$$H_k^{(0)}(\mathfrak{T}^0) \cong \begin{cases} \mathbb{Z}^2 \oplus (\mathbb{Z}/5) & \text{for } k=0, \\ \mathbb{Z} & \text{for } k=1, \\ \mathbb{Z} & \text{for } k=2. \end{cases}$$

The connecting map is defined essentially identically to the translational case, and we find that it is an isomorphism in each degree. So the approximant homology groups above are isomorphic to the ePE homology groups of the Penrose kite and dart tilings, and the generators of the approximant homology groups may be taken as generators of the ePE homology.

The most interesting feature here is the 5–torsion in degree zero, which is not found in the degree two ePE cohomology, breaking Poincaré duality over integral coefficients. It is generated by the element $t = \mathbb{1}(\mathrm{sun}) + \mathbb{1}(\mathrm{star}) - \mathbb{1}(\mathrm{queen})$, illustrated in Figure 3, where one can see that $5t$ is nullhomologous via the boundary of

$$-\mathbb{1}(\mathrm{E1}) + \mathbb{1}(\mathrm{E2}) - \mathbb{1}(\mathrm{E4}) - 2\cdot\mathbb{1}(\mathrm{E7}).$$

To calculate the ePE cohomology, or equivalently (according to Theorem 3.1) the Čech cohomology $\check{H}^\bullet(\Omega^0_{\mathfrak{T}})$, we may compute the Poincaré dual groups $H^\dagger_{2-\bullet}(\mathfrak{T}^0)$ (see Section 3.3). The method is similar to before, but now one replaces the approximant complex with the subcomplex

$$0 \leftarrow 5\mathbb{Z} \oplus 5\mathbb{Z} \oplus \mathbb{Z}^5 \xleftarrow{\partial_1} \mathbb{Z}^7 \xleftarrow{\partial_2} \mathbb{Z}^2 \leftarrow 0,$$

where the degree zero chain group is the subgroup of $C_0^{(0)}(\mathfrak{T}^0)$ which restricts the coefficients on the sun and star vertices to multiples of 5, since these vertices have 5–fold rotational symmetry and the other vertices have trivial rotational symmetry. We calculate the modified approximant homology groups in degree zero as $\mathbb{Z}^2$ and the connecting maps as isomorphisms, in agreement with the chain of isomorphisms

$$\check{H}^2(\Omega^0_{\mathfrak{T}}) \cong H^2(\mathfrak{T}^0) \cong H_0^\dagger(\mathfrak{T}^0) \cong \varinjlim(\mathbb{Z}^2, f) \cong \mathbb{Z}^2.$$

4.4.3 Non-Euclidean tilings The final generalisation which we shall discuss is to non-Euclidean tilings. Again, we shall use the pentagonal tilings of Bowers and Stephenson as our running example. There is no natural action of translation for these tilings but, as discussed in Example 3.17, there is a natural analogue of the ePE (co)homology groups. The method above, with essentially no modifications, may be used to compute them:

Example 4.10 One begins by listing the rigid equivalence classes of stars from the tiling. For a Bowers–Stephenson pentagonal tiling, there are two 0–stars, corresponding to those vertices meeting three tiles and those meeting four, there is one 1–star and one 2–star. However, the 1–star has rotational symmetry which reverses the orientation of its central 1–cell. Since we have nontrivial cell isotropy, we may only compute over suitably divisible coefficients. Over $\mathbb{Q}$ coefficients, the approximant complex is given by

$$C_\bullet^{(0)}(\mathfrak{T}^0) = 0 \leftarrow \mathbb{Q}^2 \xleftarrow{\partial_1} 0 \xleftarrow{\partial_2} \mathbb{Q} \leftarrow 0.$$

It follows that the approximant homology over $\mathbb{Q}$ coefficients is $H_k^{(0)}(\mathfrak{T}^0;\mathbb{Q}) \cong \mathbb{Q}^2,\ 0,\ \mathbb{Q}$ for $k = 0,\ 1,\ 2$, respectively. The connecting map has the analogous definition to the Euclidean case, and we find it to be an isomorphism in each degree.

To compute homology over integral coefficients, we pass to the barycentric subdivision. Now we have four 0–stars: two of them corresponding to the two 0–stars of the original tiling, one corresponding to the barycentre of each edge and one corresponding to the barycentre of each pentagon. There are three 1–stars and two 2–stars. So the approximant chain complexes over $\mathbb{Z}$ coefficients are

$$C_\bullet^{(0)}(\mathfrak{T}^0_\Delta) = 0 \leftarrow \mathbb{Z}^4 \xleftarrow{\partial_1} \mathbb{Z}^3 \xleftarrow{\partial_2} \mathbb{Z}^2 \leftarrow 0.$$

One computes $H_0^{(0)}(\mathfrak{T}^0_\Delta) \cong \mathbb{Z}^2$, $H_1^{(0)}(\mathfrak{T}^0_\Delta) \cong 0$ and $H_2^{(0)}(\mathfrak{T}_\Delta) \cong \mathbb{Z}$. So $H_1(\mathfrak{T}^0_\Delta) \cong 0$, and of course $H_2(\mathfrak{T}^0_\Delta) \cong \mathbb{Z}$ is generated by a fundamental class. One may calculate the connecting map in degree zero as having eigenvectors which span $\mathbb{Z}^2$ and have eigenvalues 1 and 6, so $H_0(\mathfrak{T}^0_\Delta) \cong \mathbb{Z} \oplus \mathbb{Z}[\frac{1}{6}]$.

To compute the analogue of the ePE cohomology of $\mathfrak{T}_\Delta$ (and hence the Čech cohomology of the associated tiling space $\Omega^0_{\mathfrak{T}}$), one may calculate the modified groups $H_\bullet^\dagger(\mathfrak{T}^0_\Delta)$ and implement Poincaré duality. At the approximant stage, this amounts to using instead the chain complexes

$$0 \leftarrow 2\mathbb{Z} \oplus 3\mathbb{Z} \oplus 4\mathbb{Z} \oplus 5\mathbb{Z} \xleftarrow{\partial_1} \mathbb{Z}^3 \xleftarrow{\partial_2} \mathbb{Z}^2 \leftarrow 0,$$

since the 0–stars of $\mathfrak{T}_\Delta$ possess isotropy of orders 2, 3, 4 and 5. After computing the connecting maps and corresponding direct limits, we find that

$$\begin{aligned}
\check{H}^0(\Omega^0_{\mathfrak{T}}) &\cong H^0(\mathfrak{T}^0_\Delta) \cong H_2^\dagger(\mathfrak{T}^0_\Delta) \cong \mathbb{Z},\\
\check{H}^1(\Omega^0_{\mathfrak{T}}) &\cong H^1(\mathfrak{T}^0_\Delta) \cong H_1^\dagger(\mathfrak{T}^0_\Delta) \cong 0,\\
\check{H}^2(\Omega^0_{\mathfrak{T}}) &\cong H^2(\mathfrak{T}^0_\Delta) \cong H_0^\dagger(\mathfrak{T}^0_\Delta) \cong \mathbb{Z} \oplus \mathbb{Z}[\tfrac{1}{6}].
\end{aligned}$$

It should be remarked that the above calculation may be performed quite painlessly by hand, the combinatorial information required being surprisingly manageable, in spite of the substitution rule not forcing the border.

4.5 Proof of Theorem 4.5

Definition 4.11 For a cellular Borel–Moore k–chain $\sigma \in C_k^{\mathrm{BM}}(\mathcal{T}_0)$, write $\sigma \in C_k^{(n)}(\mathfrak{T}^1)$ to mean that σ is determined at any k–cell c of $\mathcal{T}_0$ by the immediate surroundings of c in the level n supertiling $\mathfrak{T}_n$. More precisely, whenever there is a translation mapping k–cell a to b in $\mathcal{T}_0$, and also the patch of tiles of $\mathfrak{T}_n$ containing a to the corresponding patch at b, the k–cells a and b have the same coefficient in σ. We say that σ is *hierarchical* if $\sigma \in C_k^{(n)}(\mathfrak{T}^1)$ for some $n \in \mathbb{N}_0$ and write $C_k^{(\infty)}(\mathfrak{T}^1)$ for the collection of all hierarchical k–chains.

It is not hard to see that for $\sigma \in C_k^{(n)}(\mathfrak{T}^1)$ we also have that $\partial(\sigma) \in C_{k-1}^{(n)}(\mathfrak{T}^1)$, so we have chain complexes $C_\bullet^{(n)}(\mathfrak{T}^1)$ for every $n \in \mathbb{N}_0 \cup \{\infty\}$. Furthermore, since the level n supertiling $\mathfrak{T}_n$ is determined locally by the level $n+1$ supertiling $\mathfrak{T}_{n+1}$ via the substitution rule, for all $m \leq n$ we have an inclusion of chain complexes

$$\iota_{m,n}\colon C_\bullet^{(m)}(\mathfrak{T}^1) \hookrightarrow C_\bullet^{(n)}(\mathfrak{T}^1).$$

Importantly, every hierarchical chain is PE. Indeed, if a chain σ is determined locally at a cell c depending only on where c sits in the patch of supertiles containing c, then σ is also determined there by the i–corona of c in $\mathfrak{T}_0$, where i is chosen large enough so as to deduce the level n supertile decomposition at c (such an i exists by recognisability). We denote the inclusion of chain complexes by

$$\iota_\infty\colon C_\bullet^{(\infty)}(\mathfrak{T}^1) \hookrightarrow C_\bullet(\mathfrak{T}^1).$$

Unfortunately, it is not true that a PE chain must be hierarchical. Indeed, the i–corona of a cell c of $\mathfrak{T}_0$ need not be determined by the supertile containing c when c is interior, but close to the boundary of a supertile. On the other hand, if a k–cycle σ is PE, intuitively σ only depends on very local combinatorics of the tiles of $\mathfrak{T}_n$, relative to the sizes of the tiles of $\mathfrak{T}_n$, for n sufficiently large. One would expect that such a k–cycle could be perturbed, in a pattern-equivariant way, to a hierarchical k–cycle. More precisely, one would expect for there to exist a PE $(k+1)$–chain τ for which $\sigma + \partial(\tau)$ is supported on the k–skeleton and is still PE to a small radius relative to the sizes of the tiles, which would force $\sigma + \partial(\tau)$ to be hierarchical. To this end, we introduce the following technical lemma:

Lemma 4.12 *Let P be a d–dimensional polytope, with polytopal decomposition $\partial\mathcal{P}$ of its boundary and $\delta > 0$; there exist a function $h\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ satisfying the* *following: Let $\lambda > 1$ and the inflated polytope λP have a cellular decomposition $\mathcal{P}_\lambda$ whose cells have diameter bounded by δ. For any relative k–cycle σ of $\mathcal{P}_\lambda$ modulo the boundary $\partial\mathcal{P}_\lambda$, with $k < d$, there exists $\tau_\sigma \in C_{k+1}(\mathcal{P}_\lambda)$ for which $\sigma + \partial(\tau_\sigma)$ is supported on $\partial\mathcal{P}_\lambda$; we may choose such chains τ_σ such that, if σ_1 and σ_2 agree at distance greater than r from a subset of cells of $\lambda(\partial\mathcal{P})$, then τ_{σ_1} and τ_{σ_2} agree at distance greater than $h(r)$ from those cells.*

Proof Let $\sigma \in C_k(\mathcal{P}_\lambda)$ be a chain with $\partial(\sigma)$ supported on $\partial\mathcal{P}_\lambda$. By the homological properties of cells, there exists some τ_σ for which $\sigma + \partial(\tau_\sigma)$ is supported on $\partial\mathcal{P}_\lambda$. For any other choice of σ' which agrees with σ further than distance r from $\partial\mathcal{P}_\lambda$, we have that $\sigma' + \partial(\tau_\sigma)$ is supported on an $(r+\delta)$–neighbourhood of $\partial\mathcal{P}_\lambda$. So we may restrict attention to those relative cycles σ supported on an $(r+\delta)$–neighbourhood of $\partial\mathcal{P}_\lambda$.

So now suppose that σ is supported on an $r+\delta$ neighbourhood of $\partial\mathcal{P}_\lambda$, and let $c \in \partial\mathcal{P}$ be a $(d-1)$–cell of the boundary of $\mathcal{P}$. There exists a chain τ of $\mathcal{P}_\lambda$, supported on a $(Cr+\delta)$–neighbourhood of λc, for which $\sigma + \partial(\tau)$ is supported on $\partial\mathcal{P}_\lambda$ union a $(Cr+\delta)$–neighbourhood of the remaining $d-1$ cells of $\lambda(\partial\mathcal{P})$; here, C only depends on the polytope $\mathcal{P}$. Indeed, for sufficiently large λ, this statement would hold for $\mathcal{P}_\lambda$ replaced with a cellular decomposition of the disc of radius λ, the chain τ being induced by a radial deformation retraction of an $(r+\delta)$–neighbourhood of the boundary of the disc to its boundary. The result for the polytope P may be lifted from the case of a disc by the fact that polytopes are bi-Lipschitz equivalent to the standard unit disc. We may repeat this construction for the remaining $d-1$ cells. As a result, we construct a chain τ_σ for which $\sigma + \partial(\tau_\sigma)$ is supported on $\partial\mathcal{P}_\lambda$. For any relative cycle σ' which agrees with σ distance further than r_1 away from the $(d-2)$–skeleton of $\lambda\partial\mathcal{P}$, we have that $\sigma' + \partial(\tau_\sigma)$ is supported on some r_2–neighbourhood of the $(d-2)$–skeleton, where r_1 and r_2 depend only on r and not λ. We may now repeat this argument for those relative chains which are supported on neighbourhoods (of radius depending only on r) of the k–skeleton of $\lambda(\partial\mathcal{P})$ for successively smaller k, from which the result follows. □

Lemma 4.13 *The inclusion $\iota_\infty\colon C_\bullet^{(\infty)}(\mathfrak{T}^1) \hookrightarrow C_\bullet(\mathfrak{T}^1)$ is a quasi-isomorphism.*

Proof For a cellular Borel–Moore k–chain σ of $\mathcal{T}_0$, let us write that σ is $\mathrm{PE}_n(r)$ to mean that the values of σ at two k–cells a and b of $\mathcal{T}_0$ are equal whenever there are points $x \in a$ and $y \in b$ (as open cells) for which the patches $\mathfrak{T}_n[B_r + x]$ and $\mathfrak{T}_n[B_r + y]$ are equal up to a translation taking a to b. So a hierarchical chain is nothing other than a chain which is $\mathrm{PE}_n(0)$ for some $n \in \mathbb{N}_0$.

Let $\sigma \in C_k(\mathfrak{T}^1)$, so σ is $\mathrm{PE}_n(r)$ for all $n \in \mathbb{N}_0$ for some r. By the fact that the cells are polytopal, for sufficiently large n we have that the patch of tiles of $\mathfrak{T}_n$ within radius r of any cell c of $\mathcal{T}_0$ is determined by the star in $\mathfrak{T}_n$ of a cell of $\mathcal{T}_n$ within Cr of c. So the value of σ at any cell c of $\mathcal{T}_0$ contained in a supertile t of $\mathfrak{T}_n$ is determined by the star of a subcell of t in $\mathfrak{T}_n$ within radius Cr of c.

We may now appeal to the previous lemma. By FLC, there are only a finite number of translation classes of polytopal cells in $\mathfrak{T}_0$, and they all have diameter bounded by some $\delta > 0$. We may find a function $h\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ satisfying the result of the above lemma for each polytopal cell. Since σ only depends at subcells of supertiles on stars of cells within Cr of those cells, we may construct a chain τ for which $\sigma + \partial(\tau)$ is supported on the $(d-1)$–skeleton of $\mathfrak{T}_n$ and which only depends on stars of cells within $h(r)$ in $\mathfrak{T}_n$. So we may find a chain τ for which $\sigma + \partial(\tau)$ is supported on the $(d-1)$–skeleton of $\mathfrak{T}_n$ and which is $\mathrm{PE}_n(r')$, where r' only depends on r and not on n.

For sufficiently large n, we may repeat this process down the skeleta. As a result, we construct a chain τ for which $\sigma + \partial(\tau)$ is supported on the k–skeleton of $\mathfrak{T}_n$ (recall that k is the degree of σ) and is $\mathrm{PE}_n(r')$ with r' depending only on r. We claim that for sufficiently large n such a cycle must be hierarchical. Indeed, since σ is a cycle, it is determined across any k–cell of $\mathfrak{T}_n$ by its value on any k–cell of $\mathcal{T}_0$ contained in that cell. For sufficiently large n, for each k–cell c of $\mathcal{T}_n$ there is an interior k–cell of $\mathcal{T}_0$ for which the tiles of $\mathfrak{T}_n$ within radius r' about that cell are precisely those of the star of c in $\mathfrak{T}_n$. Since $\sigma + \partial(\tau)$ is $\mathrm{PE}_n(r')$, it follows that σ is determined at any k–cell of $\mathfrak{T}_n$ by the star of that cell. Hence, $\sigma + \partial(\tau) \in C_k^{(\infty)}(\mathfrak{T}^1)$, so we have shown that ι_* is surjective. Showing injectivity is analogous, applying the same procedure to boundaries in place of cycles. □

The lemma above allows us to work with the chain complex $C_\bullet^{(\infty)}(\mathfrak{T}^1)$ in computing the homology of $C_\bullet(\mathfrak{T}^1)$. The advantage to this is that $C_\bullet^{(\infty)}(\mathfrak{T}^1)$ possesses a natural filtration by the subchain complexes $C_\bullet^{(n)}(\mathfrak{T}^1)$. The following lemma shows that the *approximant homology groups* $H_\bullet^{(n)}(\mathfrak{T}^1)$ of these subcomplexes are all isomorphic, and in a way such that the induced inclusions $(\iota_{n,n+1})_*$ between them are the same:

Lemma 4.14 *For $n \in \mathbb{N}_0$ we have canonical quasi-isomorphisms*

$$q_n\colon C_\bullet^{(0)}(\mathfrak{T}^1) \hookrightarrow C_\bullet^{(n)}(\mathfrak{T}^1)$$

for which $(q_1)_*^{-1} \circ (\iota_{0,1})_* = (q_{n+1})_*^{-1} \circ (\iota_{n,n+1})_* \circ (q_n)_*$.

Proof Let $m \leq n$. Up to rescaling, the set of stars of $\mathfrak{T}_m$ and $\mathfrak{T}_n$ are identical. A generating element of $C_k^{(m)}(\mathfrak{T}^1)$ is an indicator chain $\mathbb{1}(c)$ of a translation class of k–cell c of $\mathcal{T}_0$, labelled by where it lies in the patch of tiles of $\mathfrak{T}_m$ containing it. We may define chain maps $q_{m,n}\colon C_\bullet^{(m)}(\mathfrak{T}^1) \to C_\bullet^{(n)}(\mathfrak{T}^1)$ by sending such an element to the sum of indicator chains $\mathbb{1}(c')$ of translation classes of k–cells c', suitably oriented, which are contained in the regions occupied by the analogous (inflated) locations of c in the tiling $\mathfrak{T}_n$.

Let us write q_n for $q_{0,n}$. A k–cycle $\sigma \in C_k^{(n)}(\mathfrak{T}^1)$ is in the image of q_n if and only if it is supported on the k–skeleton of $\mathfrak{T}_n$. We may now mimic the proof of Lemma 3.2, replacing ι with q_n, to show that each q_n is a quasi-isomorphism. The following identities are easily verified: $q_{j,k} \circ q_{i,j} = q_{i,k}$ and $\iota_{j,j+1} \circ q_j = q_{1,j+1} \circ \iota_{0,1}$ for $i \le j \le k$. It follows that $(q_{n+1})_*^{-1} \circ (\iota_{n,n+1})_* \circ (q_n)_* = (q_{n+1})_*^{-1} \circ (q_{1,n+1})_* \circ (\iota_{0,1})_* = (q_{n+1})_*^{-1} \circ ((q_{0,n+1})_* \circ (q_{0,1})_*^{-1}) \circ (\iota_{0,1})_* = (q_1)_*^{-1} \circ (\iota_{0,1})_*$. □

The chain maps $\iota_{0,1}$ and q_1 are denoted by ι and q, respectively, in the definition of the connecting map $f := (q_*)^{-1} \circ \iota_*$. We may now prove Theorem 4.5, that $H_\bullet(\mathfrak{T}^1)$ is canonically isomorphic to the direct limit $\varinjlim(H_\bullet^{(0)}(\mathfrak{T}^1), f)$. By Lemma 4.14, we have the following diagram:

$$\begin{tikzcd}
H_\bullet^{(0)}(\mathfrak{T}^1) \arrow[r, "f"] \arrow[d, "(q_0)_* = \mathrm{id}", "\cong"'] & H_\bullet^{(0)}(\mathfrak{T}^1) \arrow[r, "f"] \arrow[d, "(q_1)_*", "\cong"'] & H_\bullet^{(0)}(\mathfrak{T}^1) \arrow[r, "f"] \arrow[d, "(q_2)_*", "\cong"'] & H_\bullet^{(0)}(\mathfrak{T}^1) \arrow[r, "f"] \arrow[d, "(q_3)_*", "\cong"'] & \cdots \\
H_\bullet^{(0)}(\mathfrak{T}^1) \arrow[r, "(\iota_{0,1})_*"] & H_\bullet^{(1)}(\mathfrak{T}^1) \arrow[r, "(\iota_{1,2})_*"] & H_\bullet^{(2)}(\mathfrak{T}^1) \arrow[r, "(\iota_{2,3})_*"] & H_\bullet^{(3)}(\mathfrak{T}^1) \arrow[r, "(\iota_{3,4})_*"] & \cdots
\end{tikzcd}$$

This isomorphism of directed systems induces an isomorphism

$$\varinjlim(H_\bullet^{(0)}(\mathfrak{T}^1), f) \cong \varinjlim(H_\bullet^{(n)}(\mathfrak{T}^1), (\iota_{m,n})_*).$$

Since, by definition, $C_\bullet^{(\infty)}(\mathfrak{T}^1) = \bigcup_{n=0}^{\infty} C_\bullet^{(n)}(\mathfrak{T}^1)$, we may identify $C_\bullet^{(\infty)}(\mathfrak{T}^1)$ with the direct limit $\varinjlim(C_\bullet^{(n)}(\mathfrak{T}^1), \iota_{m,n})$. So by Lemma 4.13 we have the string of quasi-isomorphisms

$$\varinjlim(C_\bullet^{(n)}(\mathfrak{T}^1), \iota_{m,n}) \xrightarrow{\cong} C_\bullet^{(\infty)}(\mathfrak{T}^1) \xhookrightarrow{\iota_\infty} C_\bullet(\mathfrak{T}^1).$$

Applying homology and combining with the isomorphism of direct limits established above, we have that $\varinjlim(H_\bullet^{(0)}(\mathfrak{T}^1), f) \cong H_\bullet(\mathfrak{T}^1)$.

References

[1] **J E Anderson**, **I F Putnam**, *Topological invariants for substitution tilings and their associated C^*–algebras*, Ergodic Theory Dynam. Systems 18 (1998) 509–537 MR

[2] **P Arnoux**, **V Berthé**, **H Ei**, **S Ito**, *Tilings, quasicrystals, discrete planes, generalized substitutions, and multidimensional continued fractions*, from "Discrete models: combinatorics, computation, and geometry", Discrete Math. Theor. Comput. Sci. Proc. AA, Maison Inform. Math. Discrèt., Paris (2001) 59–78 MR

[3] **P Arnoux**, **G Rauzy**, *Représentation géométrique de suites de complexité $2n+1$*, Bull. Soc. Math. France 119 (1991) 199–215 MR

[4] **M Baake**, **M Schlottmann**, **P D Jarvis**, *Quasiperiodic tilings with tenfold symmetry and equivalence with respect to local derivability*, J. Phys. A 24 (1991) 4637–4654 MR

[5] **M Barge**, **B Diamond**, **J Hunton**, **L Sadun**, *Cohomology of substitution tiling spaces*, Ergodic Theory Dynam. Systems 30 (2010) 1607–1627 MR

[6] **J Bellissard**, **A Julien**, **J Savinien**, *Tiling groupoids and Bratteli diagrams*, Ann. Henri Poincaré 11 (2010) 69–99 MR

[7] **R Berger**, *The undecidability of the domino problem*, Mem. Amer. Math. Soc. No. 66, Amer. Math. Soc., Providence, RI (1966) MR

[8] **V Berthé**, **A Siegel**, *Tilings associated with beta-numeration and substitutions*, Integers 5 (2005) art. id. A2 MR

[9] **A Björner**, *Posets, regular CW complexes and Bruhat order*, European J. Combin. 5 (1984) 7–16 MR

[10] **R Bott**, **L W Tu**, *Differential forms in algebraic topology*, Graduate Texts in Mathematics 82, Springer, New York (1982) MR

[11] **P L Bowers**, **K Stephenson**, *A "regular" pentagonal tiling of the plane*, Conform. Geom. Dyn. 1 (1997) 58–68 MR

[12] **A Clark**, **J Hunton**, *Tiling spaces, codimension one attractors and shape*, New York J. Math. 18 (2012) 765–796 MR

[13] **A Clark**, **L Sadun**, *When shape matters: deformations of tiling spaces*, Ergodic Theory Dynam. Systems 26 (2006) 69–86 MR

[14] **F Durand**, *Linearly recurrent subshifts have a finite number of non-periodic subshift factors*, Ergodic Theory Dynam. Systems 20 (2000) 1061–1078 MR

[15] **A Forrest**, **J Hunton**, **J Kellendonk**, *Topological invariants for projection method patterns*, Mem. Amer. Math. Soc. 758, Amer. Math. Soc., Providence, RI (2002) MR

[16] **N P Frank**, *A primer of substitution tilings of the Euclidean plane*, Expo. Math. 26 (2008) 295–326 MR

[17] **N P Frank**, **L Sadun**, *Fusion: a general framework for hierarchical tilings of $\mathbb{R}^d$*, Geom. Dedicata 171 (2014) 149–186 MR

[18] **F Gähler**, **G R Maloney**, *Cohomology of one-dimensional mixed substitution tiling spaces*, Topology Appl. 160 (2013) 703–719 MR

[19] **D Gonçalves**, *On the K–theory of the stable C^*–algebras from substitution tilings*, J. Funct. Anal. 260 (2011) 998–1019 MR

[20] **B Grünbaum**, **G C Shephard**, *Tilings and patterns*, W H Freeman, New York (1987) MR

[21] **A Haynes**, **M Kelly**, **B Weiss**, *Equivalence relations on separated nets arising from linear toral flows*, Proc. Lond. Math. Soc. 109 (2014) 1203–1228 MR

[22] **A Haynes**, **H Koivusalo**, **J Walton**, *A characterization of linearly repetitive cut and project sets*, preprint (2015) arXiv

[23] **J Kellendonk**, *The local structure of tilings and their integer group of coinvariants*, Comm. Math. Phys. 187 (1997) 115–157 MR

[24] **J Kellendonk**, *Pattern-equivariant functions and cohomology*, J. Phys. A 36 (2003) 5765–5772 MR

[25] **J Kellendonk**, *Pattern equivariant functions, deformations and equivalence of tiling spaces*, Ergodic Theory Dynam. Systems 28 (2008) 1153–1176 MR

[26] **J Kellendonk**, **M V Lawson**, *Tiling semigroups*, J. Algebra 224 (2000) 140–150 MR

[27] **J Kellendonk**, **I F Putnam**, *Tilings, C^*–algebras, and K–theory*, from "Directions in mathematical quasicrystals" (M Baake, R V Moody, editors), CRM Monogr. Ser. 13, Amer. Math. Soc., Providence, RI (2000) 177–206 MR

[28] **J Kellendonk**, **I F Putnam**, *The Ruelle–Sullivan map for actions of $\mathbb{R}^n$*, Math. Ann. 334 (2006) 693–711 MR

[29] **M Kelly**, **L Sadun**, *Pattern equivariant cohomology and theorems of Kesten and Oren*, Bull. Lond. Math. Soc. 47 (2015) 13–20 MR

[30] **G Lafitte**, **M Weiss**, *Computability of tilings*, from "Fifth International Conference on Theoretical Computer Science" (G Ausiello, J Karhumäki, G Mauri, L Ong, editors), Int. Fed. Inf. Process. 273, Springer, New York (2008) 187–201 MR

[31] **J R Munkres**, *Elements of algebraic topology*, Addison-Wesley, Menlo Park, CA (1984) MR

[32] **N Ormes**, **C Radin**, **L Sadun**, *A homeomorphism invariant for substitution tiling spaces*, Geom. Dedicata 90 (2002) 153–182 MR

[33] **R Penrose**, *Pentaplexity: a class of nonperiodic tilings of the plane*, from "Geometrical combinatorics" (F C Holroyd, R J Wilson, editors), Res. Notes in Math. 114, Pitman (1984) 55–65 MR

[34] **N Priebe Frank**, **L Sadun**, *Fusion tilings with infinite local complexity*, Topology Proc. 43 (2014) 235–276 MR

[35] **C Radin**, *Aperiodic tilings, ergodic theory, and rotations*, from "The mathematics of long-range aperiodic order" (R V Moody, editor), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. 489, Kluwer, Dordrecht (1997) 499–519 MR

[36] **D Rust**, *An uncountable set of tiling spaces with distinct cohomology*, Topology Appl. 205 (2016) 58–81 MR

[37] **L Sadun**, *Pattern-equivariant cohomology with integer coefficients*, Ergodic Theory Dynam. Systems 27 (2007) 1991–1998 MR

[38] **L Sadun**, *Topology of tiling spaces*, University Lecture Series 46, Amer. Math. Soc., Providence, RI (2008) MR

[39] **L Sadun**, **R F Williams**, *Tiling spaces are Cantor set fiber bundles*, Ergodic Theory Dynam. Systems 23 (2003) 307–316 MR

[40] **K Schmidt**, *Multi-dimensional symbolic dynamical systems*, from "Codes, systems, and graphical models" (B Marcus, J Rosenthal, editors), IMA Vol. Math. Appl. 123, Springer, New York (2001) 67–82 MR

[41] **D Shechtman**, **I Blech**, **D Gratias**, **J W Cahn**, *Metallic phase with long-range orientational order and no translational symmetry*, Phys. Rev. Lett. 53 (1984) 1951–1953

[42] **J Walton**, *Cohomology of rotational tiling spaces*, preprint (2016) arXiv

[43] **E C Zeeman**, *Dihomology, III: A generalization of the Poincaré duality for manifolds*, Proc. London Math. Soc. 13 (1963) 155–183 MR

Department of Mathematics, University of York
Heslington, YO10 5DD, United Kingdom

jamie.walton@york.ac.uk

http://maths.york.ac.uk/www/jjw548

Received: 10 February 2014 Revised: 22 September 2016

Algebraic & Geometric Topology 17 (2017) 1375–1405

Fully irreducible automorphisms of the free group via Dehn twisting in $\sharp_k(S^2 \times S^1)$

FUNDA GÜLTEPE

By using a notion of a geometric Dehn twist in $\sharp_k(S^2 \times S^1)$, we prove that when projections of two $\mathbb{Z}$–splittings to the free factor complex are far enough from each other in the free factor complex, Dehn twist automorphisms corresponding to the $\mathbb{Z}$–splittings generate a free group of rank 2. Moreover, every element from this free group either is conjugate to a power of one of the Dehn twists or is a fully irreducible outer automorphism of the free group. We also prove that, when the projections of $\mathbb{Z}$–splittings are sufficiently far away from each other in the intersection graph, the group generated by the Dehn twists has automorphisms that are either conjugate to Dehn twists or atoroidal fully irreducible.

20F28, 20F65, 57M07

1 Introduction

Due to their dynamical properties, *fully irreducible* outer automorphisms are important to understanding the dynamics and geometric structure of $\mathrm{Out}(F_k)$ and its subgroups; see Levitt and Lustig [22], Clay and Pettet [5] and Behrstock, Bestvina and Clay [1]. Just like pseudo-Anosov surface homeomorphisms, fully irreducibles are characterized to be the class of automorphisms no power of which fixes a conjugacy class of a proper free factor of F_k. Since their dynamical properties and their role in $\mathrm{Out}(F_k)$ are similar to those of pseudo-Anosov mapping classes for the mapping class group, to construct fully irreducibles it is natural to seek ways similar to those of pseudo-Anosov constructions. In this work we will provide such a construction using *Dehn twist automorphisms*, by composing powers of Dehn twists from the free group of rank 2 that they generate. This is inspired by the work of Thurston on pseudo-Anosov mapping classes of the mapping class group of a surface; see Thurston [30]. Yet in our proof we use a ping-pong method similar to that used by Hamidi-Tehrani [12] to generalize Thurston's result to Dehn twists along multicurves.

Finding free groups of rank 2 generated by outer automorphisms and constructing fully irreducible automorphisms by composing (possibly powers of) other automorphisms and is not new to the study of $\mathrm{Out}(F_k)$. For instance, Clay and Pettet in [5] constructed

Published: 17 July 2017 DOI: 10.2140/agt.2017.17.1375

fully irreducibles by composing elements of a free group of rank 2 which was generated by powers of two Dehn twist automorphisms. However, the powers of the Dehn twists used to generate the free group were not uniform but depended on the twists; one needed to take a different power for each pair of Dehn twists to obtain a free group.

In their work, Clay and Pettet studied Dehn twists algebraically, as outer automorphisms of the free group, and they used algebraic tools to study them. As a result their construction produced the nonuniform powers of twists. In this paper, our goal is to construct fully irreducible automorphisms by studying Dehn twist automorphisms. To obtain a certain type of uniformity on the way, we first change the model for $\mathrm{Out}(F_k)$ from the 1–dimensional one to the 3–dimensional one: $M = \sharp_k(S^2 \times S^1)$. This way, we are able to understand Dehn twists geometrically, using essential embedded tori in M. This approach results in a more geometric construction of fully irreducibles.

More specifically, we will prove the following theorem using *geometric Dehn twists*.

Theorem 1.1 *Let T_1 and T_2 be two $\mathbb{Z}$–splittings of the free group F_k with rank $k > 2$ and α_1 and α_2 be two corresponding free factors in the free factor complex FF_k of the free group F_k. Let D_1 be a Dehn twist fixing α_1 and D_2 a Dehn twist fixing α_2, corresponding to T_1 and T_2, respectively. Then there exists a constant $N = N(k)$ such that whenever $d_{\mathrm{FF}_k}(\alpha_1, \alpha_2) \geq N$,*

(1) *$\langle D_1, D_2\rangle \simeq F_2$, and*
(2) *all elements of $\langle D_1, D_2\rangle$ which are not conjugate to the powers of D_1 and D_2 in $\langle D_1, D_2\rangle$ are fully irreducible.*

Now we would like to give the definitions necessary to understand the statement of Theorem 1.1 and explain the ideas used in its proof.

Splittings and $\mathrm{Out}(F_k)$–complexes A *Dehn twist automorphism* is an element of $\mathrm{Out}(F_k)$ defined by using $\mathbb{Z}$–splittings of F_k either as an amalgamated free product (eg $F_k = A *_{\langle c\rangle} B$) or as an HNN extension of the free group (eg $F_k = A*_{\langle c\rangle}$). More precisely, it is induced by the following automorphisms corresponding to each type of $\mathbb{Z}$–splitting:

$$A *_{\langle c\rangle} B\colon \begin{array}{ll} a \mapsto a & \text{for } a \in A, \\ b \mapsto cbc^{-1} & \text{for } b \in B \end{array} \qquad \text{and} \qquad A*_{\langle tct^{-1}=c'\rangle}\colon \begin{array}{ll} a \mapsto a & \text{for } a \in A, \\ t \mapsto tc. & \end{array}$$

Given a $\mathbb{Z}$–splitting of F_k as $F_k = A_1 *_{\langle c_1\rangle} B_1$ at least one of A_1, B_1 is a proper free factor. In the HNN extension case $F_k = A_1 *_{\langle c_1\rangle}$, the stable letter is a proper free factor. By Bass–Serre theory, each $\mathbb{Z}$–splitting of F_k gives rise to a tree whose quotient with respect to the action of the free group is a single edge. The edge stabilizer is $\mathbb{Z}$, and the vertex stabilizers in the amalgamated case are A_1 and B_1 while in the HNN case the vertex stabilizer is A_1. We will coarsely *project* each splitting onto the vertex which is a proper free factor. In the amalgamated case we consider the Dehn twist automorphism corresponding to the $\mathbb{Z}$–splitting which fixes this free factor and in the HNN case the Dehn twist automorphism will be the one fixing the vertex stabilizer. We study the action of this Dehn twist on the *free factor complex* FF_k of the free group F_k of rank k and we determine under which conditions the compositions of the Dehn twist automorphisms give fully irreducible automorphisms. The free factor complex is a simplicial complex whose vertices are conjugacy classes of proper free factors of F_k and the adjacency between two vertices corresponding to two free factors A and B is given whenever $A < B$ or $B < A$. This complex was first introduced by Hatcher and Vogtmann in [15] as a *curve complex* analog for $\mathrm{Out}(F_k)$ and in this work we will use its geometric properties due to its hyperbolicity, which are given in [2] by Bestvina and Handel.

There are several geometrically distinct hyperbolic simplicial complexes $\mathrm{Out}(F_k)$ acts on by simplicial automorphisms which are considered to be analogs to the curve complex for the mapping class group. Contrary to the case with the curve complex and the action of the mapping class group on it, it is not always possible to identify fully irreducible elements with respect to the way they act on a curve complex analog. For example, an element of $\mathrm{Out}(F_k)$ might act hyperbolically on a curve complex analog yet it may not be fully irreducible. In this work the free factor complex was used since loxodromic action of an automorphism on the free factor complex completely characterizes being fully irreducible for a free group automorphism. Thus, to identify fully irreducibles in a group generated by two Dehn twists, it is enough to have a loxodromic action.

By *geometric* Dehn twist we mean the following: For each equivalence class of a $\mathbb{Z}$–splitting, by Lemma 2.6, there is an associated homotopy class of a torus in M. More specifically, an amalgamated free product gives a separating torus in M whereas an HNN extension corresponds to a nonseparating torus. Hence each Dehn twist automorphism corresponds to a Dehn twist along the torus given by the $\mathbb{Z}$–splitting. The Dehn twist along a torus will be called a geometric Dehn twist.

Dehn twists and their almost fixed sets To prove the main theorem we use what we call a *ping-pong argument for an elliptic-type subgroup* since Dehn twists have fixed points in the free factor complex. To set up such an argument one needs to construct so-called ping-pong sets. Thus we need to know first that the points of the free factor complex which are not moved too far away by a power of a Dehn twist are manageable. More precisely, let $\phi \in \mathrm{Out}(F_k)$ and let

$$F_C(\phi) = \{x \in \mathrm{FF}_k : \exists n \neq 0 \text{ such that } d(x, \phi^n(x)) \leq C\}$$

be the almost fixed set corresponding to $\langle \phi \rangle$ in FF_k. The following theorem is the main ingredient in the elliptic-type ping-pong argument.

Theorem 1.2 *Let T be a $\mathbb{Z}$–splitting of the free group F_k with $k > 2$ and D_T denote a corresponding Dehn twist. Then, for all sufficiently large constants C, there exists a $C' = C'(C, k)$ such that the diameter of the almost fixed set $F_C(D_T)$ corresponding to $\langle D_T \rangle$ is bounded above by C'.*

We will drop the reference to the automorphism from the notation for the almost fixed set whenever it is clear from the context.

Relative twisting and distances along paths Now, to prove that the almost fixed sets of Dehn twists have bounded diameter, one needs to be able to calculate distances between points in the free factor complex effectively. However we cannot assume that there is a geodesic between two points in the free factor complex which is appropriate for our calculation purposes since we do not know what these geodesics are. But it is known that the *folding paths* in outer space give rise to geodesics in outer space and their projections to the free factor complex are quasigeodesics. To prove Theorem 1.2, we prove that there is a folding path whose projection to the free factor complex is at a bounded distance from the given free factor. To achieve this one would need an analog of the *annulus projection* and to be able to calculate distances on an annulus complex. Then using a version of the *bounded geodesic image theorem* of Masur and Minsky [24] one would conclude that whenever the number of twists is more than the universal constant given in that theorem, the quasigeodesic between a point and its twisted image has a vertex which does not intersect the core curve of the annulus. However, we do not have the main tool needed, which is an analog for annulus projection, since the subfactor projection is not defined for free factors of rank 1; see Bestvina and Feighn [3] and Taylor [29].

To calculate distances between points related to a rank-1 free factor without using a projection onto that free factor we refer to a theorem of Clay and Pettet. In [6] they give a pairing $\mathrm{tw}_a(G, G')$ called the *relative twisting number* between two graphs $G, G' \in \mathrm{CV}_k$ relative to some nontrivial $a \in F_k$ and it is defined using the *Guirardel core*. Using this pairing, they obtain a condition on the graphs $G, G' \in \mathrm{CV}_k$ that, when satisfied, enables them to construct a connecting geodesic between them, traveling through the thin part of CV_k.

Relative twisting along tori in $\sharp_k(S^2 \times S^1)$ We have used the interpretation of the *relative twisting number* pairing $\mathrm{tw}_a(G, G')$ for two spheres relative to an element of the free group, which is in our case the generator of the core (longitudinal) curve of a torus. Then the relative twist is a number which calculates distances between two spheres which are intersecting the same torus along its core curve. Mimicking annulus projection, the relative twisting number might be interpreted as the number of intersections between *projections* of some spheres in M onto a torus (yet we do not make a formal definition of such a projection). With the relative twisting number we are able to calculate a lower bound for the twisting number between a sphere and its Dehn-twisted image, relative to a torus hence in some sense relative to a rank-1 free factor (related to its core curve). Afterwards, a lemma of Clay and Pettet [6] guarantees the existence of a geodesic between the corresponding points in outer space along which the core curve gets short. Using a lemma of Bestvina and Feighn [2], we project this geodesic to the free factor complex and using the distance calculations we show easily that the almost fixed set of a Dehn twist automorphism has a bounded diameter. This completes the preparation for ping-pong with elliptic-type groups as it is given by Kapovich and Weidmann in [19]. Now we have ping-pong sets that we have control over.

The main argument, which also finishes the proof of our main result, is encoded in the following theorem.

Theorem 1.3 *Let G be a group acting on a δ–hyperbolic metric space X by isometries and ϕ_1, $\phi_2 \in G$. Suppose $C > 100\delta$ and the almost fixed sets $\mathcal{X}_C(\phi_1)$ and $\mathcal{X}_C(\phi_2)$ of $\langle\phi_1\rangle$ and $\langle\phi_2\rangle$, respectively, have diameters bounded above by a constant C'. Then there exists a constant C_1 such that whenever $d_X(\mathcal{X}_C(\phi_1), \mathcal{X}_C(\phi_2)) \geq C_1$,*

(1) *$\langle\phi_1, \phi_2\rangle \simeq F_2$, and*
(2) *every element of $\langle\phi_1, \phi_2\rangle$ which is not conjugate to the powers of ϕ_1 and ϕ_2 in $\langle\phi_1, \phi_2\rangle$ acts loxodromically in X.*

Finally, we *project* Dehn twists to the *intersection graph* $\mathcal{P}_k$, which is a simplicial complex with vertex set given by marked roses up to equivalence. There are two instances connecting vertices of $\mathcal{P}_k$. The first one is that whenever two roses share an edge with the same label, corresponding vertices are connected by an edge in $\mathcal{P}_k$. The second one is obtained whenever there is a marked surface with one boundary component such that the element of the fundamental group represented by the boundary crosses each edge of both roses twice. This simplicial complex is closely related to the *intersection graph* introduced by Kapovich and Lustig in [17], and it is proven to be hyperbolic by Mann in [23].

A fully irreducible automorphism is called *geometric* if it is induced by a pseudo-Anosov homeomorphism of a surface with one boundary component. A fully irreducible automorphism is *atoroidal* if no positive power of it preserves the conjugacy class of a nontrivial element of F_k. Moreover, only nongeometric fully irreducible automorphisms are atoroidal by a theorem of Bestvina and Handel in [4]. The important feature of the intersection graph for us is that the atoroidal fully irreducibles act loxodromically on this graph; see Mann [23].

We obtain the following theorem.

Theorem 1.4 *Let T_1 and T_2 be two $\mathbb{Z}$–splittings of F_k with $k > 2$ with corresponding free factors α_1 and α_2, and let D_1 and D_2 be two Dehn twists corresponding to T_1 and T_2, respectively. Then there exists a constant $N_2 = N_2(k)$ such that $\langle D_1, D_2\rangle \simeq F_2$ whenever $d_{\mathcal{P}_k}(\sigma(\alpha_1), \sigma(\alpha_2)) \geq N_2$, and all elements from this group which are not conjugate to the powers of D_1 and D_2 in $\langle D_1, D_2\rangle$ are atoroidal fully irreducible.*

Acknowledgements I would like to thank Ilya Kapovich for asking the question which led to the writing of this paper and many talks while I was writing it. I am especially indebted to Chris Leininger for his constant support and for suggesting the current shorter proof for the main theorem. I am very grateful for Matt Clay for giving me permission to use his proof in this paper. Many thanks also to Mark C Bell for reading the preprint and making many useful suggestions. Lastly I would like to thank the anonymous referee whose suggestions and corrections improved the paper substantially.

2 Preliminaries

2.1 Sphere systems and normal tori

The manifold $\sharp_k(S^2 \times S^1)$ is a reducible, connected 3–manifold which can be described as follows. We remove the interiors of $2k$ disjoint 3–balls from the 3–sphere S^3 and identify the resulting 2–sphere boundary components in pairs by orientation-reversing diffeomorphisms, creating k many $S^2 \times S^1$ summands. $\mathrm{Out}(F_k)$ is isomorphic to the mapping class group of $\sharp_k(S^2 \times S^1)$ up to twists about 2–spheres in $\sharp_k(S^2 \times S^1)$; see [21]. From now on we will let $M = \sharp_k(S^2 \times S^1)$.

Associated to M is a rich algebraic structure coming from the essential 2–spheres that M contains. A *sphere system* is a collection of isotopy classes of disjoint and nontrivial 2–spheres in M no two of which are isotopic.

We call a collection Σ of disjointly embedded essential, nonisotopic 2–spheres in M a *maximal sphere system* if every complementary component of Σ in M is a 3–punctured 3–sphere.

A fixed maximal sphere system Σ in M gives a description of the universal cover $\widetilde{M}$ of M as follows. Let $\mathbb{P}$ be the set of 3–punctured 3–spheres in M given by a maximal sphere system Σ and regard M as obtained from copies of P in $\mathbb{P}$ by identifying pairs of boundary spheres. Note that both boundary spheres in a pair might be contained in a single P, in which case the image of P in M is a once-punctured $S^2 \times S^1$. To construct $\widetilde{M}$, begin with a single copy of P and attach copies of P in $\mathbb{P}$ inductively along boundary spheres, as determined by unique path lifting. Repeating this process gives a description of $\widetilde{M}$ as a treelike union of copies of P. We remark that $\widetilde{M}$ is homeomorphic to the complement of a Cantor set in S^3.

To be able to define a concept of geometric Dehn twist we need to use the one-to-one correspondence between the equivalence classes of $\mathbb{Z}$–splittings of F_k and homotopy classes of *essential* tori in M. This important correspondence is given in Lemma 2.6.

For us, a torus in M is an embedding of a 2–torus in M so that the image of its fundamental group in $\pi_1(M)$ is a cyclic group isomorphic to $\mathbb{Z}$. Moreover, we consider only the tori which do not bound a solid torus in M and we call such a torus *essential*. There are two types of essential tori in M, depending on the type of the splitting of the free group they correspond to. Namely, for an amalgamated free product we have a separating torus in M and a nonseparating one for an HNN extension of the free group. Two examples can be seen in Figure 1.

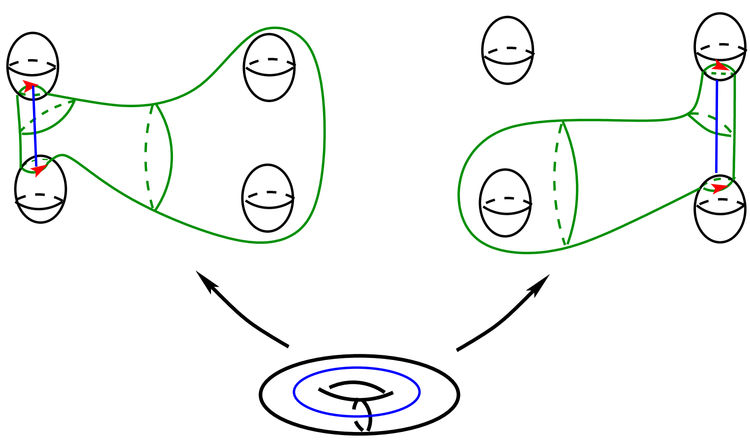

Figure 1: Embeddings with the given identifications correspond to a separating torus (left) and a nonseparating torus (right) in $\sharp_4(S^2 \times S^1)$.

Given a homotopy class of a torus, it is necessary for our purposes to identify a representative which intersects the spheres in a given maximal sphere system of M minimally. To this end, the *normal form* for tori is defined. Following Hatcher's normal form for sphere systems in [13], a normal form for tori is defined in [11] so that the intersection of the normal torus with each complementary 3–punctured 3–sphere is a disk, a cylinder or a pants piece. By [11], if a torus τ is in *normal form* with respect to a maximal sphere system Σ, the number of components of the intersection between τ and any S in Σ is minimal among all the representatives of the homotopy class $\boldsymbol{\tau}$.

In this work we will implicitly use the following existence theorem from [11].

Theorem 2.1 *Every embedded essential torus in M is homotopic to a normal torus and the homotopy process does not increase the intersection number with any sphere of a given maximal sphere system Σ.*

2.2 Geometric models, complexes and projections

Given the free group F_k on k letters, the associated *outer space* of the marked metric graphs which are homotopy equivalent to F_k was introduced by Culler and Vogtmann in [7]. We will denote by CV_k the *projectivized* outer space, in which the graphs will all have total volume 1. A marked metric graph is an equivalence class of a pair consisting of a metric graph Γ and a marking, which is a homotopy equivalence with a rose. Outer space might be thought as an analog to Teichmüller space for the mapping class group. For the details we refer the reader to [7] and [31].

The *free factor complex* of a free group is defined first by Hatcher and Vogtmann in [15] as a simplicial complex whose vertices are conjugacy classes of proper free factors and adjacency is determined by inclusion. It is hyperbolic by [2].

We will use the coarse projection $\pi\colon \mathrm{CV}_k \to \mathrm{FF}_k$ defined as follows. For each proper subgraph Γ_0 of a marked graph G that contains a circle, its image in FF_k is the conjugacy class of the smallest free factor containing Γ_0. Now by [2, Lemma 3.1], for two such proper subgraphs Γ_1 and Γ_2, we have $d_{\mathrm{FF}_k}(\pi(\Gamma_1), \pi(\Gamma_2)) \le 4$. Then for $G \in \mathrm{CV}_k$ we define

$$\pi(G) := \{\pi(\Gamma) \mid \Gamma \text{ is a proper, connected, noncontractible subgraph of } G\}.$$

We will denote the induced map $\mathrm{CV}_k \to \mathrm{FF}_k$ also by π, which is clearly a *coarse* projection in that the diameter of each $\pi(G)$ is bounded by 4. Another hyperbolic $\mathrm{Out}(F_k)$–graph we refer to is the *free bases graph* FB_k given by Kapovich and Rafi in [18]. For $k \ge 3$, this graph has vertices the free bases of F_k up to equivalence (two bases are equivalent if their Cayley graphs are F_k–equivariantly isometric), and whenever two bases representing the vertices have a common element, these vertices are connected by an edge. What is useful for us is that FB_k and FF_k are quasi-isometric.

The *intersection graph* $\mathcal{P}_k$ has vertex set consisting of marked roses up to equivalence. Two roses are connected by an edge if either they have a common edge with the same label, or there is a marked surface with one boundary component and the representative of this component crosses each edge of both roses twice. There is a Lipschitz map between FB_k and $\mathcal{P}_k$, constructed by thinking of each basis of F_k as its corresponding rose marking and observing that $\mathcal{P}_k$ shares the edges and vertices of FB_k and has some additional edges between roses.

2.3 Tori in M and $\mathbb{Z}$–splittings of the free group

In this section we will establish the correspondence between an equivalence class of a $\mathbb{Z}$–splitting and a homotopy class of a torus in M. Consider an embedded essential torus τ in M. There is a simplicial tree associated to this torus. To obtain this simplicial tree we take a neighborhood of each lift in the set of lifts $\widetilde{\tau}$ of τ and we take a vertex for each complementary component. Two complementary components are adjacent if they bound the neighborhood of the same lift. We will denote this tree by $T_{\boldsymbol{\tau}}$ and as correspondence between this tree and the torus τ we will understand the F_k–equivariant map $\widetilde{M} \to T_\tau$ which sends each complementary component of a neighborhood of a lift to a vertex and shrinks each such neighborhood to an edge. The tree constructed this way is referred to as the *Bass–Serre* tree corresponding to τ. Recall that by Bass–Serre theory, the action of F_k on T_τ gives a single-edged graph of groups decomposition of F_k, and hence a $\mathbb{Z}$–splitting of the free group F_k.

The next lemma gives the existence of an equivalence class of a $\mathbb{Z}$–splitting for each homotopy class of a torus in M and its proof is based on the notion of the *ends* of $\widetilde{M}$.

An end of a topological space is a point of the so-called *Freudenthal compactification* of the space. More precisely:

Definition 2.2 Let X be a topological space. For a compact set K, let $C(K)$ denote the set of components of the complement $X - K$. For L compact with $K \subset L$, we have a natural map $C(L) \to C(K)$. These compact sets define a *directed system* under inclusion. The set of ends $E(X)$ of X is defined to be the inverse limit of the sets $C(K)$.

The space $\widetilde{M}$ is noncompact and it has infinitely many ends. We denote the set of ends of $\widetilde{M}$ by $E(\widetilde{M})$. It is homeomorphic to a Cantor set; in particular, it is compact. For a maximal sphere system Σ in M, the set $E(T_\Sigma)$ of ends of the Bass–Serre tree T_Σ of Σ is identified with the set $E(\widetilde{M})$. By analyzing this set we were able to prove the following.

Lemma 2.3 *Let T_{τ_1} and T_{τ_2} be two Bass–Serre trees corresponding to two tori τ_1 and τ_2, respectively. If τ_1 and τ_2 are homotopic, $T_{\tau_1} = T_{\tau_2}$ and hence τ_1 and τ_2 have equivalent $\mathbb{Z}$–splittings.*

Proof Let τ be an embedded essential torus. We claim that each lift of τ is 2–sided. If it is not, then there is a nontrivial (nonhomotopic to a point) loop which intersects a lift once and connects an end of $\widetilde{M}$ to itself. By the loop theorem this loop bounds a disk in $\widetilde{M}$. Then the projection of this disk to M bounds a disk in M, which means that the torus bounds a solid torus in M. This contradicts the fact that τ is essential. Hence each lift divides $\widetilde{M}$ into two disjoint parts.

A transverse orientation on the torus gives an orientation on the spheres on each disjoint part and hence there is a labeling on the valence-2 vertices corresponding to these spheres. It is clear that each lift $L = S^1 \times \mathbb{R}$ of τ defines a decomposition of the set of ends of T_Σ into two sets L^+ and L^-, where $\mathsf{L}^+ \cap \mathsf{L}^-$ consists of two endpoints corresponding to the axis of the lift L. For each torus, let us consider all the endpoints corresponding to the axes of all lifts and eliminate them from the set of ends $E(\widetilde{M})$ of $\widetilde{M}$. Let us denote the remaining set $\widetilde{E}(\widetilde{M})$.

Now, for each lift L, we have a partition $(\mathsf{L}^+, \mathsf{L}^-)$ of the set $\widetilde{E}(\widetilde{M})$. Since lifts are disjoint, for any two lifts L_1 and L_2 we have either $\mathsf{L}_1^+ \subset \mathsf{L}_2^+$ or $\mathsf{L}_1^+ \subset \mathsf{L}_2^-$.

We construct a tree corresponding to the set of partitions as follows. For each partition $(\mathsf{L}^+, \mathsf{L}^-)$ we take a vertex. Given a pair of partitions $(\mathsf{L}_1^+, \mathsf{L}_1^-)$ and $(\mathsf{L}_2^+, \mathsf{L}_2^-)$ with $\mathsf{L}_1^+ \subset \mathsf{L}_2^+$ or $\mathsf{L}_1^+ \subset \mathsf{L}_2^-$, we connect the corresponding vertices with an edge whenever there is no collection of ends $(\mathsf{Z}^+, \mathsf{Z}^-)$ satisfying $\mathsf{L}_1^+ \subset \mathsf{Z}^+ \subset \mathsf{L}_2^+$ or $\mathsf{L}_1^+ \subset \mathsf{Z}^- \subset \mathsf{L}_2^-$. For each maximal subset of $\widetilde{E}(\widetilde{M})$ which is not separated by any lift, we take a vertex. Since the partitions of ends do not intersect, we have a tree. We will denote this tree by $T_{\mathcal{P}}$.

Since for each lift we have a partition of the ends, there is an isomorphism between the tree $T_{\mathcal{P}}$ given by the partitions and the Bass–Serre tree T_τ. To see this we first introduce another set of vertices in T_τ given by midpoints of edges and we map these "edge-midpoint" vertices of T_τ to the set of partitions, which are the components of $\widetilde{\tau}$. The image of the vertices of T_τ given by the components of $\widetilde{M} - \widetilde{\tau}$ are collections of lifts having the following property: assuming that we select the notation so that $\mathsf{L}_1^+ \subset \mathsf{L}_2^+$ for the two lifts L_1 and L_2, there is no L_3 in the collection for which $\mathsf{L}_1^+ \subset \mathsf{L}_3^+ \subset \mathsf{L}_2^+$ or $\mathsf{L}_1^+ \subset \mathsf{L}_3^- \subset \mathsf{L}_2^+$.

Now we claim that for a homotopy of embedded tori in M, the initial and final tori determine the same partition of the ends in $\widetilde{E}(\widetilde{M})$, and hence they have the same partition tree, and the same Bass–Serre tree as a result.

To see this, let τ_1 be homotopic to τ_2. To see that the lifts of τ_1 and the lifts of τ_2 give the same partition of ends we need to show that if two endpoints are separated by a component L of $\widetilde{\tau}_1$, and L is homotopic to a component L' of $\widetilde{\tau}_2$, then they are separated by L' too.

Let p and q be two endpoints separated by L. Fix an arc between them that crosses $L = S^1 \times \mathbb{R}$ in one point. During the homotopy, although no longer embedded, L moves in $\widetilde{M}$. In particular, it does not touch any endpoint. So assuming that the homotopy is transverse to the arc, its inverse image in $S^1 \times \mathbb{R} \times I$ consists of circles and arcs properly embedded in $S^1 \times \mathbb{R} \times I$. Note that if the homotopy could cross an endpoint of the arc, then an arc of the inverse image could fail to be properly embedded in $S^1 \times \mathbb{R} \times I$. But this does not happen since the homotopy between the two tori induces a homotopy between normal representatives of each tori, corresponding to a fixed maximal sphere system in M. By [11], such a homotopy is *normal* at each stage hence cannot cross an endpoint.

Back to the inverse image of the homotopy between the two tori, since only one endpoint of the inverse image of the arc is in L, there must be an odd number of endpoints in L' (ie the arc crosses L' an odd number of times) and therefore L' also separates p and q. $\square$

Recall that given a $\mathbb{Z}$–splitting of F_k, an associated *Dehn twist automorphism* of F_k is defined in the following two ways:

$$A *_{\langle c\rangle} B\colon \begin{array}{ll} a \mapsto a & \text{for } a \in A, \\ b \mapsto cbc^{-1} & \text{for } b \in B \end{array} \qquad \text{and} \qquad A*_{\langle tct^{-1}=c'\rangle}\colon \begin{array}{ll} a \mapsto a & \text{for } a \in A, \\ t \mapsto tc. & \end{array}$$

On the left is the definition when the $\mathbb{Z}$–splitting is given by an amalgamated product $F_k = A *_{\langle c\rangle} B$ and on the right is the definition when the $\mathbb{Z}$–splitting is an HNN extension $A*_{\langle c\rangle}$ of the free group F_k. Note that the Dehn twist automorphism in the amalgamated case is defined up to conjugacy since it is possible to reverse the roles of A and B.

Before we give the last lemma in this section, we give two theorems relating $\mathbb{Z}$–splittings to free splittings which will be used in the proof.

Theorem 2.4 (Shenitzer [26]) *Suppose that a free group* $\mathbb{F}$ *is an amalgamated free product* $\mathbb{F} = A *_B C$*, where* B *is cyclic. Then* B *is a free factor of* A *or a free factor of* C*.*

Theorem 2.5 (Swarup [28]) *Suppose that a free group* $\mathbb{F}$ *is an HNN extension* $\mathbb{F} = A*_B$*, where* $B \neq 1$ *is cyclic. Then* A *has a free product structure* $A = A_0 * A_1$ *in such a way that one of the following symmetric alternatives hold, where* t *is the stable letter:*

(1) $B \subset A_0$ *and there exists* $a \in A$ *such that* $t^{-1}Bt = a^{-1}A_1 a$*, or*

(2) $t^{-1}Bt \subset A_0$ *and there exists* $a \in A$ *such that* $B = a^{-1}A_1 a$*.*

Finally, the following lemma gives the converse relationship between a torus and a $\mathbb{Z}$–splitting and hence explains why we are interested in tori in M. The proof is due to Matt Clay.

Lemma 2.6 *Given a $\mathbb{Z}$–splitting Z and an associated Dehn twist automorphism, there is a torus τ in M unique up to homotopy such that $T_Z = T_\tau$, where T_Z and T_τ are the corresponding Bass–Serre trees.*

Proof In this proof we will build a homotopy class of a torus from a sphere and a loop. First we use Theorems 2.4 and 2.5, which relate a $\mathbb{Z}$–splitting of F_k to a free splitting of F_k. Then, to relate the free splitting to a homotopy class of a sphere, we use a theorem originally due to Kneser [20]. This theorem is later developed by Grushko [9], and most recently by Stallings [27], and these are the versions we will be referring to. We treat the amalgamated product and HNN-extension cases separately. The amalgamated case has schematic pictures Figure 2 and Figure 3 associated to the proof.

Case 1 We first consider the case of an amalgamated free product $F_k = A *_{\langle b\rangle} B$. By Shenitzer's theorem, Theorem 2.4, $\langle b\rangle$ is either a free factor of A or a free factor of B. Hence there is a free splitting $F_k = A * B_0$, where $B = \langle b\rangle * B_0$, or $F_k = A_0 * B$ with $A = A_0 * \langle b\rangle$. Let us assume the former, and let $S \subset M$ be an embedded (separating) sphere representing this splitting. We fix a basepoint $* \in M$ and assume it lies on S. As $b \in A$, there is an embedded loop $\gamma \subset M$ that represents $b \in F$ and only intersects S at $*$. For small ϵ, the boundary of the closed ϵ–neighborhood of $S \cup \gamma$ consists of two components: an embedded sphere isotopic to S and an embedded essential torus τ_γ.

Every torus can be written as a sphere and a loop attached to it. Hence it is clear from the construction that the splitting of F_k associated to τ_γ is the original splitting. However, there are some choices made in the construction of τ_γ and it must be shown that different choices result in homotopic tori. It is clear that changing S or γ in the construction by a homotopy results in a change of τ_γ by a homotopy.

Now since Shenitzer's theorem, Theorem 2.4, gives many possible splittings differing by automorphisms of B fixing $\langle b\rangle$ (hence Nielsen automorphisms that fix b), we need to consider two different complementary free factors B_0 and B_1 of A such that $\langle b\rangle * B_0 = \langle b\rangle * B_1 = B$ and show that the tori obtained after we add the loop to corresponding spheres are homotopic, even when the spheres themselves are not. For this, let S_0 and S_1 be the spheres representing the splittings $A * B_0$ and $A * B_1$, respectively, and τ_0 and τ_1 be the tori as constructed above using these spheres. We assume that γ intersects S_0 only at the fixed basepoint $* \in M$.

We first treat the special case that B_1 is obtained from B_0 by replacing a generator x in B_0 by xb. Fix a basis for F_k consisting of a basis for A and a basis for B_0, where x is one of the generators for B_0. This corresponds to a sphere system in M which decomposes as $\Sigma_A \cup \Sigma_{B_0}$; the sphere S_0 separates the two sets Σ_A and Σ_{B_0}. In terms of these sphere systems, we can describe a homeomorphism that takes S_0 to (a sphere isotopic to) S_1.

Denote by Σ_γ the ordered set of spheres (all in Σ_A) pierced by γ starting from the basepoint. Cut M open along the sphere β corresponding to the generator x and via a homotopy push the boundary sphere β^- through the spheres in Σ_γ in order, dragging S_0 along. After regluing β^+ and β^-, the image of S_0 is S_1 and the sphere β now corresponds to xb. By shrinking β^- and S_0, we can assume that the homotopy is the identity on τ_0 and γ. Thus, we have a homeomorphism taking S_0 to S_1, taking $S_0 \cup \gamma$ to $S_1 \cup \gamma$ and which is the identity on τ_0. As a homeomorphism takes a regular neighborhood to a regular neighborhood, τ_0 is homotopic to τ_1.

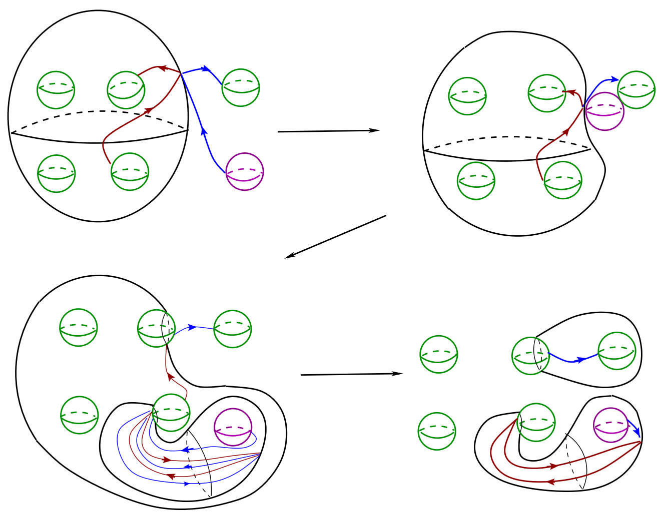

Figure 2: The homotopy which slides the pink sphere along the red loop γ representing b, where $\pi_1(M) = \langle a, b, c\rangle$. In the first picture, a sphere (black) and the b loop (red) are given, where the base point is on the sphere. c is depicted in blue in the picture.

To see a schematic picture of this homotopy we refer the reader to Figure 2. In the first picture, the black sphere (an example of S_0) and a neighborhood of the red loop give a torus (τ_0), and this torus is homotopic to the torus obtained from the last picture by taking a neighborhood of the black sphere (now an example of S_1) and the red loop.

A similar argument works if we replace x by xb^{-1}, bx or $b^{-1}x$.

The general case now follows as we can transform B_0 to B_1 by a finite sequence of the above transformations plus changes of basis that do not affect the associated spheres.

Indeed, by [32, Theorem 4.1], the subgroup of automorphisms of B that fix $b \in B$ is generated by the Nielsen automorphisms that fix b.

Last, we consider the possibility $F_k = A_0 * \langle b \rangle * B_0$, where $A = \langle A_0, b\rangle$ and $B = \langle B_0, b\rangle$. Let S_A and S_B be the spheres representing the splittings $A * B_0$ and $A_0 * B$, respectively, fix loops γ_A and γ_B representing b, and consider the neighborhoods s_A of $S_A \cup \gamma_A$ and s_B of $S_B \cup \gamma_B$. Since these neighborhoods each give a torus and a sphere, we have tori τ_A and τ_B that both represent the splitting $A *_{\langle b\rangle} B$. In this case, as a component of $M - (S_A \cup S_B)$ is $S^1 \times S^2$ with two balls removed, it is easy to see that τ_A and τ_B are homotopic. Indeed, let us model $S^1 \times S^2$ as the region between the spheres of radius 1 and 2 in $\mathbb{R}^3$ after identifying the boundary spheres. Remove a ball of radius $\frac{1}{4}$ at each of the points $\left(0,0,\frac{3}{2}\right)$ and $\left(0,0,-\frac{3}{2}\right)$. For γ_A we can choose the intersection with the positive z–axis; for γ_B we can choose the intersection with the negative z–axis. Then clearly the torus obtained from the intersection with the xy–plane is homotopic to both τ_A and τ_B. For a simple example see Figure 3.

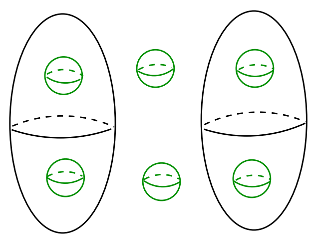

Figure 3: In this example, $F_3 = \langle a\rangle * \langle b\rangle * \langle c\rangle$ and $A = \langle A_0, b\rangle$ with $A_0 = \langle a\rangle$ and $B = \langle b, B_0\rangle$ with $B_0 = \langle c\rangle$. Spheres S_A and S_B are displayed in black.

Case 2 We now consider the case of an HNN-extension $F_k = A *_{\langle b\rangle}$. By Swarup's theorem, Theorem 2.5, there is a free factorization $A = A_0 * \langle t^{-1}bt\rangle$ for some $t \in F_k$ such that A_0 is a corank-1 free factor of F_k and such that $b \in A_0$. Let $S \subset M$ be an embedded (nonseparating) sphere representing the splitting $F_k = A_0 *_{\{1\}}$. We fix a basepoint $p \in M$ and assume it lies on S. As $b \in A_0$, there is an embedded loop $\gamma \subset M$ that represents $b \in F_k$ and only intersects S at p. Further, both ends of γ are on the same side of S and so a neighborhood of $s = S \cup \gamma$ gives a torus. Let τ be this neighborhood of s, and as in Case 1, it is clear that the splitting associated to τ is the original splitting.

Given another torus $\tau' \subset M$ that represents the same splitting, we can compress τ' to a union of a sphere and a loop $s' = S' \cup \gamma'$ such that S' represents a splitting of the form $A_1 *_{\{1\}}$, where $A = A_1 * \langle tbt^{-1}\rangle$. Then, as in Case 1, there is a sequence of transformations taking A_0 to A_1 that do not change the homotopy type of the corresponding torus. □

3 Geometric intersection and relative twisting using ends of $\widetilde{M}$

3.1 Intersection criterion and the relative twisting number

The Guirardel core $\mathcal{C}$ is a way of assigning a closed, connected CAT(0) complex to a pair of splittings which counts the number of times the corresponding Bass–Serre trees intersect. Guirardel's version unifies several notions of intersection number in the literature, including the one for two splittings of finitely generated groups given by Scott and Swarup [25]. For two F_k–trees T_0 and T_1, the core is roughly the main part of the diagonal action of F_k on $T_0 \times T_1$. For the details we refer the reader to [10] and [1].

Definition 3.1 Let T be a tree and p a point in it. A *direction* is a connected component of $T - p$. Given two trees T_0 and T_1, a *quadrant* is a product $\delta \times \delta'$ of two directions $\delta \subset T_0$ and $\delta' \subset T_1$.

We fix a basepoint $* = (*_0, *_1)$ in $T_0 \times T_1$ and we say that a quadrant Q is *heavy* if there exists a sequence $\{g_n\}$ in F_k such that $g_n(*) \in Q$ for every n and $d_{T_i}(*_i, g_n(*_i)) \to \infty$ as $n \to \infty$. A quadrant is *light* if it is not heavy.

Definition 3.2 Let T_0 and T_1 be two F_k–trees. The Guirardel core $\mathcal{C}$ is defined as

$$\mathcal{C} = \mathcal{C}(T_0 \times T_1) = (T_0 \times T_1) - \cup_I Q,$$

where I is over all of the light quadrants.

Let p be a point in T_0. Then $\mathcal{C}_p = \{x \in T_1 \mid (p, x) \in \mathcal{C}\}$ is a subtree of T_1 called the *slice of the core* above the point p. The slice which is a subtree of T_0 is defined similarly.

Given two trees T_0 and T_1, we define the *Guirardel intersection number* between them by

$$i(T_0, T_1) = \mathrm{vol}(\mathcal{C}/F_k),$$

where the right-hand side is the volume of the action of F_k on the Guirardel core $\mathcal{C}(T_0 \times T_1)$ for the product measure on $T_0 \times T_1$. Note that for simplicial trees T_0 and T_1 this volume is the number of 2–cells in $\mathcal{C}/F_k$, which will be our case.

Given two homotopy classes of sphere systems $\boldsymbol{S_1}$ and $\boldsymbol{S_2}$ in M, the intersection number between them is the number $i(\boldsymbol{S_1}, \boldsymbol{S_2})$ of components of the intersection of normal representatives of S_1 and S_2 in $\boldsymbol{S_1}$ and $\boldsymbol{S_2}$, respectively, when the intersections are transversal [13]. This is called the *geometric* intersection number.

The work of Horbez in [16] relates the geometric intersection number between two sphere systems to the Guirardel intersection number between corresponding trees. Each sphere in $\widetilde{M}$ corresponds to an edge in the associated Bass–Serre tree. By taking ϵ–neighborhoods of each edge, one obtains for each geometric intersection between two spheres a square in the product of the Bass–Serre trees. If the intersection is essential, this square is in the Guirardel core. By the definition of the Guirardel core, a square which is in the core is in a heavy quadrant. As a consequence there are 4 unbounded disjoint regions in $\widetilde{M}$ corresponding to such a square. On the other hand, each sphere $\widetilde{S}$ gives two disjoint sets $E^+(\widetilde{S})$ and $E^-(\widetilde{S})$ of ends of $\widetilde{M}$. Thus whenever two spheres $\widetilde{S}_1$ and $\widetilde{S}_2$ intersect essentially in $\widetilde{M}$ there are four disjoint sets of ends $E^+(\widetilde{S}_1)\cap E^+(\widetilde{S}_2)$, $E^+(\widetilde{S}_1)\cap E^-(\widetilde{S}_2)$, $E^-(\widetilde{S}_1)\cap E^+(\widetilde{S}_2)$ and $E^-(\widetilde{S}_1)\cap E^-(\widetilde{S}_2)$ of $\widetilde{M}$ in the complement of the intersection circle, each matching with the corresponding unbounded region.

Finally we have the following definition which we will use in the next section.

Definition 3.3 We will call the existence of four disjoint sets of ends of $\widetilde{M}$ in the complement of an intersection circle between two spheres the *intersection criterion* for that intersection circle.

According to the discussion above, an intersection circle between two spheres is essential if and only if we have the intersection criterion satisfied for that intersection circle.

Now we will define the *relative twisting number* for two intersecting sphere systems Σ_1 and Σ_2.

Given an axis of an element in $\widetilde{M}$, there are two ends of $\widetilde{M}$ which are fixed by this axis.

Definition 3.4 A sphere $\widetilde{S}$ is said to intersect an axis $\mathfrak{a}$ whenever the two ends of $\widetilde{M}$ determined by $\mathfrak{a}$ are separated by the two disjoint sets $E^+(\widetilde{S})$ and $E^-(\widetilde{S})$ of ends corresponding to $\widetilde{S}$.

Now we will define the relative twisting number between Σ_1 and Σ_2 relative to an element $a \in F_k$. This number is meaningful when both sphere systems intersect an axis $\mathfrak{a}$ of a in $\widetilde{M}$. The definition of geometric relative twisting of [6] is for two points in outer space, which are simple sphere systems in our setting. For our purposes we will translate their definition to one which is stated in terms of ends of $\widetilde{M}$.

Before we give the definition given in [6], we set up some notation first. Let Σ_1 and Σ_2 be two simple sphere systems in M and T_{Σ_1} and T_{Σ_2} the corresponding Bass–Serre trees. Let $e_{\widetilde{S}_1}$ and $e_{\widetilde{S}_2}$ be two edges in T_{Σ_1} and T_{Σ_2} corresponding to spheres $S_1 \in \Sigma_1$ and $S_2 \in \Sigma_2$, respectively. For an element $a \in F_k$, let us call $T^a_{\Sigma_1}$ and $T^a_{\Sigma_2}$ the sets of edges of T_{Σ_1} and T_{Σ_2}, respectively, whose elements intersect a fixed axis $\mathfrak{a}$ of a. Also denote by $a^k e_{\widetilde{S}_1}$ the k^{th} iterate of $e_{\widetilde{S}_1}$ under the action of $\langle a \rangle$ on T_{Σ_1} along an axis $\mathfrak{a}$.

Definition 3.5 [6] For an element $a \in F_k$, the relative twisting number $\mathrm{tw}_a(\Sigma_1, \Sigma_2)$ of Σ_1 and Σ_2 relative to a is defined to be

$$\mathrm{tw}_a(\Sigma_1, \Sigma_2) := \max_{\substack{e_{\widetilde{S}_1} \subset T^a_{\Sigma_1} \\ e_{\widetilde{S}_2} \subset T^a_{\Sigma_2}}} \{ k \mid a^k e_{\widetilde{S}_1} \times e_{\widetilde{S}_2} \in \mathcal{C} \text{ and } e_{\widetilde{S}_1} \times e_{\widetilde{S}_2} \in \mathcal{C} \}.$$

Using the fact that a geometric intersection between two spheres is a square in the Guirardel core and hence gives a separation of ends of $\widetilde{M}$ into 4 nonempty disjoint sets, we tailor the definition of [6] above to the one below to suit our needs.

Definition 3.6 For $i \in \{1, 2\}$, let $\widetilde{S}_i$ be two spheres and $E^{\mp}(\widetilde{S}_i)$ be the set of ends of $\widetilde{M}$ separated by these spheres. Assume that both spheres intersect an axis of $\boldsymbol{a} \in F_k$. Then the relative twisting number $\mathrm{tw}_a(\widetilde{S}_1, \widetilde{S}_2)$ of $\widetilde{S}_1$ and $\widetilde{S}_2$ relative to a is defined by

$$\mathrm{tw}_a(\widetilde{S}_1, \widetilde{S}_2) := \max\bigl\{k \in \mathbb{Z} \mid E^{\mp}(\widetilde{S}_i) \cap E^{\mp}(a^k \widetilde{S}_j) \neq \varnothing \\ \text{whenever } E^{\mp}(\widetilde{S}_i) \cap E^{\mp}(\widetilde{S}_j) \neq \varnothing \text{ and } \{i, j\} \in \{1, 2\}\bigr\}.$$

4 Dehn twist along a torus: the geometric picture

4.1 Definition of a Dehn twist along a torus

We will now give the definition of the Dehn twist homeomorphism about a torus in M, a description of the action of such a homeomorphism on spheres in M, and a description of the action on F_k.

To define a Dehn twist along an embedded torus τ, we will take a parametrized tubular neighborhood of the torus in M.

Definition 4.1 Let $\tau \colon \mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z} \times [0, 1] \to M$ be an embedding such that the image $\tau(\{0\} \times \mathbb{R}/\mathbb{Z} \times \{0\})$ bounds a disk in M. Denote by τ the associated torus $\tau(\mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z} \times \{0\})$. The *geometric Dehn twist* D_τ along the torus τ is the homeomorphism of M that is the identity on the complement of the image of the map τ and for which a point $p = \tau(x, y, t)$ is sent to $\tau(x + t, y, t)$.

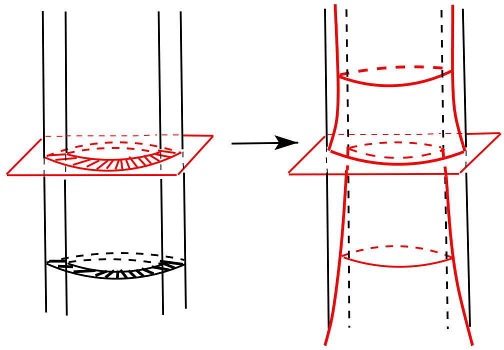

Figure 4: The image of the intersection annulus under a Dehn twist along the thick black torus

Here, the direction of the associated torus which bounds a disk in M will be called the *meridional* direction and the other one will be called the *longitudinal* direction. For a geometric description of the twist, we refer the reader to Figure 4.

Now, an ambiguity might arise in the definition of a geometric Dehn twist when it comes to determining a longitudinal curve for the parametrization. But two such choices differ by a map $x \mapsto x + ny$ for some integer n and a twist along the meridional direction. By work of Laudenbach [21], the meridional direction does not give a nontrivial homeomorphism since twists along the meridional direction correspond to twists along 2–spheres in M. It is known that such mapping classes act trivially on F_k, and hence they are in the kernel of the homomorphism $\mathrm{Map}(M) \to \mathrm{Out}(F_k)$; see [14]. Hence the induced outer automorphism $D_{\tau *}$ from the geometric Dehn twist D_τ is independent of the parametrization of the neighborhood of the torus (the image of the map τ).

Now, assume that we have a loop ξ intersecting a torus τ transversely. Then the image $D_\tau(\xi)$ of the loop under the geometric twist D_τ is obtained as follows: We surger the loop at the intersection point and insert a loop β^+ or β^- representing a generator of $\pi_1(\tau)$ in $\pi_1(M)$, depending on which side of the torus the intersection point is. Hence the induced automorphism conjugates ξ with one of β^+ or β^- and fixes the other. If τ is nonseparating, then the stable letter is multiplied by ξ.

This coincides with the action of $D_{\tau *}$ on the corresponding splitting: When the splitting is an amalgamated product of the form $A *_{\langle c \rangle} B$, the factor A is fixed whereas B is conjugated by the generator of the fundamental group of torus τ in M. (The roles of A and B might be changed.) When we have $A *_{\langle tct^{-1} = c' \rangle}$, the Dehn twist automorphism $D_{\tau *}$ fixes A and t is multiplied by c.

As a summary we have:

Lemma 4.2 *Let τ be an embedded torus and D_τ the associated geometric Dehn twist. Then $D_{\tau*}=D_Z$, where $D_{\tau*}$ is the Dehn twist automorphism induced by the homeomorphism D_τ and D_Z is the Dehn twist automorphism given by the $\mathbb{Z}$–splitting Z associated to the torus τ.*

From Lemmata 4.2 and 2.3, we easily deduce the following.

Proposition 4.3 *Let τ_1 and τ_2 be two homotopic tori. Then up to conjugacy we have $D_{\tau_1*}=D_{\tau_2*}$, where D_{τ_1*} and D_{τ_2*} are the Dehn twist automorphisms induced by the geometric Dehn twists D_{τ_1} and D_{τ_2}.*

Proof Since homotopic tori give equivalent splittings by Lemma 2.3, the Dehn twist automorphisms are equal by definition. Then by Lemma 4.2 the corresponding Dehn twists induced by the geometric Dehn twists are equal. □

For our purposes, we need to find a lower bound on $\mathrm{tw}_a(G, D^n_T(G))$ for a given simple sphere system G and a $\mathbb{Z}$–splitting T. Now we have $D_T=D_\tau$, where τ is the essential embedded torus given by the $\mathbb{Z}$–splitting T for which the Dehn twist is defined and a is the generator of the image of the fundamental group of τ in $\pi_1(M)$ under the map induced from the embedding $\iota\colon \tau\to M$. Now we take a maximal sphere system Σ containing G and homotope τ to be normal with respect to this sphere system. Then, by [11], τ intersects the spheres of Σ minimally. Now we take a lift of the torus which has an axis a conjugate of a. The relative twisting number counts the number of iterates of a sphere which intersect the image of another sphere under a Dehn twist along an axis. Hence we have

$$\mathrm{tw}_a(G, D^n_T(G))=\mathrm{tw}_\tau(G, D^n_\tau(G)).$$

5 Lower bound on the relative twisting number

In this section we will prove that the relative twisting number of a simple sphere system and its Dehn-twisted image has a lower bound, which is linear with respect to the power of the Dehn twist. We first have the following introductory lemma. Recall that a *simple* sphere system in M is one which has the complementary components in M simply connected.

Given a sphere S and a torus τ an intersection circle of $S\cap\tau$ which does not bound a disk in τ is called a *meridian* in τ.

Lemma 5.1 *Given an embedded essential torus τ and a simple sphere system G, there exists a sphere $S \in G$ such that at least one intersection between S and τ is a meridian in τ.*

Proof We first complete G to a maximal sphere system and homotope τ to be normal with respect to this maximal sphere system. Then τ intersects minimally every sphere of G it intersects [11]. Assume that no isotopy class of spheres in G intersects τ in a way that the intersection is a meridian in τ. On the other hand, since τ is essential, the image of the fundamental group of τ in $\pi_1(M)$ is nontrivial. Hence the core curve of τ exists and by the assumption it must be contained in a complementary component of G in M. But this is not possible since G is simple. □

Theorem 5.2 *Let T be a $\mathbb{Z}$–splitting of the free group F_k, let τ be the associated torus, and let D_τ be the Dehn twist along τ. For $G \in \mathrm{CV}_k$ and $n \geq 2$,*

$$\mathrm{tw}_\tau(G, D_\tau^n(G)) \geq n-1.$$

Proof We will prove this theorem by taking G as a simple sphere system instead of a marked metric graph. Let Σ be a maximal sphere system completing G and homotope τ to be the normal with respect to Σ.

Now, let $S \in G$ be such that S and τ intersect in a way that $\mu = S \cap \tau$ is a meridian in τ. Existence of a meridional intersection circle is given by Lemma 5.1. As before, let D_τ be the Dehn twist along the normal torus τ. More precisely, let $N(\tau)$ be a tubular neighborhood of τ in which D_τ is supported. Let $\widetilde{\tau}$ be the full preimage of τ in $\widetilde{M}$ and $\widetilde{\tau}_0 \in \widetilde{\tau}$. Denote also by $N(\widetilde{\tau})$ the full preimage of $N(\tau)$ and by $N(\widetilde{\tau}_0)$ the component containing $\widetilde{\tau}_0$. Let $\widetilde{S}$ be the full preimage of S and $\widetilde{S}_0$ a component such that $\widetilde{S}_0 \cap \widetilde{\tau}_0 = \widetilde{\mu}_0$, a lift of μ. Let us denote by a the generator of $\pi_1(\tau)$ corresponding to $\widetilde{\tau}_0$, and hence we have a covering transformation $a\colon \widetilde{M} \to \widetilde{M}$ which stabilizes $\widetilde{\tau}_0$. Let Δ be the region between $\widetilde{S}_0$ and $a\widetilde{S}_0$ which is the fundamental domain of $\langle a \rangle$ on $\widetilde{M}$. Set $\widetilde{S}_{0,j} = a^j \widetilde{S}_0$. Then $a^j \Delta$ is the region bounded by $\widetilde{S}_{0,j}$ and $\widetilde{S}_{0,j+1}$.

Since $\mathrm{tw}_\tau(G, D_\tau^n(G)) \geq \mathrm{tw}_\tau(S, D_\tau^n(S))$, it is sufficient to prove that

$$\mathrm{tw}_\tau(S, D_\tau^n(S)) \geq n-1.$$

Recall that we denote by $E(\widetilde{M})$ the ends of $\widetilde{M}$. As discussed in Section 2.2, there is a pair of ends of $\widetilde{M}$ fixed by a and $\widetilde{\tau}_0$ separates the remaining set of ends into two disjoint sets E_0^+ and E_0^-. Since $\widetilde{\tau}_0$ is separating in $\widetilde{M}$, there is a ray ℓ which connects an end $\mathrm{e}^+ \in E_0^+$ to an end $\mathrm{e}^- \in E_0^-$, intersecting $\widetilde{\tau}_0$ only once. Observe that since τ is essential, there is always such a ray which is disjoint from $\widetilde{\tau} - \widetilde{\tau}_0$.

Let X^+ and X^- be the components of $\widetilde{M} - N(\tilde{\tau})$ whose closures meet $N(\tilde{\tau}_0)$. Since D_τ is the identity on $M - N(\tau)$ we choose a lift $\widetilde{D}_\tau$ which is the identity on X^- and a translation on X^+. Without loss of generality we may assume this is translation by a. Hence $\widetilde{D}_\tau(\mathsf{e}^-) = \mathsf{e}^-$ and $\widetilde{D}^m_\tau(\mathsf{e}^+) = \mathsf{a}^m\mathsf{e}^+$.

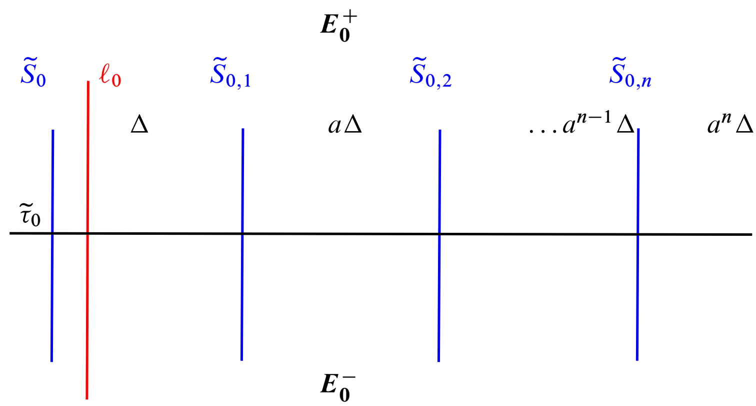

Figure 5: Fundamental domains, sets of ends separated by $\tilde{\tau}_0$ and an arc ℓ_0 connecting them

Let $E^-_{0,0} \subset E^-_0$ and $E^+_{0,0} \subset E^+_0$ be two disjoint sets of ends in Δ. Now, as described above, in this fundamental domain there is a line ℓ_0 which connects an end e^-_0 in $E^-_{0,0}$ to an end e^+_0 in $E^+_{0,0}$ intersecting $\tilde{\tau}$ once in $\tilde{\tau}_0$. (See Figure 5.) Now, after n times Dehn twisting along $\tilde{\tau}$, the image ray $\widetilde{D}^n_\tau(\ell_0)$ will connect the point e^-_0 in $E^-_{0,0}$ to the point $\widetilde{D}^n_\tau(\mathsf{e}^+_0) = \mathsf{a}^n\mathsf{e}^+_0 = \mathsf{e}^+_\mathsf{n}$ in $E^+_{0,n-1} = \widetilde{D}^n_\tau(E^+_{0,0})$. (See schematic picture Figure 6.)

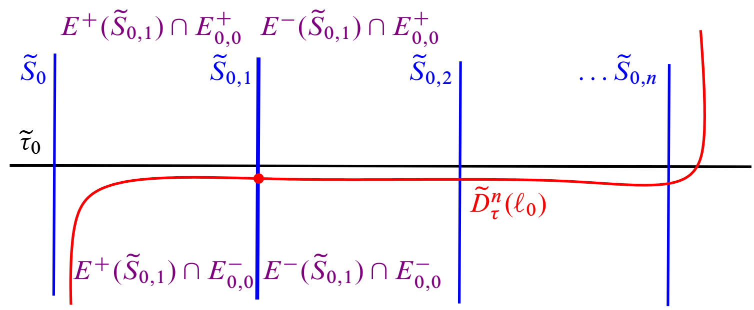

Figure 6: The image of ℓ_0 under Dehn twisting and sets of ends corresponding to the intersection $\widetilde{D}^n_\tau(\widetilde{S}_0) \cap \widetilde{S}_{0,1}$

On the other hand, let us denote by $E^+(\widetilde{S}_{0,s})$ and $E^-(\widetilde{S}_{0,s})$ the two disjoint sets of ends corresponding to the sphere $\widetilde{S}_{0,s}$ for $s \in \{1, \dots, n\}$. Now, without loss of generality, assume that $\mathsf{e}_0^- \in E^-(\widetilde{S}_{0,1})$. Then, since the image ray $\widetilde{D}_\tau^n(\ell_0)$ still intersects $\widetilde{\tau}_0$ only once, and intersects neither $\widetilde{S}_0$ (this would create a bigon) nor any other lift of the torus, $\widetilde{D}_\tau^n(\mathsf{e}_0^+) = \mathsf{e}_\mathsf{n}^+ \in E^+(\widetilde{S}_{0,1})$. So for $s=1$ we have 4 disjoint, nonempty sets of ends

$$E^+(\widetilde{S}_{0,1}) \cap E_{0,0}^-, \quad E^+(\widetilde{S}_{0,1}) \cap E_{0,0}^+, \quad E^-(\widetilde{S}_{0,1}) \cap E_{0,0}^-, \quad E^-(\widetilde{S}_{0,1}) \cap E_{0,0}^+.$$

By the intersection criterion, this shows that $\widetilde{D}_\tau^n(\ell_0)$ intersects $\widetilde{S}_{0,1}$. Since

$$E^+(\widetilde{S}_{0,1}) \subset E^+(\widetilde{S}_{0,n}) \quad \text{and} \quad E^-(\widetilde{S}_{0,n}) \subset E^-(\widetilde{S}_{0,0}),$$

we similarly have the nonempty disjoint sets of ends

$$E^+(\widetilde{S}_{0,n}) \cap E_{0,n-1}^-, \quad E^+(\widetilde{S}_{0,n}) \cap E_{0,n-1}^+, \quad E^-(\widetilde{S}_{0,n}) \cap E_{0,n-1}^-, \quad E^-(\widetilde{S}_{0,n}) \cap E_{0,n-1}^+.$$

Again by the intersection criterion this means that $\widetilde{D}_\tau^n(\ell_0)$ intersects $\widetilde{S}_{0,n}$ as well. As a result, $\widetilde{D}_\tau^n(\ell_0)$ intersects all iterates $\widetilde{S}_{0,s}$ of $\widetilde{S}_0$, where $s \in \{1, \dots, n\}$.

Hence $\widetilde{D}_\tau^n(\widetilde{S}_0)$ intersects all n iterates of $\widetilde{S}_0$. More precisely,

$$\widetilde{D}_\tau^n(\widetilde{S}_0) \cap \widetilde{S}_{0,j} \neq \varnothing \quad \text{for } j = 1, \dots, n \text{ with } n \geq 2,$$

and thus we conclude that

$$\mathrm{tw}_\tau(S, D_\tau^n(S)) \geq n-1. \qquad \square$$

6 The almost fixed set

In this section we will prove the following theorem which says that there is an upper bound on the diameter of the almost fixed set of a Dehn twist and this upper bound depends only on the rank of the free group.

Theorem 1.2 *Let T be a $\mathbb{Z}$–splitting of F_k with $k > 2$ and D_T denote the corresponding Dehn twist. Then, for all sufficiently large constants C, there exists a $C' = C'(C, k)$ such that the diameter of the almost fixed set*

$$F_C = \{x \in \mathrm{FF}_k : \exists n \neq 0 \text{ such that } d(x, D_T^n(x)) \leq C\}$$

corresponding to $\langle D_T \rangle$ is bounded above by C'.

6.1 Finding the folding line

For G_1 and G_2 in CV_k, let $m(G_1, G_2)$ be the infimum of the set of maximal slopes of all *change of marking* maps (maps linear on edges) $f\colon G_1 \to G_2$. Then we define a function $d_L\colon \mathrm{CV}_k \times \mathrm{CV}_k \to \mathbb{R}_{\geq 0}$ by

$$d_L(G_1, G_2) = \log m(G_1, G_2).$$

Despite being nonsymmetric, since this is its only failure to be a distance, d_L is referred to as the *Lipschitz metric* on CV_k.

For an interval $I\subset\mathbb{R}$ the folding lines $\tilde{g}\colon I\to\mathrm{CV}_k$ are the paths connecting any given two points in the interior of CV_k, obtained as follows:

For $G_1,G_2\in\mathrm{CV}_k$ let $f\colon G_1\to G_2$ be a change of marking map whose Lipschitz constant realizes the maximal slope. We find a path based at G_1 which is contained in an open simplex of unprojectivized outer space and parametrize it by arclength. Then we concatenate this path with another geodesic path outside the open simplex obtained by the folding process. The resulting path $g\colon[0,d_L(G_1,G_2)]\to\mathrm{CV}_k$ is a geodesic by Francaviglia and Martino [8] and is called the *folding line*.

For a given F_k–tree Γ and an element $a\in F_k$, let $\ell_\Gamma(a)$ denote the minimal translation length of a in Γ. To prove Theorem 1.2, we use the following result from [6].

Theorem 6.1 [6] *Suppose $G,G'\in\mathrm{CV}_k$ with $d=d_L(G,G')$ such that $\mathrm{tw}_a(G,G')$ is at least $n+2$ for some $a\in F_k$. Then there is a geodesic (folding line) $g\colon[0,d]\to\mathrm{CV}_k$ such that $g(0)=G$ and $g(d)=G'$ and for some $t\in[0,d]$, we have $\ell_{g(t)}(a)\le 1/n$. In other words, $g([0,d])\cap\mathrm{CV}_k^{1/n}\neq\varnothing$.*

6.2 Converting short length to distance

Let π be the coarse projection $\pi\colon\mathrm{CV}_k\to\mathrm{FF}_k$.

Lemma 6.2 [2, Lemma 3.3] *Let $a\in F_k$ be a simple class and G a point in CV_k so that the loop corresponding to a in G intersects some edge $\le m$ times. Then*

$$d_{\mathrm{FF}_k}(\alpha,\pi(G))\le 6m+13,$$

where α is the smallest free factor containing the conjugacy class of a.

Using this lemma we prove the following:

Lemma 6.3 *Let α be a free factor containing the conjugacy class of an element $a\in F_k$ and G a point in CV_k, and suppose $\ell_G(a)\le m$. Then there is a constant B and a number A depending only on the rank of the free group such that*

$$d_{\mathrm{FF}_k}(\alpha,\pi(G))\le Am+B.$$

Proof Let e be the edge of G with the greatest length. Hence $\ell(e)\ge 1/(3k+3)$. Then α crosses e less than $(3k+3)m$ times. Therefore $d_{\mathrm{FF}_k}(\alpha,\pi(G))\le 6(3k+3)m+13$ by Lemma 6.2. Here $A=6(3k+3)$ clearly depends only on the rank of the free group. □

6.3 Proof of Theorem 1.2

Given $x \in F_C$, let us assume that $G \in \mathrm{CV}_k$ is a point which is projected to x. We will write $\pi(G) = x$. Let τ be the essential embedded torus in M corresponding to the given $\mathbb{Z}$–splitting T.

Let n be such that $d_{\mathrm{FF}_k}(G, D_T^n(G)) \leq C$. Up to replacing C by a constant we can assume that $n \geq 4$. By Theorem 5.2, we have $\mathrm{tw}_\tau(G, D_\tau^n(G)) \geq n-1$. Since $\mathrm{tw}_\tau(G, D_\tau^n(G)) = \mathrm{tw}_a(G, D_T^n(G))$, with $a \in F_k$ representing the core curve of the torus τ, we use the theorem of Clay and Pettet, Theorem 6.1, to deduce that there is a folding path $G_t\colon [0,d] \to \mathrm{CV}_k$ such that $G_0 = G$ and $G_d = D_\tau^n(G)$ and such that $\ell_{G_t}(a) \leq 1/(n-3)$ for some $t \in [0,d]$.

Now, by [2] and [18], the projection $\pi(\{G_t\})$ of the folding path $\{G_t\}_t$ onto FF_k is a quasigeodesic in FF_k between $\pi(D_\tau^n(G))$ and $\pi(G) = x$. Let α be the smallest free factor containing a. Since $n \geq 4$ we use Lemma 6.3 to deduce that

$$d_{\mathrm{FF}_k}(\alpha, \pi(G_t)) \leq \frac{A}{n-3} + 13,$$

where A is the same as it is given in the lemma. Since $\pi(\{G_t\})$ is uniformly Hausdorff-close to a geodesic and $d_{\mathrm{FF}_k}\big(\pi(G), \pi(D_\tau^n(G))\big) \leq C$, by the triangle inequality we have

$$d_{\mathrm{FF}_k}(\pi(G), \alpha) \leq \frac{A}{n-3} + C + H + 13 \leq A + C + H + 13,$$

where $H = H(k)$ is the distance between the geodesic and the unparametrized quasigeodesic $\pi(\{G_t\})$. Hence we have

$$C' = 2\,(A + C + H + 13). \qquad \square$$

7 Constructing fully irreducibles

In this section we will prove the main theorem of this paper.

Theorem 1.1 *Let T_1 and T_2 be two $\mathbb{Z}$–splittings of the free group F_k with rank $k > 2$ and α_1 and α_2 be two corresponding free factors in the free factor complex FF_k of the free group F_k. Let D_1 be a Dehn twist fixing α_1 and D_2 a Dehn twist fixing α_2, corresponding to T_1 and T_2, respectively. Then there exists a constant $N = N(k)$ such that whenever $d_{\mathrm{FF}_k}(\alpha_1, \alpha_2) \geq N$,*

1. *$\langle D_1, D_2 \rangle \simeq F_2$, and*
2. *all elements of $\langle D_1, D_2 \rangle$ which are not conjugate to the powers of D_1 and D_2 in $\langle D_1, D_2 \rangle$ are fully irreducible.*

We will start with some basic definitions and lemmata that are standard for δ–hyperbolic spaces.

Let (X, d) be a metric space. For $x, y, z \in X$, the *Gromov product* $(y, z)_x$ is defined as

$$(y, z)_x := \tfrac{1}{2}(d(y, x) + d(z, x) - d(y, z)).$$

If (X, d) is a δ–hyperbolic space, the initial segments of length $(y, z)_x$ of any two geodesics $[x, y]$ and $[x, z]$ stay close to each other. In other words, these geodesics are in 2δ–neighborhoods of each other. Hence the Gromov product measures how long two geodesics stay close together. This characterization of δ–hyperbolicity will be used in our work as the definition of being δ–hyperbolic for a metric space.

The Gromov product $(y, z)_x$ also approximates the distance between x and the geodesic $[y, z]$ within 2δ:

$$(y, z)_x \le d(x, [y, z]) \le (y, z)_x + 2\delta.$$

Definition 7.1 A path $\sigma\colon I \to X$ is called a (λ, ϵ)–quasigeodesic if σ is parametrized by arclength and if, for any $s_1, s_2 \in I$, we have

$$|s_1 - s_2| \le \lambda d(\sigma(s_1), \sigma(s_2)) + \epsilon.$$

If the restriction of σ to any subsegment $[a, b] \subset I$ of length at most ℓ is a $(\lambda, \mathcal{L})$–quasigeodesic, then we call σ an ℓ–local $(\lambda, \mathcal{L})$–quasigeodesic.

Let X be a geodesic metric space and $Y \subset X$. We say that Y is c–quasiconvex if for all $y_1, y_2 \in Y$ the geodesic segment $[y_1, y_2]$ lies in the c–neighborhood of Y.

For any $x \in X$ we call $p_x \in Y$ an *ϵ–quasiprojection* of x onto Y if

$$d(x, p_x) \le d(x, Y) + \epsilon.$$

The lemma below shows that in hyperbolic spaces, quasiprojections onto quasiconvex sets are *quasiunique*.

Lemma 7.2 *Let X be a δ–hyperbolic metric space and let $Y \subset X$ be c–quasiconvex. Let $x \in X$ and let p_x and $p_{x'}$ be two ϵ–quasiprojections of x onto Y. Then*

$$d(p_x, p_{x'}) \le 2c + 4\delta + 2\epsilon.$$

For a δ–hyperbolic geodesic G–space X, consider the almost fixed set $X_C(g)$ corresponding to a subgroup $\langle g \rangle$ for $g \in G$. Then the *quasiconvex hull* $\mathcal{X}_C(g)$ of $X_C(g)$ is defined to be the union of all geodesics connecting any two points of $X_C(g)$. From now on we will work with quasiconvex hulls of almost fixed sets. The following is standard for δ–hyperbolic spaces.

Lemma 7.3 [19, Lemma 3.9] *$\mathcal{X}_C(g)$ is g–invariant and 4δ–quasiconvex.*

The following lemma appears as Lemma 3.12 in [19].

Lemma 7.4 *Let (X,d) be a δ–hyperbolic space and let $[x_p,x_q]$ be a geodesic segment in X. Let $p,q \in X$ be such that x_p is a projection of p on $[x_p,x_q]$ and such that x_q is a projection of q on $[x_p,x_q]$. Then if $d(x_p,x_q) > 100\delta$, the path $[p,x_p]\cup[x_p,x_q]\cup[x_q,q]$ is a $(1,30\delta)$–quasigeodesic.*

In the proof of Theorem 1.3 we use Lemma 7.4, which assumes that the projections onto almost fixed sets exist. However, given a geodesic metric space X and $Y \subset X$ a subset which is not necessarily closed in X, the closest point projection onto Y may not exist. To fix this we will use quasiprojections which exist quasiuniquely when there is quasiconvexity, by Lemma 7.2.

Now we state and prove the following theorem, which essentially proves Theorem 1.1.

Theorem 1.3 *Let G be a group acting on a δ–hyperbolic metric space X by isometries and $\phi_1,\phi_2 \in G$. Suppose $C > 100\delta$ and the almost fixed sets $\mathcal{X}_C(\phi_1)$ and $\mathcal{X}_C(\phi_2)$ of $\langle\phi_1\rangle$ and $\langle\phi_2\rangle$, respectively, have diameters bounded above by a constant C'. Then there exists a constant C_1 such that, whenever $d_X(\mathcal{X}_C(\phi_1),\mathcal{X}_C(\phi_2)) \geq C_1$,*

1. *$\langle\phi_1,\phi_2\rangle \simeq F_2$, and*
2. *every element of $\langle\phi_1,\phi_2\rangle$ which is not conjugate to the powers of ϕ_1 and ϕ_2 in $\langle\phi_1,\phi_2\rangle$ acts loxodromically in X.*

Proof Let $p_1 \in \mathcal{X}_C(\phi_1)$ and $p_2 \in \mathcal{X}_C(\phi_2)$ be two points such that

$$d_X(p_1,p_2) = d_X(\mathcal{X}_C(\phi_1),\mathcal{X}_C(\phi_2)).$$

To prove the theorem we will pick a random word ω and construct a ping-pong argument involving the sets $\mathcal{X}_C(\phi_1)$ and $\mathcal{X}_C(\phi_2)$. The goal is to show that the iterates of $[p_1,p_2]$ under ω give a local quasigeodesic, hence a quasigeodesic. Without loss of generality, for $g \in \langle\phi_1\rangle$, we will start with proving that the path $[p_2,p_1]\cup[p_1,gp_1]\cup g[p_1,p_2]$ is a quasigeodesic.

Let $\pi(p_2)$ and $\pi(gp_2)$ be quasiprojections of the points p_2 and gp_2 on the geodesic segment $[p_1,gp_1]$. Then p_1 and $\pi(p_2)$ are both 4δ–quasiprojections. This is true for $\pi(gp_2)$ and gp_1 also. Since the difference is negligible we will assume p_1 and gp_1 are closest point projections.

Now we prove that $d(p_1, gp_1) \geq C$. To see this suppose that $d_X(p_1, gp_1) < C$. We take a point x on the geodesic segment $[p_1, p_2]$ such that $d(x, p_1) < \epsilon$, where $\epsilon = \frac{1}{2}(C - d(p_1, gp_1))$. Then we have

$$d(x, gx) \leq 2\epsilon + d(p_1, gp_1) = C,$$

which contradicts the assumption that p_1 is the closest point of $\mathcal{X}_C(\phi_1)$ to $\mathcal{X}_C(\phi_2)$. Since we proved that $d(p_1, gp_1) \geq C$, we apply Lemma 7.4 to conclude that the path $[p_2, p_1] \cup [p_1, gp_1] \cup g[p_1, p_2]$ is a quasigeodesic.

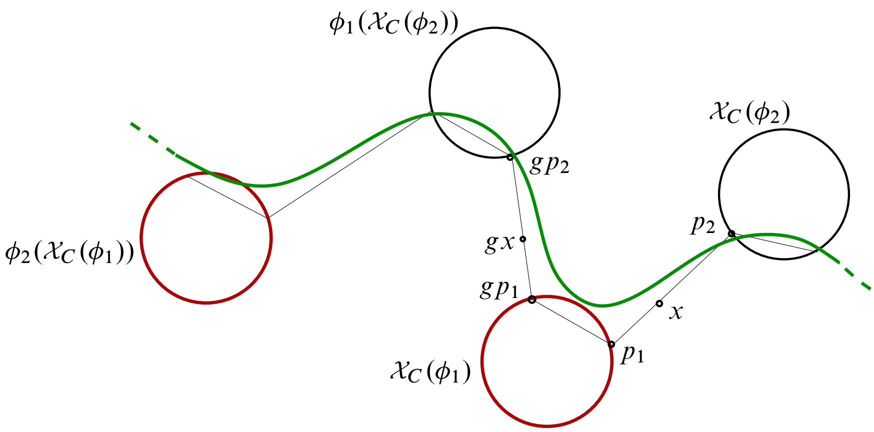

Figure 7: The ping-pong sets and a quasigeodesic between iterated points

Given two points $p_1 \in \mathcal{X}_C(\phi_1)$ and $p_2 \in \mathcal{X}_C(\phi_2)$ as above, a geodesic connecting them is called a *bridge* and it is not unique. However, by [19, Lemma 5.2], when two sets are sufficiently far apart it is *almost unique*. Hence we first assume that

$$d(p_1, p_2) \geq C.$$

Since $C > 100\delta$ this is sufficient to have a quasiunique bridge. Now we take a word $\omega = \phi_2^{m_l} \phi_1^{s_l} \cdots \phi_2^{m_1} \phi_1^{s_1}$ and consider the iterates of the quasiunique bridge $[p_1, p_2]$ under ω.

It is known that by the hyperbolicity of X, given $(\lambda, \mathcal{L})$, there exists $\ell > 0$ such that an ℓ–local $(\lambda, \mathcal{L})$–quasigeodesic is a $(\lambda', \mathcal{L}')$–quasigeodesic, where $\lambda' = \lambda'(\lambda, \mathcal{L}, \ell)$ and $L' = L'(\lambda, L, \ell)$. Since we have $(1, 30\delta)$–quasigeodesic pieces, we have such an ℓ which satisfies $\ell = \ell(30\delta)$. Hence we let

$$d(p_1, p_2) \geq C_1 := \max\{100\delta, \ell\}.$$

Now consisting of previously given quasigeodesic pieces, the path

$$\gamma := [p_1, p_2] \cup [p_1, \phi_1^{s_1} p_1] \cup [\phi_1^{s_1} p_1, \phi_1^{s_1} p_2] \cup [\phi_1^{s_1} p_2, \phi_2^{m_1} \phi_1^{s_1} p_2] \cup \cdots \cup [\omega p_1, \omega p_2]$$

is an ℓ–local $(1, 100\delta)$–quasigeodesic, and as a result, it is a quasigeodesic.

In particular, for any word ω in $\langle \phi_1, \phi_2 \rangle$ we have $d(\omega(x), x) \geq |\omega|$, where $|\omega|$ denotes the syllable length, up to conjugation. Now, it follows that $\langle \phi_1, \phi_2 \rangle$ is free. Since the path which is obtained by iterating any segment between two almost fixed sets under ω is a quasigeodesic, ω is loxodromic. □

Proof of Theorem 1.1 Consider the action of $\mathrm{Out}(F_k)$ on the free factor complex FF_k. By Theorem 1.2, for a sufficiently large constant $C = C(k)$ there exists a $C' = C'(C)$ such that the diameter of the almost fixed set of a Dehn twist is bounded above by C'. Let C_1 be the constant from Theorem 1.3.

Now assume that D_1 and D_2 are the Dehn twists so that $d_{\mathrm{FF}_k}(\alpha_1, \alpha_2) \geq 2C' + C_1$, where α_1 and α_2 are the projections of the given $\mathbb{Z}$–splittings T_1 and T_2 to FF_k, respectively. Since Theorem 1.2 implies $\mathrm{diam}\{F_C\} \leq A + C + H + 13 = C'$, we have

$$d(F_C(D_1), F_C(D_2)) \geq C_1.$$

Hence Theorem 1.3 applies to $\langle D_1, D_2 \rangle$ with $N = 2\,\mathrm{diam}\{F_C\} + C_1 = 2C' + C_1$. It is clear that $N = N(k)$. As a result, since loxodromically acting elements in the free factor complex are fully irreducible, every element from the group $\langle D_1, D_2 \rangle$ is either conjugate to powers of the twists or fully irreducible. □

8 Constructing atoroidal fully irreducibles

In this section we prove the following theorem which produces *atoroidal* fully irreducibles. Recall $\mathrm{FF}_k^{(1)}$ is the 1–skeleton of the free factor complex (free factor graph) and that $\sigma\colon \mathrm{FF}_k^{(1)} \to \mathcal{P}_k$ is a coarse surjective map and that both graphs have the same vertex sets.

Theorem 1.4 *Let T_1 and T_2 be two $\mathbb{Z}$–splittings of F_k with $k > 2$ with corresponding free factors α_1 and α_2, and let D_1 and D_2 be two Dehn twists corresponding to T_1 and T_2, respectively. Then there exists a constant $N_2 = N_2(k)$ such that $\langle D_1, D_2 \rangle \simeq F_2$ whenever $d_{\mathcal{P}_k}(\sigma(\alpha_1), \sigma(\alpha_2)) \geq N_2$, and all elements from this group which are not conjugate to the powers of D_1 and D_2 in $\langle D_1, D_2 \rangle$ are atoroidal fully irreducible.*

Proof To see that $\langle D_1, D_2\rangle \simeq F_2$ we use Theorem 1.3. To do that, we need to show that given a $\mathbb{Z}$–splitting T and a constant C there is another constant depending only on C which bounds from above the diameter of the almost fixed set $\mathcal{P}_C$ of a Dehn twist corresponding to T.

Let x be a point in $\mathcal{P}_C$ and $\sigma\pi(G) = x$ for some $G \in \mathrm{CV}_k$. Then, as before, we use the theorem of Clay and Pettet, Theorem 6.1, to obtain a folding line $\{G_t\}_t$ between G and $D^n_\tau(G)$ along which a is short, where a is the generator of the fundamental group of the torus τ in M. Then, since distances in the intersection graph are shorter, the lemma of Bestvina and Feighn, Lemma 6.2, applies, and

$$d_{\mathcal{P}_k}(\sigma\alpha, \sigma\pi(G)) \le d_{\mathrm{FF}_k}(\alpha, \pi(G)) \le 6m + 13.$$

Hence we can convert the short length to distance in the intersection graph. Thus there exists constants A and B such that

$$d_{\mathcal{P}_k}(\alpha, \pi(G_t)) \le Am + B,$$

where G_t is the point along which a is short. Then the rest of the proof follows the same since the image of the folding line in $\mathcal{P}_k$ is a quasigeodesic and it is Hausdorff-close to a geodesic by [18] and [23]. Hence the diameter of $\mathcal{P}_C$ is uniformly bounded above by a constant, and Theorem 1.3 applies to $\langle D_1, D_2\rangle$ with $N_2 = 2\,\mathrm{diam}\{\mathcal{P}_C\} + C_1$.

Since in $\mathcal{P}_k$ loxodromically acting automorphisms are atoroidal fully irreducible [23], an element of $\langle D_1, D_2\rangle$ which is not conjugate to the powers of D_1 and D_2 in $\langle D_1, D_2\rangle$ is *atoroidal* fully irreducible. □

References

[1] **J Behrstock**, **M Bestvina**, **M Clay**, *Growth of intersection numbers for free group automorphisms*, J. Topol. 3 (2010) 280–310 MR

[2] **M Bestvina**, **M Feighn**, *Hyperbolicity of the complex of free factors*, Adv. Math. 256 (2014) 104–155 MR

[3] **M Bestvina**, **M Feighn**, *Subfactor projections*, J. Topol. 7 (2014) 771–804 MR

[4] **M Bestvina**, **M Handel**, *Train tracks and automorphisms of free groups*, Ann. of Math. 135 (1992) 1–51 MR

[5] **M Clay**, **A Pettet**, *Twisting out fully irreducible automorphisms*, Geom. Funct. Anal. 20 (2010) 657–689 MR

[6] **M Clay**, **A Pettet**, *Relative twisting in outer space*, J. Topol. Anal. 4 (2012) 173–201 MR

[7] **M Culler**, **K Vogtmann**, *Moduli of graphs and automorphisms of free groups*, Invent. Math. 84 (1986) 91–119 MR

[8] **S Francaviglia**, **A Martino**, *Metric properties of outer space*, Publ. Mat. 55 (2011) 433–473 MR

[9] **I Gruschko**, *On the bases of a free product of groups*, Rec. Math. [Mat. Sbornik] N.S. 8 (50) (1940) 169–182 MR In Russian

[10] **V Guirardel**, *Cœur et nombre d'intersection pour les actions de groupes sur les arbres*, Ann. Sci. École Norm. Sup. 38 (2005) 847–888 MR

[11] **F Gültepe**, *Normal tori in* $\sharp_n(S^2 \times S^1)$, Topology Appl. 160 (2013) 953–959 MR

[12] **H Hamidi-Tehrani**, *Groups generated by positive multi-twists and the fake lantern problem*, Algebr. Geom. Topol. 2 (2002) 1155–1178 MR

[13] **A Hatcher**, *Homological stability for automorphism groups of free groups*, Comment. Math. Helv. 70 (1995) 39–62 MR

[14] **A Hatcher**, **D McCullough**, *Finite presentation of* 3*–manifold mapping class groups*, from "Groups of self-equivalences and related topics" (R A Piccinini, editor), Lecture Notes in Math. 1425, Springer (1990) 48–57 MR

[15] **A Hatcher**, **K Vogtmann**, *The complex of free factors of a free group*, Quart. J. Math. Oxford Ser. 49 (1998) 459–468 MR

[16] **C Horbez**, *Sphere paths in outer space*, Algebr. Geom. Topol. 12 (2012) 2493–2517 MR

[17] **I Kapovich**, **M Lustig**, *Geometric intersection number and analogues of the curve complex for free groups*, Geom. Topol. 13 (2009) 1805–1833 MR

[18] **I Kapovich**, **K Rafi**, *On hyperbolicity of free splitting and free factor complexes*, preprint (2012) arXiv

[19] **I Kapovich**, **R Weidmann**, *Freely indecomposable groups acting on hyperbolic spaces*, Internat. J. Algebra Comput. 14 (2004) 115–171 MR

[20] **H Kneser**, *Geschlossene Flächen in dreidimensionalen Mannigfaltigkeiten*, Jahresber. Dtsch. Math.-Ver. 38 (1929) 248–260

[21] **F Laudenbach**, *Topologie de la dimension trois: homotopie et isotopie*, Société Mathématique de France, Paris (1974) MR

[22] **G Levitt**, **M Lustig**, *Irreducible automorphisms of* F_n *have north-south dynamics on compactified outer space*, J. Inst. Math. Jussieu 2 (2003) 59–72 MR

[23] **B Mann**, *Some hyperbolic* $Out(F_N)$*–graphs and nonunique ergodicity of very small* F_N *trees*, PhD thesis, University of Utah (2014)

[24] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves, II: Hierarchical structure*, Geom. Funct. Anal. 10 (2000) 902–974 MR

[25] **P Scott**, **G A Swarup**, *Splittings of groups and intersection numbers*, Geom. Topol. 4 (2000) 179–218 MR

[26] **A Shenitzer**, *Decomposition of a group with a single defining relation into a free product*, Proc. Amer. Math. Soc. 6 (1955) 273–279 MR

[27] **J R Stallings**, *A topological proof of Grushko's theorem on free products*, Math. Z. 90 (1965) 1–8 MR

[28] **G A Swarup**, *Decompositions of free groups*, J. Pure Appl. Algebra 40 (1986) 99–102 MR

[29] **S J Taylor**, *A note on subfactor projections*, Algebr. Geom. Topol. 14 (2014) 805–821 MR

[30] **W P Thurston**, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Amer. Math. Soc. 19 (1988) 417–431 MR

[31] **K Vogtmann**, *Automorphisms of free groups and outer space*, Geom. Dedicata 94 (2002) 1–31 MR

[32] **R D Wade**, *Folding free-group automorphisms*, Q. J. Math. 65 (2014) 291–304 MR

Department of Mathematics, University of Illinois at Urbana-Champaign
1409 W Green St, Urbana, IL 61801, United States

fgultepe@illinois.edu

http://www.math.illinois.edu/~fgultepe

Received: 28 January 2015 Revised: 19 May 2016

Algebraic & Geometric Topology 17 (2017) 1407–1444

Pair of pants decomposition of 4–manifolds

MARCO GOLLA
BRUNO MARTELLI

Using tropical geometry, Mikhalkin has proved that every smooth complex hypersurface in $\mathbb{CP}^{n+1}$ decomposes into *pairs of pants*: a pair of pants is a real compact $2n$–manifold with cornered boundary obtained by removing an open regular neighborhood of $n+2$ generic complex hyperplanes from $\mathbb{CP}^n$.

As is well-known, every compact surface of genus $g \geqslant 2$ decomposes into pairs of pants, and it is now natural to investigate this construction in dimension 4. Which smooth closed 4–manifolds decompose into pairs of pants? We address this problem here and construct many examples: we prove in particular that every finitely presented group is the fundamental group of a 4–manifold that decomposes into pairs of pants.

57M99, 57N13

Introduction

The decomposition of surfaces into pairs of pants is an extraordinary instrument in geometric topology that furnishes, among many other things, a nice parametrization for Teichmüller spaces. Mikhalkin has generalized this notion in [4] to all even dimensions as follows: he defines the $2n$*–dimensional pair of pants* as the manifold obtained by removing $n+2$ generic hyperplanes from $\mathbb{CP}^n$. One actually removes open regular neighborhoods of the hyperplanes to get a compact real $2n$–manifold with stratified cornered boundary: when $n=1$, we get $\mathbb{CP}^1$ minus three points, whence the usual pair of pants.

Using some beautiful arguments from tropical geometry, Mikhalkin has proved in [4] that every smooth complex hypersurface in $\mathbb{CP}^{n+1}$ decomposes into pairs of pants. We address here the following natural question:

Question 0.1 Which smooth closed manifolds decompose into pairs of pants?

The question makes sense, of course, only for real smooth manifolds of even dimension $2n$. It is natural to expect the existence of many smooth manifolds that decompose into pairs of pants and are not complex projective hypersurfaces: for instance, the

Published: 17 July 2017 DOI: 10.2140/agt.2017.17.1407

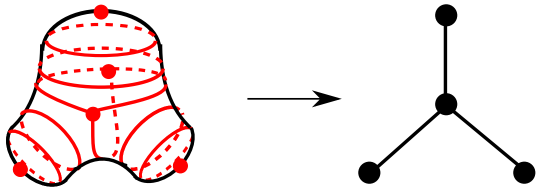

Figure 1: The model fibration $\mathbb{CP}^1 \to \Pi_1$

hypersurfaces of $\mathbb{CP}^2$ are precisely the closed orientable surfaces of genus $g = \frac{1}{2}(d-1)(d-2)$ for some $d > 0$, while by assembling pairs of pants, we obtain closed orientable surfaces of any genus.

In this paper, we study pants decompositions in (real) dimension 4. We start by constructing explicit pants decompositions for some simple classes of closed 4–manifolds: S^4, torus bundles over surfaces, circle bundles over 3–dimensional graph manifolds, toric manifolds, the simply connected manifolds $\#_k(S^2 \times S^2)$ and $\#_k\mathbb{CP}^2\#_h\overline{\mathbb{CP}}^2$. Then we prove the following theorem, which shows that the 4–manifolds that decompose into pairs of pants form a quite large class:

Theorem 0.2 *Every finitely presented group is the fundamental group of a closed* 4*–manifold that decomposes into pairs of pants.*

We get in particular plenty of non-Kähler, and hence nonprojective, 4–manifolds. We expose a more detailed account of these results in the remaining part of this introduction.

Pair of pants decompositions Mikhalkin's definition of a pair of pants decomposition is slightly more flexible than the usual one adopted for surfaces: the boundary of a pair of pants (of any dimension $2n$) is naturally stratified into circle fibrations, and an appropriate collapse of these circles is allowed. With this language, the sphere S^2 has a pants decomposition with a single pair of pants, where each boundary component is collapsed to a point.

More precisely, a *pair of pants decomposition* of a closed $2n$–manifold M^{2n} is a fibration $M^{2n} \to X^n$ over a compact n–dimensional cell complex X^n which is locally diffeomorphic to a *model fibration* $\mathbb{CP}^n \to \Pi_n$ derived from tropical geometry. The fiber of a generic (smooth) point in X^n is a real n–torus.

When $n = 1$, the model fibration is drawn in Figure 1, and some examples of pair of pants decompositions are shown in Figure 2. The reader is invited to look at these pictures that, although quite elementary, describe some phenomena that will also be present in higher dimensions. When $n = 1$, the base cell complex X^1 may be of these

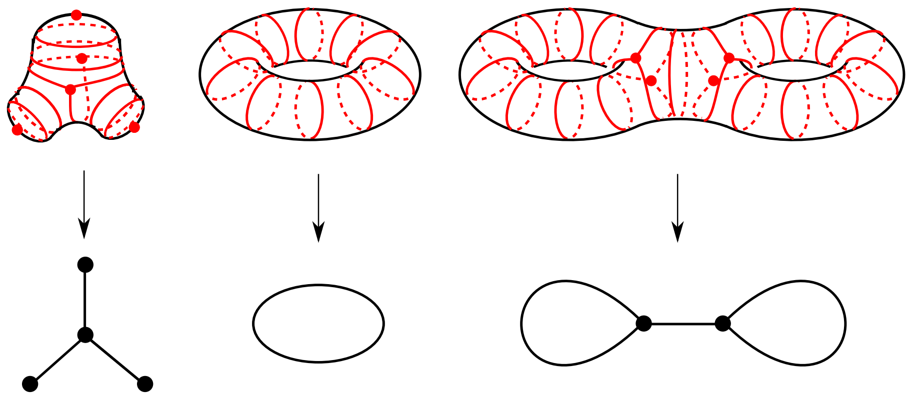

Figure 2: Pair of pants decompositions of surfaces

limited types: either a circle, or a graph with vertices of valence 1 and 3. There are three types of points x in X^1 (smooth, a vertex with valence 1, or a vertex with valence 3), and the fiber over x depends only on its type (a circle, a point, or a θ–shaped graph, respectively).

In dimension $2n$, the model cell complex Π_n is homeomorphic to the cone of the $(n-1)$–skeleton of the $(n+1)$–simplex; the model fibration sends $n+2$ generic hyperplanes onto the base $\partial\Pi_n$ of the cone and the complementary pair of pants onto its interior $\Pi_n \setminus \partial\Pi_n$. We are interested here in the case $n = 2$.

Dimension 4 In dimension 4, a pair of pants is $\mathbb{CP}^2$ minus (the open regular neighborhood of) four generic lines: it is a 4–dimensional compact manifold with cornered boundary; the boundary consists of six copies of $P \times S^1$, where P is the usual 2–dimensional pair-of-pants, bent along six 2–dimensional tori.

The model fibration $\mathbb{CP}^2 \to \Pi_2$ is sketched in Figure 3: the cell complex Π_2 is homeomorphic to the cone over the 1–skeleton of a tetrahedron, and there are 6 types of points in Π_2; the fiber of a generic (ie smooth) point of Π_2 is a torus, and the fibers over the other 5 types are: a point, S^1, θ, $\theta \times S^1$, and a more complicated 2–dimensional cell complex F_2 fibering over the central vertex of Π_2. The central fiber F_2 is a spine of the 4–dimensional pair of pants and is homotopically equivalent to a punctured 3–torus. (Likewise, when $n = 1$, the fiber of the central vertex in $\mathbb{CP}^1 \to \Pi_1$ is a θ–shaped spine of the 2–dimensional pair of pants and is homotopically equivalent to a punctured 2–torus.)

A pair of pants decomposition of a closed 4–manifold M^4 is a map $M^4 \to X^2$ locally diffeomorphic to this model. The cell complex X^2 is locally diffeomorphic to Π_2, and the fiber over a generic point of X^2 is a torus.

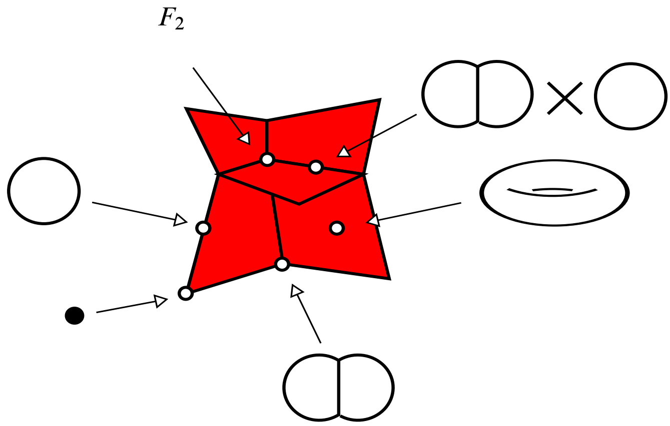

Figure 3: Fibers of the model fibration $\mathbb{CP}^2 \to \Pi_2$

We note that pants decompositions are similar to (but different from) Turaev's *shadows* (see [7] and also Costantino and Thurston [1]), which are fibrations $M^4 \to X^2$ onto similar cell complexes, where the generic fiber is a disc and M^4 is a 4–manifold with boundary that collapses onto X^2.

The main object of this work is to introduce many examples of 4–manifolds that decompose into pairs of pants. These examples are far from being exhaustive, and we are very far from having a satisfactory answer to Question 0.1: for instance, we are not aware of any obstruction to the existence of a pants decomposition, so we do not know if there is a closed 4–manifold which does not admit one.

Sketch of the proof of Theorem 0.2 Once we set up the general theory, Theorem 0.2 is proved as follows. We first solve the problem of determining all the possible fibrations $M^4 \to X^2$ on a given X^2 by introducing an appropriate system of *labelings* on X^2. We note that the same X^2 may admit fibrations $M^4 \to X^2$ of different kinds, sometimes with pairwise nondiffeomorphic total spaces M^4, and each such fibration is detected by some labeling on X^2. This combinatorial encoding is interesting on its own because it furnishes a complete presentation of all pants decompositions in dimension 4.

We then use these labelings to construct a large class of complexes X^2 for which there exist fibrations $M^4 \to X^2$ that induce isomorphisms on fundamental groups. Finally, we show that every finitely presented G has an X^2 in this class with $\pi_1(X^2) = G$.

Structure of the paper We introduce pair of pants decompositions in all dimensions in Section 1, following and expanding from Mikhalkin [4] and focusing mainly on the 4–dimensional case. In Section 2, we construct some examples.

In Section 3, we study in detail the simple case when X is a polygon. In this case, $M \to X$ looks roughly like the moment map on a toric manifold, and every fibration $M \to X$ is encoded by some labeling on X. We then extend these labelings to more general complexes X in Section 4.

In Section 5, we prove Theorem 0.2.

Acknowledgements We thank Filippo Callegaro and Giovanni Gaiffi for stimulating discussions, and the referee for helpful comments and for finding a mistake in an earlier proof of Theorem 0.2.

1 Definitions

We introduce here simple complexes, tropical fibrations, and pair of pants decompositions. We describe these objects with some detail in dimensions 2 and 4.

Recall that a *subcomplex* $X \subset M$ of a smooth manifold M is a subcomplex of some smooth triangulation of M.

We work in the category of smooth manifolds: all the objects we consider are subcomplexes of some $\mathbb{R}^N$, and a map between two such complexes is smooth if it locally extends to a smooth map on some open set.

1.1 The basic cell complex Π_n

Let Δ be the standard $(n+1)$–simplex

$$\Delta = \{(x_0, \dots, x_{n+1}) \in \mathbb{R}^{n+2} \mid x_0 + \cdots + x_{n+1} = 1,\ x_i \geqslant 0\}.$$

We use the barycentric coordinates on Δ; that is, for every nonzero vector $x = (x_0, \dots, x_{n+1}) \in \mathbb{R}^{n+2}_{\geqslant 0}$, we denote by $[x_0, \dots, x_{n+1}]$ the unique point in Δ that is a multiple of x. Every point $p \in \Delta$ has a unique description as $[x_0, \dots, x_{n+1}]$ with $\max x_i = 1$, and we call it the *normal form* of p.

Definition 1.1 Let $\Pi_n \subset \Delta$ be the following subcomplex:

$$\Pi_n = \{[x_0, \dots, x_{n+1}] \mid 0 \leqslant x_i \leqslant 1 \text{ and } x_i = 1 \text{ for at least two values of } i\}.$$

The subcomplex $\Pi_n \subset \Delta$ may be interpreted as the cut-locus of the vertices of Δ; see Π_1 in Figure 4. Every point $x \in \Pi_n$ has a *type* (k, l) with $0 \leqslant k \leqslant l \leqslant n$, which is determined by the following requirements: the normal form of x contains $l - k + 2$

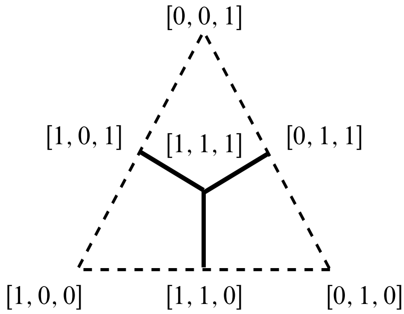

Figure 4: The subcomplex Π_1 inside the standard simplex Δ

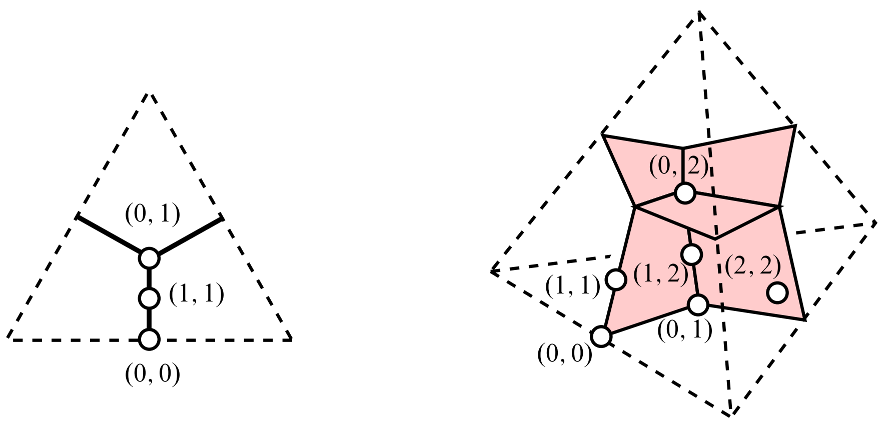

Figure 5: The subcomplexes Π_1 and Π_2. Every point is of some type (k, l) with $0 \leqslant k \leqslant l \leqslant n$, and points of the same type define strata. Here k is the dimension of the stratum and $l+1$ is the dimension of the face of Δ containing it.

different 1–entries and $n-l$ different 0–entries. More concretely, see Figure 5 for the cases $n = 1$ and 2, which are of interest for us here.

Points of the same type (k, l) form some open k–cells, and these cells stratify Π_n. Geometrically, a point x of type (k, l) is contained in the k–stratum of Π_n and in the $(l+1)$–stratum of Δ. An open star neighborhood of x in Π_n is diffeomorphic to the subcomplex

$$\Pi_{l,k} = \mathbb{R}^k \times \Pi_{l-k} \times [0, +\infty)^{n-l}.$$

A point of type (n, n) is a *smooth* point, while the points with $l < n$ form the *boundary*

$$\partial \Pi_n = \Pi_n \cap \partial \Delta.$$

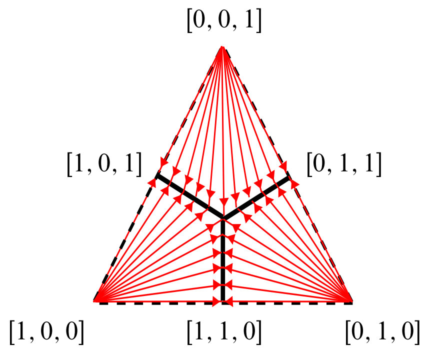

Figure 6: The projection of Δ^* onto Π_n

1.2 Tropical fibration of $\mathbb{CP}^n$

Using tropical geometry, Mikhalkin has constructed in [4] a map

$$\pi\colon \mathbb{CP}^n \to \Pi_n.$$

The map is defined as a composition $\pi = \pi_2 \circ \pi_1$ of two projections. The first one is a restriction of the projection

$$\mathbb{CP}^{n+1} \xrightarrow{\pi_1} \Delta, \quad [z_0, \dots, z_{n+1}] \mapsto [|z_0|, \dots, |z_{n+1}|].$$

We identify $\mathbb{CP}^n$ with the hyperplane $H \subset \mathbb{CP}^{n+1}$ defined by $z_0 + \cdots + z_{n+1} = 0$ and restrict π_1 to H. The image $\pi_1(H)$ is a region in Δ called an *amoeba*, which contains Π_n as a spine [5]. There is a simple projection that retracts the amoeba onto its spine Π_n: it is the restriction of a map

$$\pi_2\colon \Delta^* \to \Pi_n,$$

where Δ^* is Δ minus its vertices. The map π_2 is drawn in Figure 6 and is defined as follows: up to permuting the coordinates, we suppose for simplicity that $x = [x_0, \dots, x_{n+1}]$ with $x_0 \geqslant x_1 \geqslant \dots \geqslant x_{n+1}$, and we define

$$\pi_2(x) = [x_1, x_1, x_2, \dots, x_{n+1}].$$

The composition $\pi = \pi_2 \circ \pi_1$ is a map that sends $\mathbb{CP}^n = H$ onto Π_n.

The map π_2 is only piecewise smooth; it can then be smoothened as explained in [4, Section 4.3], so that the composition π is also smooth. In the following sections, we study π_2 in the cases $n = 1$ and $n = 2$ before the smoothening, because it is easier to determine the fibers of π concretely using the nonsmoothed version of π_2. We remark that in dimension 4, wherein lies our interest, every piecewise-linear object can be easily smoothened, so this will not be an important issue anyway.

1.3 The case $n = 1$

We now explicitly describe the fibration

$$\pi\colon \mathbb{CP}^1 \to \Pi_1.$$

Recall that $\mathbb{CP}^1$ is identified with the line $H = \{z_0 + z_1 + z_2 = 0\}$ in $\mathbb{CP}^2$, and that Π_1 contains points of type $(0,0)$, $(1,1)$ and $(0,1)$.

Proposition 1.2 *The fiber $\pi^{-1}(x)$ of a point $x \in \Pi_1$ is*

- *a point if x is of type $(0,0)$,*
- *a piecewise-smooth circle if x is of type $(1,1)$,*
- *a θ–shaped smooth graph if x is of type $(0,1)$.*

Proof Up to reordering we have $x = [1,1,0]$, $[1,1,t]$, or $[1,1,1]$ with $0 < t < 1$ depending on the type. Using the calculation made in Figure 7 (left) we can describe the fibers explicitly:

$$\begin{aligned}
\pi^{-1}([1,1,0]) &= [1,1,0],\\
\pi^{-1}([1,1,t]) &= \Big(\big\{[x,e^{i\theta},t] \;\big|\; |x| \geqslant 1\big\} \cup \big\{[e^{i\theta},x,t] \;\big|\; |x| \geqslant 1\big\}\Big) \cap H\\
&= \big\{[-e^{i\theta}-t,e^{i\theta},t] \;\big|\; \cos\theta \geqslant -\tfrac{t}{2}\big\} \cup \big\{[e^{i\theta},-e^{i\theta}-t,t] \;\big|\; \cos\theta \geqslant -\tfrac{t}{2}\big\},\\
\pi^{-1}([1,1,1]) &= \Big(\big\{[x,e^{i\theta},1] \;\big|\; |x| \geqslant 1\big\} \cup \big\{[e^{i\theta},1,x] \;\big|\; |x| \geqslant 1\big\}\\
&\qquad\qquad \cup \big\{[1,x,e^{i\theta}] \;\big|\; |x| \geqslant 1\big\}\Big) \cap H\\
&= \big\{[-e^{i\theta}-1,e^{i\theta},1] \;\big|\; \cos\theta \geqslant -\tfrac{1}{2}\big\} \cup \big\{[e^{i\theta},1,-e^{i\theta}-1] \;\big|\; \cos\theta \geqslant -\tfrac{1}{2}\big\}\\
&\qquad\qquad \cup \big\{[1,-e^{i\theta}-1,e^{i\theta}] \;\big|\; \cos\theta \geqslant -\tfrac{1}{2}\big\}.
\end{aligned}$$

The fiber $\pi^{-1}([1,1,t])$ consists of two arcs with disjoint interiors but coinciding endpoints $[e^{\pm i\theta}, e^{\mp i\theta}, t]$ with $\cos\theta = -t/2$; therefore, $\pi^{-1}([1,1,t])$ is a piecewise smooth circle. Analogously $\pi^{-1}([1,1,1])$ consists of three arcs joined at their endpoints $[e^{\pm 2\pi i/3}, e^{\mp 2\pi i/3}, 1]$ to form a θ–shaped graph. □

The fibration π is homeomorphic to the one drawn in Figure 8. The smoothing described in [4, Section 4.3] transforms the piecewise smooth circles into smooth circles, so that the resulting fibration is diffeomorphic to the one shown in the picture.

We note that the θ–shaped graph is a spine of the pair of pants, and is also homotopic to a once-punctured 2–torus. Both of these facts generalize to higher dimensions.

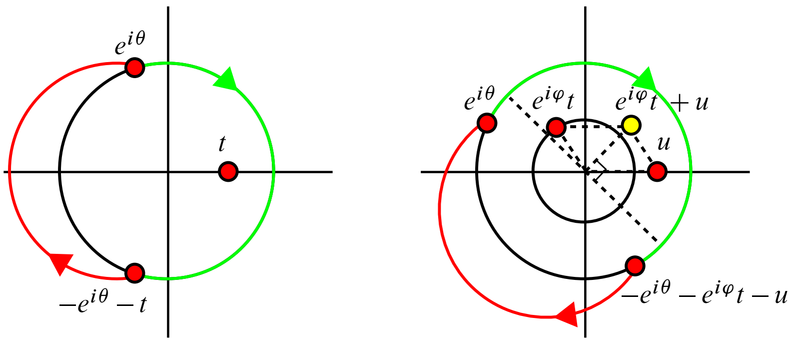

Figure 7: The points $[-e^{i\theta}-t, e^{i\theta}, t]$ with $|-e^{i\theta}-t| \geqslant 1$: when $\cos\theta = -t/2$, we get $-e^{i\theta}-t = e^{-i\theta}$; as the point $e^{i\theta}$ moves along the green arc of the unit circle, the point $-e^{i\theta}-t$ moves along the red arc, and hence has norm bigger than 1. This identifies one of the two arcs in $\pi^{-1}([1,1,t])$ (left). The fiber $\pi^{-1}([1,1,t,u])$ is considered similarly, with $e^{i\varphi}t+u$ instead of t (right).

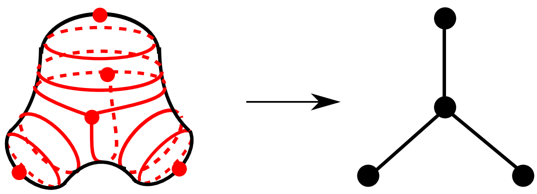

Figure 8: The tropical fibration $\mathbb{CP}^1 \to \Pi_1$

1.4 The case $n = 2$

We now study the fibration $\pi\colon \mathbb{CP}^2 \to \Pi_2$, and our main goal is to show that its fibers are as in Figure 3.

Recall that we identify $\mathbb{CP}^2$ with the plane $H = \{z_0+z_1+z_2+z_3 = 0\}$ in $\mathbb{CP}^3$. The subcomplex Π_2 has points of type $(0,0)$, $(1,1)$ and $(0,1)$ on its boundary, and of type $(2,2)$, $(1,2)$ and $(0,2)$ in its interior.

Proposition 1.3 *The fiber $\pi^{-1}(x)$ of a point $x \in \Pi_2$ is*

- *a point if x is of type $(0,0)$,*
- *a piecewise-smooth circle if x is of type $(1,1)$,*
- *a θ–shaped smooth graph θ if x is of type $(0,1)$,*
- *a piecewise-smooth torus if x is of type $(2,2)$,*
- *a piecewise-smooth product $\theta \times S^1$ if x is of type $(1,2)$,*
- *some 2–dimensional cell complex F_2 if x is of type $(0,2)$.*

Proof Up to reordering, the point x is one of the following:

$$[1,1,0,0],\ [1,1,t,0],\ [1,1,1,0],\ [1,1,t,u],\ [1,1,1,t],\ [1,1,1,1],$$

with $1 > t \geqslant u > 0$.

Let f_1, f_2, f_3 and f_4 be the faces of Δ, with $f_i = \{x_i = 0\}$. The preimage $\pi_1^{-1}(f_i)$ is the plane $\{z_i = 0\}$ in $\mathbb{CP}^3$ and intersects H in a line l_i. The four lines l_1, l_2, l_3 and l_4 are in general position in H and intersect pairwise in the six points obtained by permuting the coordinates of $[1,-1,0,0]$.

The map π sends l_i onto the Y–shaped graph $f_i \cap \Pi_2$ exactly as described in the previous section; see Figure 8. The map π sends the four lines l_i onto $\partial\Pi_2$, each line projected onto its own Y–shaped graph; the six intersection points are sent bijectively to the six points of type $(0,0)$ of Π_2.

It remains to understand the map π over the interior of Π_2. Similarly as in the 1–dimensional case, Figure 7 (right) shows that

$$\begin{aligned}
&\pi^{-1}([1,1,t,u])\\
&\quad= \left(\left\{[x, e^{i\theta}, e^{i\varphi}t, u] \,\middle|\, |x| \geqslant 1\right\} \cup \left\{[e^{i\theta}, x, e^{i\varphi}t, u] \,\middle|\, |x| \geqslant 1\right\}\right) \cap H\\
&\quad= \left\{[-e^{i\theta} - e^{i\varphi}t - u, e^{i\theta}, e^{i\varphi}t, u] \,\middle|\, \cos(\theta - \arg(e^{i\varphi}t + u)) \geqslant -\tfrac{1}{2}|e^{i\varphi}t + u|\right\}\\
&\quad\cup \left\{[e^{i\theta}, -e^{i\theta} - e^{i\varphi}t - u, e^{i\varphi}t, u] \,\middle|\, \cos(\theta - \arg(e^{i\varphi}t + u)) \geqslant -\tfrac{1}{2}|e^{i\varphi}t + u|\right\}.
\end{aligned}$$

For every fixed $e^{i\varphi} \in S^1$, we get two arcs parametrized by θ with the same endpoints, thus forming a circle as in the 1–dimensional case. Therefore, the fiber over $[1,1,t,u]$ is a (piecewise smooth) torus.

Analogously, the fiber over $[1,1,1,t]$ is a piecewise smooth product of a θ–shaped graph and S^1. Finally, the fiber over $[1,1,1,1]$ is a more complicated 2–dimensional cell complex F_2. □

The different fibers are shown in Figure 3. Let F_i be the fiber over a point of type $(0,i)$. The fibers F_0, F_1 and F_2 are a point, a θ–shaped graph and some 2–dimensional complex. These fibers "generate" all the others: the fiber over a point of type (k,l) is piecewise-smoothly homeomorphic to $F_l \times (S^1)^k$.

1.5 More on dimension 4

The fibration $\mathbb{CP}^2 \to \Pi_2$ plays the main role in this work, and we need to fully understand it. We consider here a couple of natural questions.

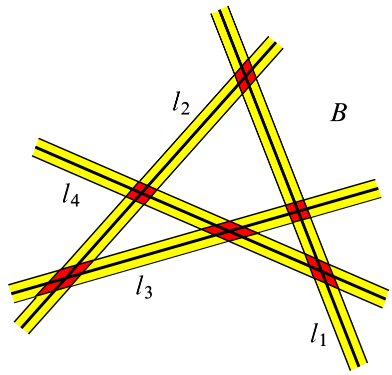

Figure 9: A regular neighborhood of the four lines. It decomposes into six pieces diffeomorphic to $D^2 \times D^2$ (red) and four pieces diffeomorphic to $P \times D^2$ (yellow), where P is a 2–dimensional pair of pants. (Every yellow piece is a D^2–bundle over P, and every such bundle is trivial. Note however that the normal bundle of each line is not trivial.)

How does the fibration look on a collar of $\partial\Pi_2$? It sends a regular neighborhood of the four lines l_1, l_2, l_3 and l_4 shown in Figure 9 onto a regular neighborhood of $\partial\Pi_2$ as drawn in Figure 10. Note that the regular neighborhood of the lines decomposes into pieces diffeomorphic to $D^2 \times D^2$ and $P \times D^2$, where P is a pair of pants; see Figure 9. On the red regions, the fibration sends $D^2 \times D^2$ to $[0,1]\times[0,1]$ as $(w,z) \mapsto (|w|,|z|)$. On the yellow zone, each piece $P \times D^2$ is sent to $Y \times [0,1]$ as $(x,z) \mapsto (\pi(x),|z|)$, where Y is a Y–shaped graph.

What is the fiber F_2? By construction, it is a 2–dimensional spine of $\mathbb{CP}^2$ minus the four lines. It is a well-known fact (proved for instance using the Salvetti complex [6]) that the complement of four lines in general position in $\mathbb{CP}^2$ is homotopically equivalent to a punctured 3–torus. More generally, the fiber F_n is homotopic to a once-punctured $(n+1)$–torus (compare the case $n=1$). We have determined F_2 only up to homotopy, but this is sufficient for us.

1.6 Simple complexes

Always following Mikhalkin, we use the fibration π as a standard model to define more general fibrations of manifolds onto complexes.

Definition 1.4 A *simple n–dimensional complex* is a compact connected space $X \subset \mathbb{R}^N$ such that every point has a neighborhood diffeomorphic to an open subset of Π_n.

For example, a simple 1–dimensional complex is either a circle or a graph with vertices of valence 1 and 3.

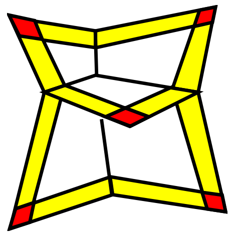

Figure 10: A regular neighborhood of the four lines projects onto a regular neighborhood of $\partial\Pi_2$. Yellow and red blocks from Figure 9 project to the yellow and red portions in Π_2 drawn here. Note that there is a sixth sheet with a sixth red block behind the five that are shown.

Every point in X inherits a type (k,l) from Π_n, and points of the same type form a k–manifold called the (k,l)–*stratum* of X. As opposed to Π_n, a connected component of a (k,l)–stratum need not be a cell: for instance, a closed smooth n–manifold is a simple complex where every point is smooth, ie of type (n,n).

We use the word "simple" because it is largely employed to denote 2–dimensional complexes with generic singularities; see for instance [3].

1.7 Pants decomposition

Let M be a closed smooth manifold of dimension $2n$. Following [4], we define a *pants decomposition* for M to be a map

$$p\colon M \to X$$

over a simple n–dimensional complex X which is locally modeled on the fibration $\pi\colon \mathbb{CP}^n \to \Pi_n$; that is, the following holds: for every point $x \in X$ there are an open neighborhood U of x, a point y in an open subset $V \subset \Pi_n$, a diffeomorphism $(U,x) \to (V,y)$, and a fiber-preserving diffeomorphism $\pi^{-1}(V) \to p^{-1}(U)$ such that the resulting diagram commutes:

$$\begin{tikzcd} \pi^{-1}(V) \arrow[r] \arrow[d, "\pi"'] & p^{-1}(U) \arrow[d, "p"] \\ V \arrow[r] & U \end{tikzcd}$$

When $n=1$, a pants decomposition is a fibration $p\colon M \to X$ of a closed surface onto a 1–dimensional simple complex. If X is not a circle and contains no 1–valent vertices, the fibration induces on M a pants decomposition in the usual sense: the complex X decomposes into Y–shaped subgraphs whose preimages in M are pairs

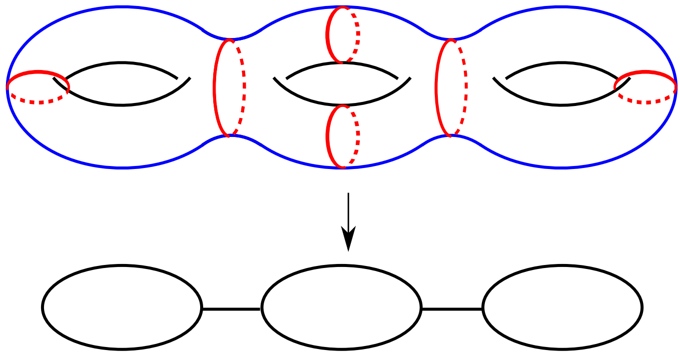

Figure 11: A pants decomposition of a surface M in the usual sense induces a fibration $M \to X$ onto a simple complex.

of pants; see Figure 11. Conversely, every usual pants decomposition of M defines a fibration $M \to X$ of this type.

In general, the base complex X may be quite flexible; for instance, it might just be an n–manifold. Therefore, every smooth n–torus fibration on a n–manifold X is a pants decomposition. Mikhalkin has proved the following remarkable result:

Theorem 1.5 (Mikhalkin [4]) *Every smooth complex hypersurface in* $\mathbb{CP}^{n+1}$ *admits a pants decomposition.*

As stated in the introduction, we would like to understand which manifolds of even dimension admit a pants decomposition. In dimension 2, every closed orientable surface has a pants decomposition. Those having genus $g > 1$ admit a usual one, while the sphere and the torus admit one in the more generalized sense introduced here: they fiber respectively over a segment (or a Y–shaped graph, or any tree) and a circle.

We now focus on the case $n = 2$; that is, we look at smooth 4–manifolds fibering over simple 2–dimensional complexes.

2 Four-manifolds

We now construct some closed 4–dimensional manifolds M that decompose into pairs of pants, that is, that admit a fibration $M \to X$ onto some simple complex X locally modeled on $\mathbb{CP}^2 \to \Pi_2$. In the subsequent sections, we will study fibrations on a given X in a more systematic way.

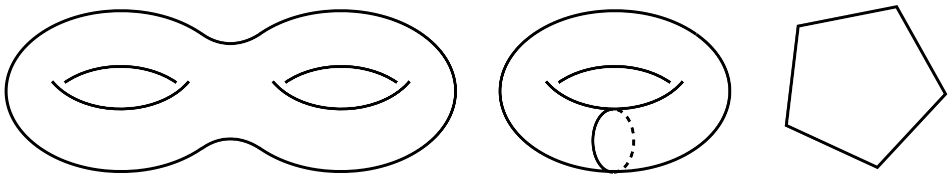

Figure 12: Three simple 2–dimensional complexes: a closed surface (all points are smooth), a surface with a disc attached (all points are of type $(1,2)$ or $(2,2)$), and a polygon (all points are of type $(0,0)$, $(1,1)$ or $(2,2)$)

2.1 Some examples

We construct three families of examples of fibrations $M \to X$ that correspond to three simple types of complexes X shown in Figure 12: surfaces, surfaces with triple points, and polygons.

If X is a closed surface, the fibrations $M \to X$ are precisely the torus bundles over X.

If X contains points of type $(1,2)$ and $(2,2)$, we obtain more manifolds. Recall that a Waldhausen *graph manifold* [8] is any 3–manifold that decomposes along tori into pieces diffeomorphic to $P \times S^1$ and $D^2 \times S^1$, where P is the pair of pants. For example, all lens spaces and Seifert manifolds are graph manifolds.

Proposition 2.1 *Let $p\colon M \to N$ be a circle bundle over an orientable closed graph manifold N. The closed manifold M has a pants decomposition $M \to X$ for some X consisting of points of type $(1,2)$ and $(2,2)$ only.*

Proof It is proved in [1, Proposition 3.31] that every orientable graph manifold N admits a fibration π over some simple complex X, called a *shadow*, that consists of points of type $(1,2)$ and $(2,2)$ only, with fibers respectively diffeomorphic to a θ–shaped graph and a circle. The composition of the two projections $\pi \circ p\colon M \to X$ is a pair of pants decomposition. □

If X has only points of type $(0,0)$, $(1,1)$ and $(2,2)$, then it is a surface with polygonal boundary consisting of vertices and edges. We also get interesting manifolds in this case.

Proposition 2.2 *A closed 4–dimensional toric manifold M has a pants decomposition $M \to X$ for some polygonal disc X. In particular, $\mathbb{CP}^2$ fibers over the triangle.*

Proof The moment map $M \to X$ is a fibration onto a polygon X locally modeled on $\mathbb{CP}^2 \to \Pi_2$ near a vertex of type $(0,0)$. □

The 4–dimensional closed toric varieties are $S^2 \times S^2$ and $\mathbb{CP}^2 \# h\overline{\mathbb{CP}}^2$; see [2]. In all the previous examples, the base complex X has no vertex of type $(0,2)$.

Problem 2.3 Classify all the pair of pants decompositions $M \to X$ onto simple complexes X without vertices of type $(0,2)$.

This is a quite interesting set of not-too-complicated 4–manifolds, which contains torus bundles over surfaces, circle bundles over graph manifolds, and toric manifolds.

2.2 Smooth hypersurfaces

We now turn to more complicated examples where X contains vertices of type $(0,2)$. Mikhalkin's theorem [4, Theorem 1] produces the following manifolds.

Theorem 2.4 *The smooth hypersurface M of degree d in $\mathbb{CP}^3$ has a pants decomposition $M \to X$ on a simple complex X with d^3 vertices of type $(0,2)$.*

Recall that the diffeomorphism type of M depends only on the degree d. When $d = 1,2,3,4$, the manifold M is $\mathbb{CP}^2$, $S^2 \times S^2$, $\mathbb{CP}^2 \# 6\overline{\mathbb{CP}}^2$, and the $K3$ surface, respectively.

2.3 Euler characteristic

The Euler characteristic of a pants decomposition can be easily calculated, and it depends only on the base X.

Proposition 2.5 *Let $M \to X$ be a pants decomposition. We have*

$$\chi(M) = n_0 - n_1 + n_2,$$

where n_i is the number of points of type $(0,i)$ in X.

Proof All fibers have zero Euler characteristic, except the fibers F_i above vertices of type $(0,i)$, that have $\chi(F_i) = (-1)^i$. □

2.4 The nodal surface

Proposition 2.6 *Let $p\colon M \to X$ be a pants decomposition. The preimage $S = p^{-1}(\partial X)$ is an immersed smooth compact surface in M.*

Proof The fibration p is locally modeled on the tropical fibration $\mathbb{CP}^2 \to \Pi_2$, and the preimage of $\partial\Pi_2$ in $\mathbb{CP}^2$ is an immersed surface consisting of four lines intersecting transversely in six points lying above the vertices of type $(0,0)$. $\square$

We call S the *nodal surface* of the fibration p. It is an immersed surface in M with one transverse self-intersection above each point of type $(0,0)$ of X. Every such self-intersection is called a *node*.

Remark 2.7 We note that a vertex of type $(0,1)$ connected to three vertices of type $(0,0)$ determines an embedded sphere in S. Two vertices of type $(0,0)$ connected by an edge also determine an embedded sphere.

3 Polygons

Let X be a 2–dimensional simple complex. Is there a combinatorial way to encode all the pants decompositions $M \to X$ fibering over X? Yes, there is one, at least in the more restrictive case where every connected stratum in X is a cell: every fibration is determined by some *labeling* on X, which is roughly the assignment of some 2×2 matrices to the connected 1–strata of X satisfying some simple requirements. We describe this method here in the simple case when X is a polygon. We will treat the general case in the next section.

3.1 Fibrations over polygons

Let X be a n–gon as in Figure 13 (left), that is, a simple 2–dimensional complex homeomorphic to a disc with $n \geqslant 1$ points of type $(0,0)$ called *vertices*. The strata of type $(0,1)$ form n *edges* (or *sides*).

Let $\pi\colon M \to X$ be a pair of pants decomposition. We first make some topological considerations.

Proposition 3.1 *The manifold M is simply connected, and $\chi(M) = n$. The nodal surface consists of n spheres.*

Proof We have $\chi(M) = n$ by Proposition 2.5. The manifold M is simply connected because X is, and every loop contained in some fiber $\pi^{-1}(x)$ is homotopically trivial: it suffices to push x to a vertex v of X and the loop contracts to the point $\pi^{-1}(v)$.

Thanks to Remark 2.7, the nodal surface consists of n spheres, one above each edge of X. $\square$

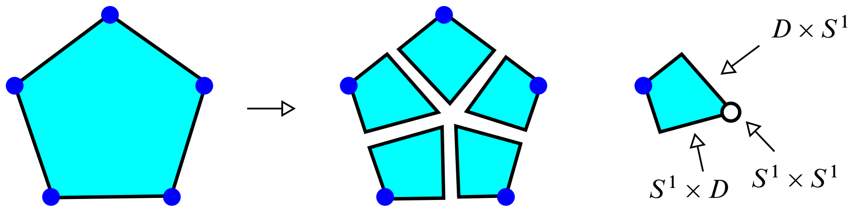

Figure 13: A fibering $M \to X$ over a pentagon (left) can be broken into n basic pieces (center). The fibering over each basic piece (right).

3.2 Orientations

In this paper, we will be often concerned with orientations on manifolds, their products and their boundaries. This can be an annoying source of mistakes, so we need to be careful. We will make use of the following formula on oriented manifolds M and N:

$$\partial(M \times N) = (\partial M \times N) \cup (-1)^{\dim M}(M \times \partial N). \tag{1}$$

Moreover, recall that the map

$$M \times N \to N \times M \tag{2}$$

that interchanges the two factors is orientation-preserving if and only if $\dim M \cdot \dim N$ is even.

3.3 The basic fibration

Again, let $M \to X$ be a fibration over a polygon. We now break the given fibration $M \to X$ into some basic simple pieces and show that $M \to X$ can be described by some simple combinatorial data.

We break the n–gon into n star neighborhoods of the vertices as in Figure 13 (center). Above each star neighborhood, the fibration is diffeomorphic to the *basic fibration*

$$D^2 \times D^2 \to [0,1] \times [0,1]$$

that sends (w, z) to $(|w|, |z|)$, which we encountered in Section 1.5 and sketched in Figure 13 (right). The whole fibration $M \to X$ is constructed by gluing n such basic fibrations as suggested in Figure 14 (left). We only need to find a combinatorial encoding of these gluings to determine $M \to X$.

Consider a single basic fibration $D^2 \times D^2 \to [0,1] \times [0,1]$ as in Figure 13 (right). The point $(0,0)$ is the fiber of $(0,0)$, the blue vertex in the figure. The boundary of

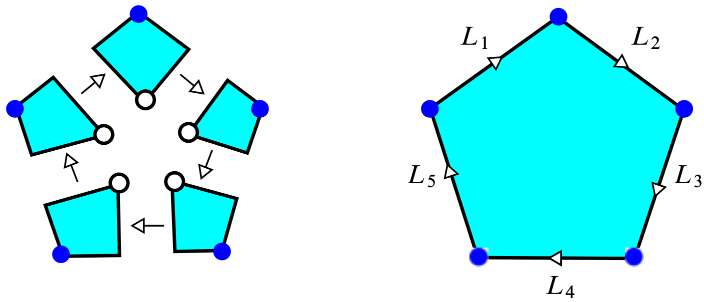

Figure 14: The fibering $M \to X$ may be reconstructed by gluing the basic fibrations (left). The gluing can be determined by some label matrices L_i (right).

$D^2 \times D^2$ is

$$\partial(D^2 \times D^2) = (D^2 \times S^1) \cup (S^1 \times D^2),$$

which is two solid tori (we call them *facets*) cornered along the torus $S^1 \times S^1$ (a *ridge*). The manifold $D^2 \times D^2$ is naturally oriented, and by (1) and (2), both solid tori inherit from $D^2 \times D^2$ their natural orientations, which are invariant if we swap the factors D^2 and S^1. The ridge torus $S^1 \times S^1$, however, inherits opposite orientations from the two facets.

The ridge torus $S^1 \times S^1$ is the fiber of $(1,1)$, the white dot in the figure, and the two facets, the solid tori, fiber over the two adjacent sides $\{1\} \times [0,1]$ and $[0,1] \times \{1\}$.

Every arrow in Figure 14 (left) indicates a diffeomorphism $\psi\colon D^2 \times S^1 \to S^1 \times D^2$ between two facets of two consecutive basic fibrations. It is convenient to write ψ as a composition

$$D^2 \times S^1 \xrightarrow{\psi'} D^2 \times S^1 \xrightarrow{j} S^1 \times D^2,$$

where j simply interchanges the two factors. By standard 3–manifold theory, the diffeomorphism ψ' is determined (up to isotopy) by its restriction to the boundary torus $S^1 \times S^1$, which is in turn determined (up to isotopy) by the integer invertible matrix $L \in \mathrm{GL}(2,\mathbb{Z})$ that encodes its action on $H_1(S^1 \times S^1) = \mathbb{Z} \times \mathbb{Z}$. The only requirement is that L must preserve the meridian; that is, it must send $(1,0)$ to $(\pm 1, 0)$. Summing up, we have the following.

Proposition 3.2 *The isotopy class of ψ' is determined by a matrix*

$$L = \begin{pmatrix} \varepsilon & k \\ 0 & \varepsilon' \end{pmatrix}$$

with $\varepsilon, \varepsilon' = \pm 1$ and $k \in \mathbb{Z}$.

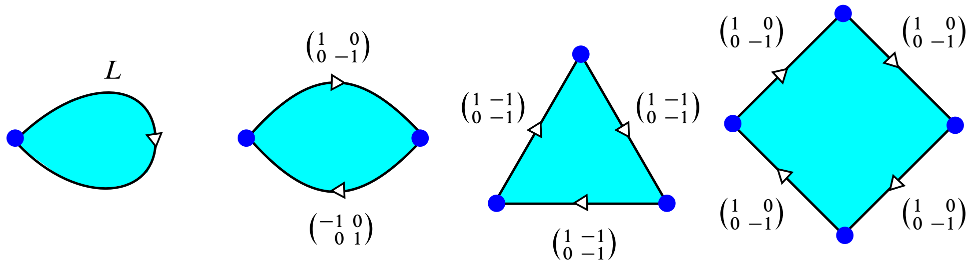

Figure 15: The monogon has no admissible labeling. The other admissible labelings shown here represent S^4, $\mathbb{CP}^2$ and $S^2 \times S^2$.

We can encode all the gluings by assigning labels $L_1, \ldots, L_n$ of this type to the n oriented edges of X as in Figure 14 (right). We call such an assignment a *labeling* of the polygon X. We define the matrices

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad J = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Not every labeling defines a fibration $M \to X$. A necessary and sufficient condition is that the global monodromy around the central torus must be trivial.

Proposition 3.3 *The labeling defines a fibration $M \to X$ if and only if*

$$JL_n JL_{n-1} \cdots JL_1 = I.$$

If $\det L_i = -1$ *for all i, the manifold M is oriented.*

Proof We only need to ensure that the monodromy around the central torus $S^1 \times S^1$ is isotopic to the identity, that is, $JL_n \cdots JL_1 = I$. (The composition $\psi = \psi' \circ j$ translates into JL.) If $\det L_i = -1$, the standard orientations of the pieces $D^2 \times D^2$ match to induce an orientation for M. □

We say that the labeling is *admissible* if $L_n J \cdots L_1 J = I$ and *oriented* if $\det L_i = -1$ for all i. Summing up, we have proved the following.

Proposition 3.4 *Every fibration $M \to X$ over an n–gon X is obtained by some admissible labeling on X.*

Some examples are shown in Figure 15. The monogon in Figure 15 (left) has no admissible labeling L because $LJ \neq I$ for every $L = \begin{pmatrix} \varepsilon & k \\ 0 & \varepsilon' \end{pmatrix}$. The figure shows some oriented admissible labelings on the bigon, the triangle and the square (admissibility is easily checked). Each determines a fibration $M \to X$.

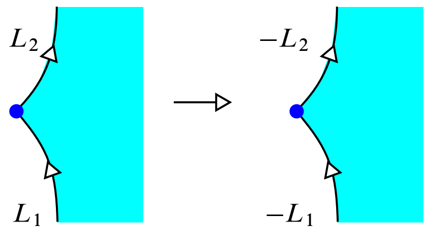

Figure 16: If we change the signs of the labels of two consecutive edges, the fibration $M \to X$ remains unaffected.

Proposition 3.5 *The bigon in Figure 15 represents S^4.*

Proof The bigon X decomposes into two pieces $[0,1]\times[0,1]$, and M decomposes correspondingly into two pieces $D^2\times D^2$. The manifold M decomposes smoothly into two 4–discs and is diffeomorphic to S^4. □

We have discovered that S^4 decomposes into pairs of pants. We will soon prove that the triangle and square in Figure 15 represent $\mathbb{CP}^2$ and $S^2\times S^2$, respectively.

Recall that we work entirely in the smooth (or equivalently, piecewise-linear) category.

3.4 Moves

We now introduce some moves on admissible labelings.

Let $L_1,\dots,L_n$ be a fixed oriented admissible labeling on the n–gon X with edges $e_1,\dots,e_n$. We know that it determines an oriented fibration $\pi\colon M\to X$. We start by noting that different labelings may yield the same fibration.

Proposition 3.6 *The move in Figure 16 produces a new oriented admissible labeling that encodes the same fibration $M\to X$.*

Proof The fibration $D^2\times D^2\to[0,1]\times[0,1]$ has the orientation-preserving automorphism $(z,w)\mapsto(\bar z,\bar w)$, which acts on $S^1\times S^1$ like $-I$. By employing it, we see that the move produces isomorphic fibrations $M\to X$. □

Since $\det L_i=-1$ by hypothesis, every label L_i is either $\left(\begin{smallmatrix}1 & k_i\\ 0 & -1\end{smallmatrix}\right)$ or $\left(\begin{smallmatrix}-1 & k_i\\ 0 & 1\end{smallmatrix}\right)$, and we correspondingly say that L_i is *positive* or *negative*. By applying the move of Figure 16 iteratively on the vertices of X, we may require that all labels L_i are positive except at most one. Positive labels are preferable because of the following.

Let S_i be the sphere in the nodal surface lying above the edge e_i of X. Note that two spheres S_i and S_j with $i \neq j$ intersect if and only if e_i and e_j are consecutive edges, and in that case, they intersect transversely at the point (a node) projecting to the common vertex.

Proposition 3.7 *If L_i is positive, the sphere S_i has a natural orientation. If L_i and L_{i+1} are positive, then $S_i \cdot S_{i+1} = +1$.*

Proof The label L_i represents the gluing of two pieces $D^2 \times D^2$ and $D^2 \times D^2$ along a map $\psi\colon D^2 \times S^1 \to S^1 \times D^2$ that sends the core $\{0\} \times S^1$ to $S^1 \times \{0\}$. The sphere S_i decomposes into two discs as $(\{0\} \times D^2) \cup_\psi (D^2 \times \{0\})$.

If L_i is positive, then ψ identifies $\{0\} \times S^1$ to $S^1 \times \{0\}$ orientation-reversingly, and hence the natural orientations of the two discs match to give an orientation for S_i.

The intersection of two consecutive S_i and S_j is transverse and positive (when they are both naturally oriented) because they intersect like $\{0\} \times D^2$ and $D^2 \times \{0\}$ inside $D^2 \times D^2$. □

Recall that the self-intersection $S_i \cdot S_i$ is independent of the chosen orientation for S_i and is hence defined for all i, no matter whether L_i is positive or not. The self-intersection of S_i is easily detected by the labeling as follows.

Proposition 3.8 *For each i, we have*

$$L_i = \begin{pmatrix} \pm 1 & \mp(S_i \cdot S_i) \\ 0 & \mp 1 \end{pmatrix}.$$

Proof Up to using the move in Figure 16, we may restrict to the positive case $L_i = \left(\begin{smallmatrix} 1 & k \\ 0 & -1 \end{smallmatrix}\right)$, and we need to prove that $S_i \cdot S_i = -k$. We calculate $S_i \cdot S_i$ by counting (with signs) the point in $S_i \cap S_i'$ where S_i' is isotopic and transverse to S_i.

Recall that $S_i = (\{0\} \times D^2) \cup_\psi (D^2 \times \{0\})$. We construct S_i' by taking the discs $\{1\} \times D^2$ and $D^2 \times \{1\}$: their boundaries do not match in $S^1 \times S^1$ because they form two distinct longitudes in the boundary of the solid torus $S^1 \times D^2$, of type $(1,0)$ and $(1,k)$. We can isotope the former longitude to the latter inside the solid torus, at the price of intersecting the core $S^1 \times 0$ some $|k|$ times. In this way, we get an S_i' that intersects S_i transversely into these $|k|$ points, always with the same sign.

We have proved that $S_i \cdot S_i = \pm k$. To determine the sign, it suffices to consider one specific case. We pick the triangle X in Figure 15, where all labels are $\left(\begin{smallmatrix} 1 & -1 \\ 0 & -1 \end{smallmatrix}\right)$. Here $\chi(M) = 3$, and M is simply connected; therefore, $H_2(M) = \mathbb{Z}$. The nodal surface contains three spheres S_1, S_2 and S_3 that represent elements in $H_2(M)$ with

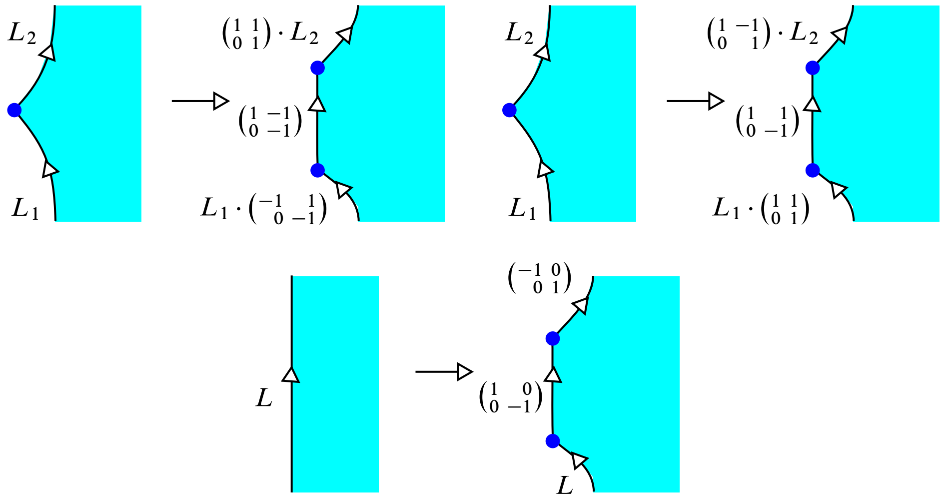

Figure 17: Three moves that transform M by connected sum with $\mathbb{CP}^2$ (top left), $\overline{\mathbb{CP}}^2$ (top right), and $S^2\times S^2$ (bottom center).

$S_i\cdot S_i=\varepsilon=\pm1$ for each i, and $S_i\cdot S_j=1$ when $i\neq j$. In particular, S_i is a generator of $H_2(M)$ for each i. Since $S_1\cdot S_2=S_1\cdot S_3=1$, then $S_2=S_3=\varepsilon S_1$, and hence $1=S_2\cdot S_3=\varepsilon^2\cdot S_1\cdot S_1=\varepsilon$; thus $S_i\cdot S_i=+1$. □

We now consider two more moves on positive admissible labelings, shown in Figure 17. It is easily checked that they both transform $L_1,\dots,L_n$ into a new positive admissible labeling on a bigger polygon.

Proposition 3.9 *The three moves in Figure 17 respectively transform M into*

$$M\,\#\,\mathbb{CP}^2,\quad M\,\#\,\overline{\mathbb{CP}}^2\quad\text{and}\quad M\,\#\,(S^2\times S^2).$$

Proof Both moves transform a fibration $M\to X$ into a new fibration $M'\to X'$. The first two moves substitute a vertex v of X with a new edge e. The preimages of v and e in M and M' are a point $x\in M$ and a sphere $S\subset M'$ with $S\cdot S=+1$ or $S\cdot S=-1$ depending on the move. Substituting x with S amounts to making a topological blowup, that is, a connected sum with $\mathbb{CP}^2$ and $\overline{\mathbb{CP}}^2$, respectively.

The third move substitutes a point x contained in some edge of X with a new edge e. The preimages of x and e in M and M' are a circle γ and a sphere $S\subset M'$ with $S\cdot S=0$. The substitution of γ with S is called a *surgery*, and since M is simply connected, the effect is a connected sum with $S^2\times S^2$. □

In particular, the triangle and square from Figure 15 represent the oriented smooth 4–manifolds $\mathbb{CP}^2$ and $S^2\times S^2$.

Corollary 3.10 *If $M = \#_h\mathbb{CP}^2 \#_k \overline{\mathbb{CP}}^2$ or $M = \#_h(S^2 \times S^2)$, then M decomposes into pairs of pants; more precisely, M fibers over the n–gon, with $n = \chi(M)$.*

These oriented manifolds are, in fact, all we can get from a polygon X.

Proposition 3.11 *Every oriented labeling on a polygon X represents one of the manifolds of Corollary 3.10.*

Proof Every label is of type $L_i = \left(\begin{smallmatrix} \pm1 & h_i \\ 0 & \mp1 \end{smallmatrix}\right)$. If $|h_i| \leqslant 1$, we can simplify X via one of the moves from Figure 17 and proceed by induction. If $|h_i| \geqslant 2$ for all i, it is easy to show that the coloring cannot be admissible, because the product $L_n J \cdots L_1 J$ cannot be equal to I.

Indeed, we have $M_i = L_i J = \left(\begin{smallmatrix} h_i & \pm1 \\ \mp1 & 0 \end{smallmatrix}\right)$. The matrix M_1 sends the vector $\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ to some $\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$ with $|a| > |b| > 0$, and any such vector is sent by any M_i to a vector $\left(\begin{smallmatrix} a' \\ b' \end{smallmatrix}\right)$ with $|a'| > |b'| > 0$ again, so $M_n \cdots M_1 \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) \neq \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$. □

4 The general case

We now extend the discussion of the previous section from polygons to more general simple complexes X. For the sake of simplicity, we restrict our investigation to a class of complexes called *special*, whose strata are all discs.

Definition 4.1 A simple complex X is *special* if the connected components of all the (k,l)–strata are open k–cells.

For instance, the polygons and the model complex Π_n are special. Every connected component of each stratum in a special 2–dimensional complex X is a cell, called *vertex*, *edge* or *face* according to its dimension. Vertices are of type $(0,0)$, $(0,1)$ or $(0,2)$, and edges are of type $(1,1)$ or $(1,2)$. Each face is a polygon with m edges and m vertices for some m, and the vertices may be of different types.

4.1 The basic fibrations

Let $M \to X$ be a fibration over some special complex X. We now extend the discussion of the previous section to this more general setting: we break $M \to X$ into basic fibrations of three types, and we show that $M \to X$ may be encoded by some combinatorial labeling on X that indicates the way these basic fibrations match along their (cornered) boundaries.

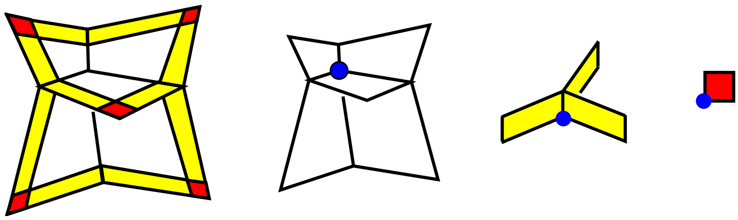

Figure 18: The complex Π_1 (left) decomposes into the star neighborhoods of its vertices (right).

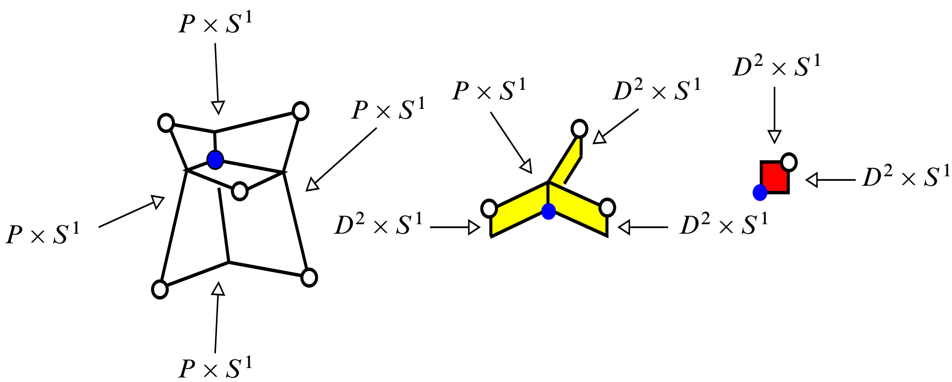

Figure 19: Every block M_v is a compact 4–manifold with corners; its boundary is a closed connected 3–manifold cornered along tori. There is one corner torus above each white dot, and the tori decompose the 3–manifold ∂M_v into pieces diffeomorphic to $S^1 \times P$ or $S^1 \times D^2$. Here P indicates the pair of pants.

A n–gon breaks into n star neighborhoods of its vertices as in Figure 13; analogously, every special complex X decomposes into star neighborhoods S_v of its vertices v, which are now of three different types $(0,0)$, $(0,1)$ and $(0,2)$. For instance, the model complex Π_2 decomposes into 11 pieces as shown in Figure 18: these are 6, 4 and 1 stars of vertices of type $(0,0)$, $(0,1)$ and $(0,2)$, respectively.

The fibration $\pi\colon \mathbb{CP}^2 \to \Pi_2$ decomposes correspondingly into $6+4+1=11$ *basic fibrations* $M_v \to S_v$ above the star neighborhood S_v of each vertex v. Every manifold M_v is a regular neighborhood in $\mathbb{CP}^2$ of the fiber $\pi^{-1}(v)$ of v, and its topology is deduced from Figures 9 and 10.

There are three basic fibrations $M_v \to S_v$ to analyze, depending on the type of the vertex v. If v is of type $(0,0)$ or $(0,1)$, the fibration $M_v \to S_v$ is respectively diffeomorphic to

$$D^2 \times D^2 \to [0,1] \times [0,1], \quad (z,w) \mapsto (|z|,|w|)$$

or

$$D^2 \times P \to [0,1] \times Y, \quad (z,x) \mapsto (|z|, \pi(x)),$$

where Y is a Y–shaped graph and $\pi\colon P \to Y$ is the tropical fibration; see Figure 19. Both D^2 and P are naturally oriented as subsets of some complex line in $\mathbb{CP}^2$. In both cases, M_v is a product, and its boundary is

$$\partial(D^2 \times D^2) = (D^2 \times S^1) \cup (S^1 \times D^2)$$

or

$$\partial(D^2 \times P) = (D^2 \times \partial P) \cup (S^1 \times P),$$

respectively. Recall the orientation formulas (1) and (2). The boundary consists of some *facets* cornered along tori (the *ridges*). The facets are either solid tori or $S^1 \times P$. Once and for all, we orientation-preservingly identify every boundary component of P with S^1, so that $D^2 \times \partial P$ is identified to three copies of $D^2 \times S^1$. There are three corner tori in $S^1 \times \partial P$.

4.2 The pair of pants

If v is of type $(0,2)$, the block M_v is not a product: it is the compact *pair of pants*, as named by Mikhalkin [4], diffeomorphic to the complement of an open regular neighborhood of four generic lines $l_1, \dots, l_4$ in $\mathbb{CP}^2$. Its boundary has four facets $f_1, \dots, f_4$, each diffeomorphic to $S^1 \times P$, cornered along six tori, one for each pair l_i, l_j of distinct lines.

The facet f_i is an S^1–bundle over some pair of pants $P_i \subset l_i$ obtained from l_i by removing open discs containing the intersection points with the other lines. The bundle is necessarily trivial since it is a circle bundle over a compact orientable surface with nonempty boundary; hence f_i is diffeomorphic to $S^1 \times P$, but unfortunately *not* in a canonical way (not even up to isotopy): the diffeomorphism depends (up to isotopy) on the choice of a *section* of the bundle, and on an orientation of the fibers (this is a standard fact on circle bundles over surfaces with boundary).

A natural way to construct a section goes as follows. Pick a line $r \in \mathbb{CP}^2$ that intersects l_i in one of the three points $l_i \cap l_j$, for some $j \neq i$. The line r provides a section of the normal bundle of l_i that vanishes only at $l_i \cap l_j$, and hence a section of the circle bundle over P_i. In fact, the isotopy class of the section does not depend on the chosen line r, but only on the point $l_i \cap l_j$, so there are three possible choices.

We now fix an arbitrary partition $\{l_1, l_2\}, \{l_3, l_4\}$ of the four lines into two pairs, and define r to be the line passing through the points $l_1 \cap l_2$ and $l_3 \cap l_4$. We use the line r to define sections on all the four facets f_i simultaneously as just explained.

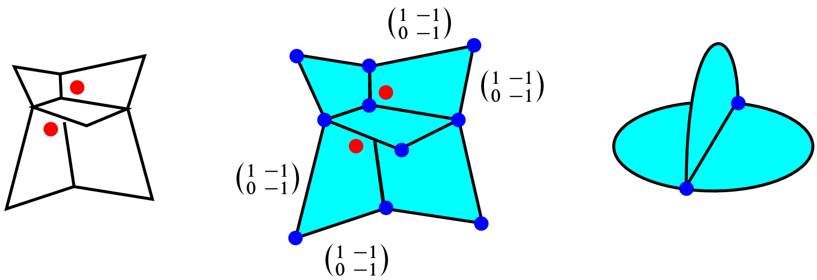

Figure 20: At every vertex of type $(0,2)$, we fix two (of the six) opposite faces and we mark them with red dots (left). An admissible oriented labeling on X that represents the tropical fibration $\mathbb{CP}^2 \to \Pi_2$ (center) and one that represents $(S^1 \times S^3) \# (S^1 \times S^3)$ (right).

Each section is oriented as a subset of r and identified with P. To complete the identification of f_i with $S^1 \times P$ we need to orient the fibers: we orient them so that $S^1 \times P$ gets the correct orientation as a boundary portion of the block M_v (which is in turn oriented as a domain in $\mathbb{CP}^2$).

Remark 4.2 By taking an affine chart that sends r to infinity, we see that

$$\mathbb{CP}^2 \setminus (l_1 \cup l_2 \cup l_3 \cup l_4 \cup r) \cong (\mathbb{C}\setminus\{0,1\}) \times (\mathbb{C}\setminus\{0,1\}),$$

so M_v minus an open neighborhood of r is naturally diffeomorphic to a product $P \times P$. This diffeomorphism furnishes the identifications of each f_i with $S^1 \times P$ just described.

There are, of course, three possible partitions of $\{l_1, l_2, l_3, l_4\}$ to choose from. To indicate on X which partition we use, we mark with a dot the two opposite faces near v that correspond to the pairs l_1, l_2 and l_3, l_4, as in Figure 20 (left). This mark fixes an identification of every facet f_i with the product $S^1 \times P$.

4.3 Labeling

Every fibration $M \to X$ decomposes into basic fibrations, glued along facets that are either $D \times S^1$ or $S^1 \times P$. We now encode every such gluing with an appropriate labeling on X that extends the one introduced in Section 3 for polygons.

A typical face f of X is shown in Figure 21: it may have vertices and edges of various kinds, and its closure need not be embedded (it may also be adjacent multiple times to the same edge or vertex). We want to assign labels L_i to the oriented edges (that is, sides) of f as shown in the figure.

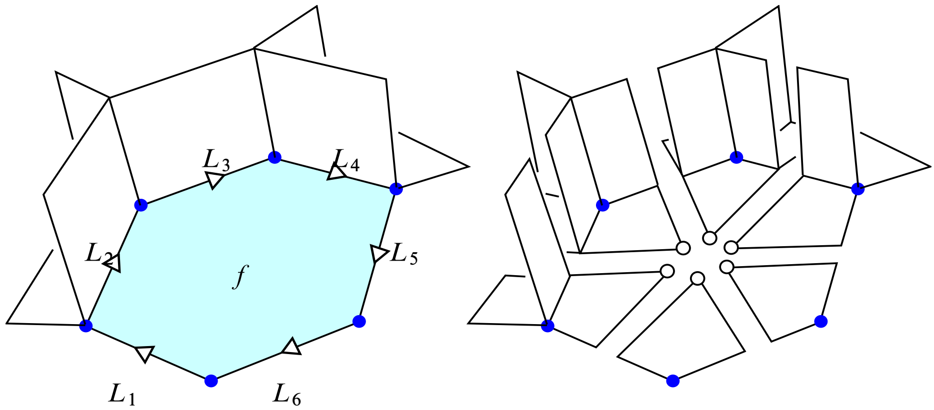

Figure 21: A face f of a special complex X, with vertices and edges of various types: here f has two vertices of each type $(0,0)$, $(0,1)$ and $(0,2)$, and three edges of each type $(1,1)$ and $(1,2)$. We label the oriented edges with some matrices L_i (left), and we break f into star neighborhoods of its vertices (right).

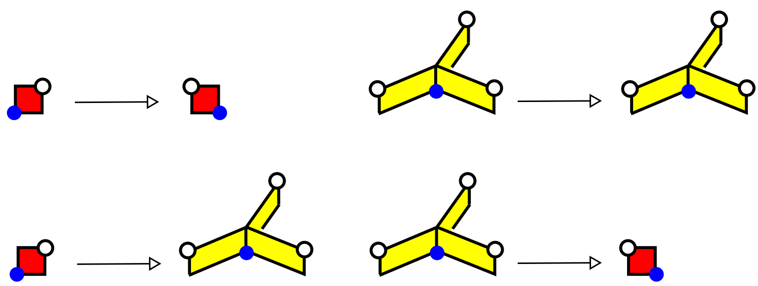

Figure 22: Four possible gluings along an oriented edge of type $(1,1)$

An edge e_i of f can be of either type $(1,1)$ or type $(1,2)$, and we respectively call it an *interior edge* or a *boundary edge*. A boundary edge is contained in ∂X and connects two vertices v and v' that may be of type $(0,0)$ or $(0,1)$. There are four possible cases, shown in Figure 22. In any case, the fibrations $M_v \to S_v$ and $M_{v'} \to S_{v'}$ get identified along some diffeomorphism $\psi\colon D^2 \times S^1 \to D^2 \times S^1$ identifying two solid torus facets. As in Section 3, we encode this diffeomorphism unambiguously (up to isotopy) via a label

$$L_i = \begin{pmatrix} \varepsilon & k \\ 0 & \varepsilon' \end{pmatrix}$$

with $\varepsilon, \varepsilon' = \pm 1$ and $k \in \mathbb{Z}$. This label is assigned to the side e_i of f.

If e_i is an interior edge, it connects two vertices v and v' that may be of type $(0,1)$ or $(0,2)$. The two fibrations $M_v \to S_v$ and $M_{v'} \to S_{v'}$ are now glued along a diffeomorphism $\psi\colon S^1 \times P \to S^1 \times P$ between two facets.

The restriction of ψ to the boundary torus lying above f is a diffeomorphism $S^1 \times S^1 \to S^1 \times S^1$ whose isotopy class is encoded by a matrix $L_i \in \mathrm{GL}(2,\mathbb{Z})$. This is the label that we assign to e_i.

Since the fiber generates the center of $\pi_1(S^1 \times P)$, the diffeomorphism $\psi\colon S^1 \times P \to S^1 \times P$ must preserve the fiber (up to reversing the orientation). Therefore, the label L_i has the same nice form as in the previous case:

$$L_i = \begin{pmatrix} \varepsilon & k \\ 0 & \varepsilon' \end{pmatrix}.$$

Summing up, a *labeling* of X is simply the assignment of a matrix $\left(\begin{smallmatrix} \varepsilon & k \\ 0 & \varepsilon' \end{smallmatrix}\right)$ to every oriented side e of every face f in X.

We implicitly agree that the orientation reversal of the side e changes the label from L to L^{-1}. Note that an interior edge e inherits three labels, one for each incident face, while a boundary edge has only one label.

4.4 The fibration of $\mathbb{CP}^2$ over Π_2

As an example, we now analyze in detail the labeling on Π_2 induced by the tropical fibration $\pi\colon \mathbb{CP}^2 \to \Pi_2$; the answer is depicted in Figure 20 (center), where every unlabeled edge is tacitly assumed to have label $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. This analysis is not necessary for the rest of the paper, so the reader may skip it and jump to Section 4.5.

Recall that the preimage of all points of type (k,l) with $l \leqslant 1$ is the union of four lines in $\mathbb{CP}^2$, intersecting in the six points of type $(0,0)$. Call these lines $l_1,\dots,l_4$, and denote by $q_1,\dots,q_4$ the four points of type $(0,1)$ corresponding to $l_1,\dots,l_4$, respectively.

Fix an ordered pair (i,j). At the intersection point $p_{ij} = l_i \cap l_j$, we have an identification of a neighborhood N_{ij} of p_{ij} with $D^2 \times D^2$ such that $D^2 \times \{0\}$ is the intersection of l_i with N_{ij}, and $\{0\} \times D^2$ is the intersection of l_j with N_{ij}. (The identification is sensitive to swapping i and j.) Set $q_{ij} = \pi(p_{ij})$.

We fix as above an auxiliary line r going through the points $l_1 \cap l_2$ and $l_3 \cap l_4$. The line r induces a section of the normal bundles of the four lines, and we use it to fix an identification of all the other facets involved with $S^1 \times P$. With this identification, every internal edge gets a label $\left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$; one only needs to check signs by looking at orientations. Using a move that will be described in Proposition 4.7, we can change all these labels with $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$.

We need to determine the labels on the external edges. Consider the point q_{13}. We are interested in the isotopy class of the section above l_1 in the boundary of N_{13}. Since all lines going through p_{12} are isotopic, the section induced by r on N_{13} is parallel to the curve $S^1 \times \{1\}$ in ∂N_{13}. Therefore, the label on the edge connecting q_1 to q_{13} is diagonal, and by looking at the orientations, we get $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. Likewise, all edges incident to q_{14}, q_{23} and q_{24} are labeled with $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$.

At the point p_{12}, the section determined by r on l_1 is no longer parallel to $S^1 \times \{1\}$ in ∂N_{12}. However, one checks that the section is parallel to the diagonal curve S^1 in the corner torus $S^1 \times S^1$ in N_{12}, and we get $\left(\begin{smallmatrix} 1 & -1 \\ 0 & -1 \end{smallmatrix}\right)$.

Notice that in no case do we need to specify an orientation of the edges, since $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)^2 = \left(\begin{smallmatrix} 1 & -1 \\ 0 & -1 \end{smallmatrix}\right)^2 = I$.

4.5 Admissibility

As in the polygonal case, every fibration is encoded by some (nonunique) labeling of X, but not every labeling defines a fibration: some simple conditions need to be verified.

Let f be a face of X, with oriented sides $e_1, \dots, e_n$. Let v_i be the vertex of f adjacent to e_i and e_{i+1}. We assign a matrix J_i to v_i as follows:

- if v_i is of type $(0,0)$, then $J_i = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$;
- if v_i is of type $(0,1)$, then $J_i = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$;
- if v_i is of type $(0,2)$ and is not dotted, then $J_i = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$;
- if v_i is of type $(0,2)$ and is dotted, then $J_i = \left(\begin{smallmatrix} -1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$.

Recall that we have fixed two dots at every vertex of type $(0,2)$ as in Figure 20. Note that in all cases, we get $J_i^2 = I$.

Proposition 4.3 *A labeling defines a fibration $M \to X$ if and only if the following hold:*

(1) *at every oriented interior edge, the three labels of the incident faces are*

$$\begin{pmatrix} \varepsilon & k_1 \\ 0 & \varepsilon' \end{pmatrix}, \quad \begin{pmatrix} \varepsilon & k_2 \\ 0 & \varepsilon' \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \varepsilon & k_3 \\ 0 & \varepsilon' \end{pmatrix}$$

for some constants $\varepsilon, \varepsilon' = \pm 1$, with the condition $k_1 + k_2 + k_3 = 0$;

(2) *at every face f, we have*

$$J_n L_n \cdots J_1 L_1 = I.$$

If $\det L_i = -1$ *for all i, the manifold M is oriented.*

Proof At every interior edge we need to build a diffeomorphism $\psi\colon S^1 \times P \to S^1 \times P$, and it is a standard fact in three-dimensional topology that such a diffeomorphism exists if and only if it acts on the boundary tori $S^1 \times S^1$ as specified by condition (1).

Condition (2) is that the monodromy around the central torus must be the identity. The role of J_i is to translate between the two bases of the same corner torus, used by the two adjacent facets. A careful case by case analysis is needed here:

- if v_i is of type $(0,0)$, the facets are $S^1 \times D^2$ and $D^2 \times S^1$, so $J_i = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$;
- if v_i is of type $(0,1)$, the facets are $D^2 \times S^1$ and $S^1 \times P$, so $J_i = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$;
- if v_i is of type $(0,2)$, both facets are $S^1 \times P$ and there are two cases:
  - if v_i is not dotted, the factors in $S^1 \times P$ are interchanged as in the case of $(0,0)$, so we get $J_i = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$;
  - if v_i is dotted, the boundaries $S^1 \times \partial P$ of the two sections coincide, and we get $J_i = \left(\begin{smallmatrix} -1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$.

In the latter case, we have three complex lines l_1, l_2 and r passing through a point p and determining three oriented curves γ_1, γ_2 and μ in the corner torus $S^1 \times S^1$. The bases to be compared are (γ_1, μ) and (γ_2, μ) and we have $\mu = \gamma_1 + \gamma_2$; hence $\gamma_2 = \mu - \gamma_1$, and we get $J_i = \left(\begin{smallmatrix} -1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$. □

A labeling on X satisfying the requirements of Proposition 4.3 is *admissible*. If $\det L_i = -1$, then it is *oriented*. An oriented label is either $L = \left(\begin{smallmatrix} 1 & k \\ 0 & -1 \end{smallmatrix}\right)$ or $\left(\begin{smallmatrix} -1 & k \\ 0 & 1 \end{smallmatrix}\right)$, and we have respectively called them positive and negative. Note that $L = L^{-1}$, and hence we do not need to orient the edge when assigning it an oriented label. Also in this setting, positive labels are preferable (at least on boundary edges).

Proposition 4.4 *If all labels are oriented and positive, the nodal surface S is naturally oriented. Every nodal point has positive intersection $+1$.*

Proof This is the same as for Proposition 3.7, with $P \times D^2$ replacing $D^2 \times D^2$. □

We now turn to self-intersection. The nodal surface S is the union of some closed surfaces $S_1 \cup \cdots \cup S_k$ intersecting transversely, such that the abstract resolution of each S_i is connected.

Proposition 4.5 *If the labels are oriented and positive, and S_i is embedded, then*

$$S_i \cdot S_i = -\sum_j k_j$$

as $L_j = \left(\begin{smallmatrix} 1 & k_j \\ 0 & -1 \end{smallmatrix}\right)$ varies among all labels on edges onto which S_i projects.

Proof This is the same as for Proposition 3.8. □

Example 4.6 Consider the two labelings in Figure 20 (center) and (right), where every unlabeled edge is tacitly assumed to have label $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. Both labelings are oriented and admissible: the three labels at the interior edges are equal to $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ and hence condition (1) is fulfilled. In the central figure, there are two kinds of faces; the nondotted ones give

$$\begin{aligned} &J_4 L_4 \cdots J_1 L_1 \\ &\quad = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = I, \end{aligned}$$

and on the dotted ones we get

$$\begin{aligned} &J_4 L_4 \cdots J_1 L_1 \\ &\quad = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix} = I. \end{aligned}$$

As seen above, this labeling represents the tropical fibration $\mathbb{CP}^2 \to \Pi_2$.

On the right figure, we note that there are only two vertices v, both of type $(0,1)$, and at every face we have

$$J_2 L_2 J_1 L_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = I.$$

The manifold M here is the double of the basic piece M_v along its boundary. The fiber above v is a θ–shaped graph θ and M_v is a regular neighborhood of θ, that is, a handlebody with one 0–handle and two 1–handles. The double of such a manifold is $M = (S^3 \times S^1) \# (S^3 \times S^1)$.

4.6 Moves

Let X be a special complex equipped with an admissible labeling, defining a fibration $M \to X$. The moves described in Section 3.4 apply also here, and there are more moves that involve vertices of type $(0,1)$ and $(0,2)$ that modify a labeling without affecting the fibration $M \to X$.

Proposition 4.7 *Let v be a vertex of X, of any type $(0,0)$, $(0,1)$ or $(0,2)$. If we change the signs simultaneously of the labels on all the (two, three or four) edges incident to v, we get a new admissible labeling that encodes the same fibration $M \to X$.*

Proof The manifolds $D^2 \times D^2$, $D^2 \times P$ and the four-dimensional pair-of-pants B have orientation-preserving self-diffeomorphisms that act like $-I$ on the homologies of all the corner tori in the boundary.

To see this for B, consider B as the complement of some lines in $\mathbb{CP}^2$ defined by equations with real coefficients. The map $[z_0, z_1, z_2] \mapsto [\overline{z}_0, \overline{z}_1, \overline{z}_2]$ preserves B and acts as required. □

We note in particular that Proposition 3.9 is still valid in this context.

Proposition 4.8 *The three moves in Figure 17 respectively transform M into*

$$M \# \mathbb{CP}^2, \quad M \# \overline{\mathbb{CP}}^2 \quad \text{and} \quad M \# (S^2 \times S^2).$$

Remark 4.9 In this section, we have dealt only with special complexes, as this simplifies the labelings, but an extension of Propositions 4.3 and 4.8 to all simple complexes can be done quite easily. In the first proposition, condition (1) is local, and is required also when dealing with nonspecial complexes. Condition (2), on the other hand, is only needed to ensure that the torus fibration on the boundary extends to the interior of the cell; if a connected component of the $(2,2)$–stratum is not a disc, we need to require that the fibration on its boundary extends to the interior. Notice that this extension is not unique in general; hence a labeling in the above sense does not determine a fibration $M \to X$. In order to get uniqueness, we need to specify the monodromy on the boundary as well as its extension. We do not explore this further here.

Remark 4.10 A 3–manifold decomposing into pieces diffeomorphic to $D^2 \times S^1$ and $P \times S^1$ was called a *graph manifold* by Waldhausen [8]: such 3–manifolds are classified and well understood.

5 Fundamental groups

In the previous section, we have made some effort in defining some labelings that encode all pants decompositions $M \to X$ over a given special complex X. We now use them to prove the following, which is the main result of this paper.

Theorem 5.1 *For every finitely presented group G, there is a pants decomposition $M \to X$ with $\pi_1(M) = G$.*

5.1 Even complexes

We say that a special complex is *even* if every face is incident to an even number of vertices (counted with multiplicity). Recall that there are three types of vertices, $(0,0)$, $(0,1)$ and $(0,2)$, and each of these must be counted. For instance, the complex Π_2 is even: every 2–cell is incident to four vertices.

Even complexes are particularly useful here because of the following.

Proposition 5.2 *If X is even, there is a pants decomposition $M \to X$.*

Proof We note that every face f in any simple complex contains an even number of vertices of type $(0,1)$. So the evenness hypothesis on X says that the number of vertices of type $(0,0)$ or $(0,2)$ is even for every f.

Every vertex v of type $(0,2)$ in X is adjacent to six faces, and we assign dots to two opposite ones arbitrarily.

We will first try to assign trivial labels $L = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ everywhere. Condition (1) of Proposition 4.3 is trivially satisfied, and at every face f we get a product monodromy $J_{2n}L_{2n} \cdots J_1L_1 = J_{2n} \cdots J_1$ that we now compute.

If there were no dots in f, we would get $J_{2n} \cdots J_1 = J^{2k} = I$ with $J = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ and $2k \leqslant 2n$ is the number of vertices of type $(0,0)$ or $(0,2)$. In that case, condition (2) would also be satisfied.

If there are some dots, we adjust the labeling so that the above construction still works. For every maximal string of dotted corners of odd length in a polygonal face, we put a label $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ at the two oriented edges incoming and outcoming the string, both oriented towards the string; eg if there is an isolated dotted corner v, the two labels on the edges incoming into v will have label $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, while if there are two connected dotted corners w and w' isolated from all other dotted corners, the label on all the edges incident to w or w' will simply be $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

This works for the compatibility condition (2) since two consecutive dotted corners contribute with $\left(\begin{smallmatrix} -1 & 0 \\ 1 & 1 \end{smallmatrix}\right)^2 = I$; in an even chain, the product is trivial, while in an odd chain of $2k+1$ dotted vertices, we obtain

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 1 & 1 \end{pmatrix}^{2k+1}\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

In either case, after the simplification, we are left with a power of $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ for each chain of odd length, and the global monodromy will be trivial for parity reasons (because X is even).

The addition of these labels, however, may have destroyed condition (1). Consider an oriented interior edge e that connects either two vertices of type $(0,2)$ or one of type $(0,1)$ and one of type $(0,2)$.

If e is incident to two vertices of type $(0,2)$, there are two possibilities: either the dots are on the same face of X incident to e, or they are on different faces.

In the former case, the three labels of e are left unchanged, namely $L = \left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$, and condition (1) is trivially satisfied. In the latter, two of its three labels have been modified to $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}1&-1\\0&1\end{smallmatrix}\right)$ and then condition (1) still holds (note the role of the edge directions).

If e is incident to a vertex v_2 of type $(0,2)$ and one v_1 of type $(0,1)$, it is incident to three faces, exactly one of which has a dotted corner at v_2; denote this face with f. If the label of e as part of ∂f is the trivial label $\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$, condition (1) is again automatically satisfied.

Suppose now the label of e has been changed. Then condition (1) is violated along e since exactly one label has been modified to $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$. We need to modify the labeling further, and we do so by modifying both the labels at such edges e and on some boundary edges that share a vertex with them.

Consider the set E_1 of all external edges e_1 with the following property: e_1 shares exactly one vertex with an interior edge e such that the label on e on the face f_1 that they span is nontrivial; ie it is $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$. Let E_2 be the set of all external edges e_2 with the following property: e_2 shares both endpoints with two interior edges e' and e'', and the labels on e' and e'' on the face f_2 that they span are both nontrivial; ie they are $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$. By construction, E_1 and E_2 are disjoint, and so are the associated sets of interior edges. Also, notice that the faces and edges denoted by f_1 and e (respectively, f_2, e' and e'') are all determined by e_1 (resp. e_2).

For every edge e_1 in E_1, we orient it towards e, we replace the label of e_1 with $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ and let e, seen as part of the boundary of f, have the trivial label $\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$. For every edge e_2 in E_2, we replace the two labels on the two associated edges e' and e'' (as part of the boundary of f_2) by the trivial label $\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$, and leave the label of e_2 unchanged, ie trivial.

It is readily checked that now both conditions (1) and (2) are satisfied. $\square$

There are many even complexes:

Proposition 5.3 *Every finitely presented group is the fundamental group of an even complex without boundary.*

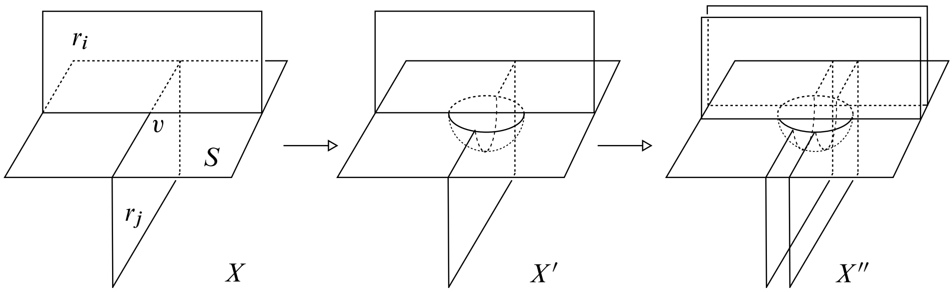

Figure 23: How to construct an even complex. The base surface S here is horizontal, and the relator faces r_i and r_j are attached vertically.

Proof Every finitely presented group $G = \langle g_1, \dots, g_k \mid r_1, \dots, r_s \rangle$ is the fundamental group of some special complex X without boundary, constructed by attaching discs to a genus-k surface S. To see this, first attach discs to S to transform the fundamental group of S into a free group F_k with k generators (for instance, you may take the meridians of a handlebody with boundary S). Then attach discs on S along s generic curves that represent the relators $r_1, \dots, r_s$ in a sufficiently complicated way, so that S is cut into polygons by them (add a trivial relator r_1 in case there are none).

We now modify X to an even complex X'' with the same fundamental group G. The modification is depicted in Figure 23 and consists of two steps: the first is a local modification at every vertex v of S, where two relator faces r_i and r_j intersect. Note that every face in X' is even, except the new small triangles created by the move. Then we double each relator r_i as shown in the figure (that is, for every $i = 1, \dots, s$ we attach two parallel discs), Now triangles are transformed into squares: the final polyhedron X'' is even and has the same fundamental group G of X and X'. □

5.2 Fundamental group

How can we calculate the fundamental group of M by looking at the fibration $M \to X$? We answer this question in some cases. We start by showing that in dimension 4, any facet of the compact pair of pants carries the fundamental group of the whole block (in contrast with dimension 2).

Lemma 5.4 *Let B be the compact 4–dimensional pair of pants and $F \cong P \times S^1$ any of its four facets. The map $\pi_1(F) \to \pi_1(B)$ induced by inclusion is surjective.*

Proof Recall that B is $\mathbb{CP}^2$ minus the open regular neighborhood of four lines l_1, l_2, l_3 and l_4. Let F correspond to l_4. Using the Salvetti complex [6], we see that $\pi_1(B) \cong \mathbb{Z}^3$ is generated by three loops turning around any three of these lines, say l_1, l_2 and l_3. These loops can be homotoped inside $F \cong P \times S^1$, where they correspond to three meridians on the boundary tori. □

Let X^1 denote the 1–stratum of X, that is, the set of all nonsmooth points of X.

Proposition 5.5 *Let $M \to X$ be a pants decomposition. Then the induced map $\pi_1(M) \to \pi_1(X)$ on fundamental groups is surjective. It is also injective, provided the following hold:*

- *X is not a surface,*
- *every connected component of $X^1 \setminus \partial X$ is incident to a vertex in ∂X whose fiber is contained in a (possibly immersed) spherical component of the nodal surface.*

Proof The map $\pi_1(M) \to \pi_1(X)$ is surjective because all fibers are connected and arcs lift from X to M.

Let $F_x = \pi^{-1}(x)$ be the fiber of x and let G_x be the image of the map $\pi_1(F_x) \to \pi_1(M)$ induced by inclusion (with some basepoint in F_x). It is easy to prove that if G_x is trivial for every $x \in X$, then $\pi_1(M) \to \pi_1(X)$ is an isomorphism. We now prove that the additional assumptions listed above force all groups G_x to be trivial.

We use the term *connected stratum* to denote a connected component of some (k,l)–stratum of X. If G_x is trivial for some x, then $G_{x'}$ is trivial for all points x' lying in the same connected stratum of x, and we say that the connected stratum is *trivial*. We now show that the triviality propagates along incident connected strata in most (but not all!) cases. Let s and t be two incident connected strata, that is, such that either $s \subset \overline{t}$ or $t \subset \overline{s}$. Suppose that s is trivial. We claim that, if any of the following conditions holds, then t is also trivial:

(1) $\dim t > \dim s$;

(2) $t \subset \partial X$, $s \not\subset \partial X$, and $\dim t = \dim s - 1$;

(3) t is a vertex of type $(0,2)$ and s is an edge of type $(1,2)$.

To prove the claim, pick $x \in s$ and $y \in t$; by assumption, G_x is trivial.

(1) We have $s \subset \overline{t}$, and the fiber F_y can be isotoped to $F_{y'}$ where y' is close to x, so $F_{y'}$ lies in a regular neighborhood of F_x; therefore, G_y is naturally a subgroup of G_x, hence trivial.

(2) In this particular case, $F_x \cong F_y \times S^1$, and F_y can be isotoped inside F_x.

(3) It follows from Lemma 5.4.

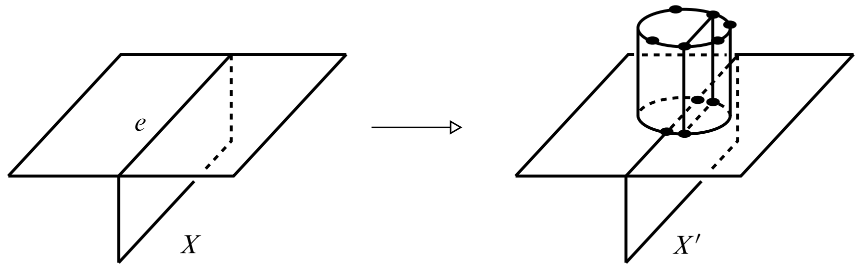

Figure 24: How to create some boundary on an even complex, preserving evenness and the fundamental group. The complex X' is constructed by attaching a product $\theta \times [0,1]$ to X along $\theta \times 0$ as shown, where θ is a θ–shaped graph. The boundary $\partial X' = \theta \times 1$ contains four vertices of type $(0,0)$ and two of type $(0,1)$; all dotted points are vertices of some type.

By assumption, every connected component C of $X^1 \setminus \partial X$ is incident to a vertex v of type $(0,1)$ in ∂X, whose fiber F_v is contained in a sphere; therefore, G_v is trivial. By property (2), the edge of type (1,2) adjacent to v is also trivial, and we can use (1) and (3) to propagate the triviality along all the connected strata of C.

Since X is not a surface, every 2–dimensional connected stratum of X is incident to $X^1 \setminus \partial X$, and is hence trivial by property (1). Finally, the triviality extends to the rest of ∂X by (2). □

Corollary 5.6 *Let $M \to X$ be a pants decomposition. If $X^1 \setminus \partial X$ is connected, $\partial X \neq \varnothing$, and the nodal surface consists of (possibly immersed) spheres, the map $\pi_1(M) \to \pi_1(X)$ is an isomorphism.*

The homomorphism $\pi_1(M) \to \pi_1(X)$ may not be injective in general: Figure 20 (right) shows a fibration $M \to X$ with $\pi_1(M) = \mathbb{Z} * \mathbb{Z}$ and $\pi_1(X) = \{e\}$.

5.3 Proof of the main theorem

We can finally prove the main result of this paper, Theorem 5.1.

Proof of Theorem 5.1 For every finitely presented group G there is an even special complex X without boundary and with $\pi_1(X) = G$ by Proposition 5.3. We slightly modify X to a complex X' with nonempty boundary by choosing an arbitrary edge e and modifying X near e as shown in Figure 24.

We have $\pi_1(X) = \pi_1(X')$, and X' is still even. Note that $\partial X'$ is a θ–shaped graph with two vertices of type $(0,1)$ and also four vertices of type $(0,0)$ as indicated in the picture. Note also that X^1 is connected because X is special without boundary, and hence $(X')^1 \setminus \partial X'$ is also connected.

By Proposition 5.2, there is a pants decomposition $M \to X'$. By looking at $\partial X'$, we see that the nodal curve consists of three spheres. Corollary 5.6 hence applies and gives $\pi_1(M) = \pi_1(X') = G$. $\square$

References

[1] **F Costantino**, **D Thurston**, *3–manifolds efficiently bound 4–manifolds*, J. Topol. 1 (2008) 703–745 MR

[2] **S Fischli**, **D Yavin**, *Which 4–manifolds are toric varieties?*, Math. Z. 215 (1994) 179–185 MR

[3] **S Matveev**, *Algorithmic topology and classification of 3–manifolds*, 2nd edition, Algorithms and Computation in Mathematics 9, Springer, Berlin (2007) MR

[4] **G Mikhalkin**, *Decomposition into pairs-of-pants for complex algebraic hypersurfaces*, Topology 43 (2004) 1035–1065 MR

[5] **M Passare**, **H Rullgård**, *Amoebas, Monge–Ampère measures, and triangulations of the Newton polytope*, Duke Math. J. 121 (2004) 481–507 MR

[6] **M Salvetti**, *Topology of the complement of real hyperplanes in* $\mathbf{C}^N$, Invent. Math. 88 (1987) 603–618 MR

[7] **V G Turaev**, *Quantum invariants of knots and 3–manifolds*, 3rd edition, de Gruyter Studies in Mathematics 18, de Gruyter, Berlin (2010) MR

[8] **F Waldhausen**, *Eine Klasse von 3–dimensionalen Mannigfaltigkeiten, I*, Invent. Math. 3 (1967) 308–333 MR

Department of Mathematics, Uppsala University
Box 480, SE-751 06 Uppsala, Sweden

Mathematics Department "Tonelli", Università di Pisa
Largo Pontecorvo 5, I-56127 Pisa, Italy

marco.golla@math.uu.se, martelli@dm.unipi.it

Received: 30 June 2015 Revised: 18 May 2016

Algebraic & Geometric Topology 17 (2017) 1445–1525

Heegaard Floer homology of spatial graphs

SHELLY HARVEY
DANIELLE O'DONNOL

We extend the theory of combinatorial link Floer homology to a class of oriented spatial graphs called transverse spatial graphs. To do this, we define the notion of a grid diagram representing a transverse spatial graph, which we call a graph grid diagram. We prove that two graph grid diagrams representing the same transverse spatial graph are related by a sequence of graph grid moves, generalizing the work of Cromwell for links. For a graph grid diagram representing a transverse spatial graph $f\colon G \to S^3$, we define a relatively bigraded chain complex (which is a module over a multivariable polynomial ring) and show that its homology is preserved under the graph grid moves; hence it is an invariant of the transverse spatial graph. In fact, we define both a minus and hat version. Taking the graded Euler characteristic of the homology of the hat version gives an Alexander type polynomial for the transverse spatial graph. Specifically, for each transverse spatial graph f, we define a balanced sutured manifold $(S^3 \setminus f(G), \gamma(f))$. We show that the graded Euler characteristic is the same as the torsion of $(S^3 \setminus f(G), \gamma(f))$ defined by S Friedl, A Juhász, and J Rasmussen.

57M15; 05C10

In memory of Tim Cochran

1 Introduction

Knot Floer homology, introduced by P Ozsváth and Z Szabó [18], and independently by J Rasmussen [20], is an invariant of knots in S^3 that categorifies the Alexander polynomial. Knot Floer homology is widely studied because of its many applications in low-dimensional topology. For example, it detects the unknot (Ozsváth and Szabó [17]), whether a knot is fibered (P Ghiggini [5] and Y Ni [15]) and the genus of a knot [17]. The theory was generalized to links by Ozsváth and Szabó [19]. The primary goal of this paper is to extend link Floer homology to a class of oriented spatial graphs in S^3, called transverse spatial graphs.

Originally, knot Floer homology was defined as the homology of a chain complex obtained by counting certain holomorphic disks in a $2g$–dimensional symplectic manifold

Published: 17 July 2017 DOI: 10.2140/agt.2017.17.1445

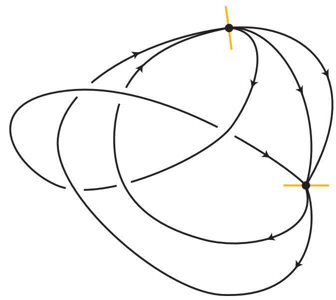

Figure 1: Example of a diagram of a transverse spatial graph; the lines at each vertex indicate the grouping of the incoming and outgoing edges.

with some boundary conditions that arose from a (doubly pointed) Heegaard diagram for S^3 compatible with the knot. As such, the chain *groups* were combinatorial but one could not, a priori, compute the boundary map. However, Sucharit Sarkar discovered a criterion that would ensure that the count of said holomorphic disks is combinatorial. This crucial idea was used by C Manolescu, P Ozsváth and S Sarkar [11] to give a combinatorial description of link Floer homology using grid diagrams. Using this description, C Manolescu, P Ozsváth, Z Szabó and D Thurston [12] gave a combinatorial proof that link Floer homology is an invariant. In this paper, we generalize the combinatorial description of Heegaard Floer homology and proof in [11; 12] to transverse spatial graphs. Specifically, we define a relatively bigraded chain complex (a combinatorial minus version) which is a module over $\mathbb{F}[U_1, \dots, U_V]$, where V is the number of vertices of the transverse spatial graph and $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$ is the field with two elements. We then show that it is well defined up to quasi-isomorphism. We note that, independently, Y Bao [1] defined a (noncombinatorial version of) Heegaard Floer homology for balanced bipartite spatial graphs with a balanced orientation; see Section 6 for the relationship to our theory.

Informally, a transverse spatial graph is an oriented spatial graph where the incoming (respectively outgoing) edges are grouped at each vertex and any ambient isotopy must preserve this grouping. See Figure 1 for an example. Details can be found in Section 2. To define the chain complex, we first introduce the notion of a graph grid diagram representing a transverse spatial graph in Section 3. Roughly, a graph grid diagram is an $n \times n$ grid of squares each of which is decorated with an X, an O (sometimes decorated with $*$) or is empty, and satisfies the following conditions. Like for links, there is precisely one O per row and column. There are no restrictions on the X's but if an O shares a row or column with multiple (or no) X's then it must be decorated with $*$. Moreover, each *connected component* must contain an O decorated with $*$. See Section 3.1 for a precise definition and Figure 2 for an example.

	O		X		
		O	X		
			O*	X	X
X				O	
X					O
O*	X	X			

Figure 2: Example of a graph grid diagram

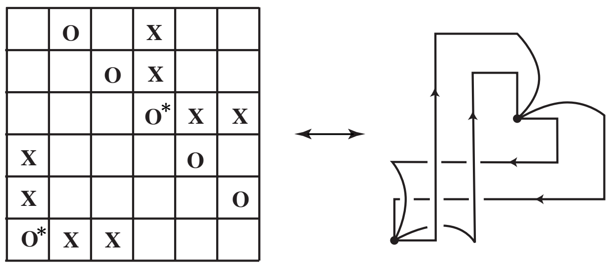

Figure 3: Associating a transverse spatial graph to a graph grid diagram

To each graph grid diagram we associate a transverse spatial graph by connecting the X's to the O's vertically and the O's to the X's horizontally with the convention that the vertical strands go over the horizontal strands. See Section 3.2 for more details and Figure 3 for an example.

We prove that every transverse spatial graph can be represented by a graph grid diagram.

Proposition 3.3 *Let $f\colon G \to S^3$ be a transverse spatial graph. Then there is a graph grid diagram g representing f.*

However, this representative is not unique. We define a set of moves on graph grid diagrams (cyclic permutation, commutation$'$, and stabilization$'$), called graph grid moves, generalizing the grid moves for links. See Section 3.2 for their definitions. We prove that any two representatives for the same transverse spatial graph are related by a sequence of graph grid moves.

Theorem 3.6 *If g and g' are two graph grid diagrams representing the same transverse spatial graph, then g and g' are related by a finite sequence of graph grid moves.*

In Section 4, to each (saturated) graph grid diagram g, we assign a chain complex $(C^-(g), \partial^-)$. Saturated means that there is at least one X per row and column, and these graph grid diagrams correspond to transverse spatial graphs where the underlying graphs have neither sinks nor sources. We need this technical condition to show that $\partial^- \circ \partial^- = 0$. The chain group consists of bigraded free $\mathbb{F}[U_1, \ldots, U_V]$–modules, where V is the number of O's decorated with $*$. Like in link Floer homology, the generators of the chain groups are unordered tuples of intersection points between the horizontal and vertical line segments in the grid, with exactly one point on each horizontal and vertical line segment. The Maslov grading is defined exactly as in [12]. Note that this is possible since it only depends on the set of O's on the grid. For links, the Alexander grading lives in $\mathbb{Z}^m$. For transverse spatial graphs, we define an Alexander grading that has values in $H_1(S^3 \setminus f(G))$, where $f\colon G \to S^3$ is the transverse spatial graph associated to g. To compute this, for each point in the lattice of the grid, we define an element of $H_1(S^3 \setminus f(G))$, called the generalized winding number. It is defined so that if you can get from one point to another by passing an edge of the projection of $f(G)$ coming from g, then the difference between their values is (plus or minus) the homology class of the meridian of that edge. The Alexander grading of a generator is obtained by taking the sum of the generalized winding numbers of the points of the generator. Each U_i is associated with an O, and we define the Alexander grading so that multiplication by U_i corresponds to lowering the Alexander grading by the element of $H_1(S^3 \setminus f(G))$ represented by the meridian of the O. See Section 4.3 for more details. The ∂^- map is defined by counting empty rectangles in the (toroidal) grid that do not contain X's. We show in Section 4 that $\partial^- \circ \partial^- = 0$ and so the homology of $(C^-(g), \partial^-)$ gives a well-defined invariant for each saturated graph grid diagram.

For a given transverse spatial graph, there are infinitely many graph grid diagrams representing it. In Section 5, we show that the homology of the bigraded chain complex is independent of the choice of saturated graph grid diagram. To prove this, we show that the quasi-isomorphism type of the chain complex is preserved under the three graph grid moves.

Theorem 4.22 *If g_1 and g_2 are saturated graph grid diagrams representing the same transverse spatial graph $f\colon G \to S^3$ then $(C^-(g_1), \partial^-)$ is quasi-isomorphic to $(C^-(g_2), \partial^-)$ as relatively absolutely $(H_1(E(f)), \mathbb{Z})$–bigraded $\mathbb{F}[U_1, \ldots, U_V]$–modules. In particular,* $\mathrm{HFG}^-(g_1)$ *is isomorphic to* $\mathrm{HFG}^-(g_2)$ *as relatively absolutely $(H_1(E(f)), \mathbb{Z})$–bigraded $\mathbb{F}[U_1, \ldots, U_V]$–modules.*

Thus, we can define the *graph Floer homology* of the sinkless and sourceless transverse spatial graph f, denoted $\mathrm{HFG}^-(f)$, to be $\mathrm{HFG}^-(g)$ for any saturated grid diagram g representing f. We also define a hat variant, $\widehat{\mathrm{HFG}}(f)$ by taking the homology of the chain complex obtained by setting $U_1, \ldots, U_V$ to zero. See Sections 4.2–4.4 for more details.

As far as the authors are aware, there have been no papers in the past that have defined an Alexander polynomial for an arbitrary spatial graph that do not just depend on the fundamental group of the exterior. Kinoshita [9] defines the Alexander polynomial of an oriented spatial graph as the Alexander polynomial of its exterior. In contrast, Litherland [10] defined the Alexander polynomial of a spatial graph where the underlying graph is a theta graph and his definition does not depend solely on the exterior of the embedding. At the same time as this paper was being written, Bao [1] independently defined an Alexander polynomial of a balanced bipartite spatial graphs with a balanced orientation and produces a state sum formula for the polynomial. Bao's invariant is essentially the same as ours; see Section 6 for more details.

In Section 6, we define an Alexander polynomial for any transverse spatial graph $f\colon G \to S^3$. To do this, we associate a balanced sutured manifold $(S^3 \setminus f(G), \gamma(f))$ to f. We then define the Alexander polynomial of f, $\Delta_f \in \mathbb{Z}[H_1(S^3 \setminus f(G))]$, to be the torsion invariant associated to balanced sutured manifold $(S^3 \setminus f(G), \gamma(f))$ defined by S Friedl, A Juhász, and J Rasmussen [4]. We show that the graded Euler characteristic of $\widehat{\mathrm{HFG}}(f)$ is essentially Δ_f. Note that $\bar{\Delta}_f$ is the image of Δ_f under the mapping that sends each element of $H_1(S^3 \setminus f(G))$ to its inverse.

Corollary 6.8 *If $f\colon G \to S^3$ is a sinkless and sourceless transverse spatial graph where G has no cut edges, then*

$$\chi(\widehat{\mathrm{HFG}}(f)) \doteq \bar{\Delta}_f.$$

That is, they are the same up to multiplication by units in $\mathbb{Z}[H_1(S^3 \setminus f(G))]$.

To prove this, we first prove the stronger result that the hat version of our graph Floer homology is essentially the same as the sutured Floer homology of $(S^3 \setminus f(G), \gamma(f))$. We say $\mathrm{rSHF}(E(f), \gamma(f))$ to mean $\mathrm{SFH}(E(f), \gamma(f))$ considered as a bigraded $(H_1(E(f)), \mathbb{Z}_2)$–module but with the $H_1(S^3 \setminus f(G))$ Alexander grading changed by a negative sign.

Theorem 6.6 *Let $f\colon G \to S^3$ be a sinkless and sourceless transverse spatial graph where G has no cut edges. Then*

$$\widehat{\mathrm{HFG}}(f) \cong \mathrm{rSHF}(E(f), \gamma(f))$$

as relatively $(H_1(E(f)), \mathbb{Z}_2)$–bigraded $\mathbb{F}$–vector spaces.

To complete the proof of Corollary 6.8, we use the theorem of S Friedl, A Juhász and J Rasmussen stating that the decategorification of sutured Floer homology is their torsion invariant [4].

Acknowledgements We would like to thank Sucharit Sarkar, András Juhász, Eamonn Tweedy, Tim Cochran, Katherine Vance, Adam Levine, Ciprian Manolescu, Matt Hedden, Robert Lipshitz, and Dylan Thurston for helpful conversations. We would also like to thank the referee for helpful comments.

Harvey was partially supported by National Science Foundation grants DMS-0748458 and DMS-1309070 and by grant #304538 from the Simons Foundation. O'Donnol was partially supported by National Science Foundation grants DMS-1406481 and DMS-1600365 and by an AMS-Simons travel grant.

2 Transverse disk spatial graphs

Graph Floer homology is a version of Heegaard Floer homology defined for transverse spatial graphs. In this section we will define the term *transverse spatial graph* and the notion of equivalence of transverse spatial graphs. We will also discuss their diagrams and Reidemeister moves.

We will work in the PL category. A *graph* Y is a 1–dimensional complex, consisting of a finite set of vertices (0–simplices) and edges (1–simplices) between them. A *spatial graph* $f\colon Y \to S^3$ is an embedding of a graph Y in S^3. A *diagram* of a spatial graph is a projection of $f(Y)$ to S^2 with only transverse double points away from vertices, where the over and under crossings are indicated. Two spatial graphs f_1 and f_2 are *equivalent* if there is an ambient isotopy between them. Notice that the ambient isotopy gives a map $h\colon S^3 \to S^3$ which sends $f_1(Y)$ to $f_2(Y)$ sending edges to edges and vertices to vertices.

Theorem 2.1 (Kauffman [8]) *Let f_1 and f_2 be spatial graphs. Then f_1 is ambient isotopic to f_2 if and only if any diagram of f_2 can be obtained from any diagram of f_1 by a finite number of graph Reidemeister moves (shown in* Figure 4*) and planar isotopy.*

An *oriented graph* is a graph together with orientations given on each of the edges. Let $\mathcal{D}$ be the 2–complex obtained by gluing three copies of the 2–simplex $[e_0, e_1, e_2]$ together so that their union is a disk and with all three of the e_0's identified to a single point in the interior of $\mathcal{D}$. Note that e_0 is the unique vertex in the interior of $\mathcal{D}$. We say that $\mathcal{D}$ is a *standard disk* and e_0 is the *vertex associated to* $\mathcal{D}$. An *oriented disk graph* G is a 2–complex constructed as follows. Start with an oriented graph Y. Then, for each vertex v of Y, glue a standard disk $\mathcal{D}$ to Y by identifying the vertex associated to $\mathcal{D}$ with v. We note that Y is a subset of the oriented disk graph, which we call the *underlying oriented graph* of the oriented disk graph (or the *underlying graph* of G if we do not want to consider the orientations). We say that a vertex of an oriented disk

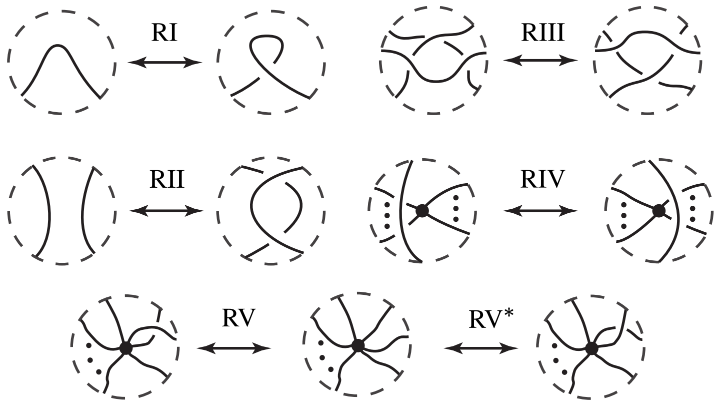

Figure 4: The graph Reidemeister moves. Reidemeister moves RI, RII, and RIII are the same as those for knots and links. Reidemeister move RIV moves an edge past a vertex, either over or under (over is pictured here). Reidemeister move RV swaps two of the edges next to the vertex.

graph is a *graph vertex* (respectively *graph edge*) if it is a vertex (respectively edge) of its underlying oriented graph. When it is clear, we will just refer to them as vertices and edges of the oriented disk graph (and will not refer to the other 0 and 1–simplices of the oriented disk graph as vertices or edges). For an oriented disk graph G and a given graph vertex v of G, the set of graph edges of G with orientation going towards v are called the *incoming edges* of v and the set of edges with the orientation going away from v are called the *outgoing edges* of v.

Definition 2.2 A *transverse spatial graph* is an embedding $f\colon G \to S^3$ of an oriented disk graph G into S^3 where each vertex of the graph locally looks like Figure 5, each standard disk of G lies in a plane, and locally the disk separates the incoming and outgoing edges of the given vertex. We call the image of each of the standard disks of G a *disk* of f, and the embedding of the underlying graph of G the *underlying spatial graph* of f. Two transverse spatial graphs are *equivalent* if there is an ambient isotopy between them.

Note that in a transverse spatial graph the incoming and outgoing edges are each grouped together. In the ambient isotopy, at each vertex, both the set of incoming and the set of outgoing edges can move freely. However, in an ambient isotopy, the incoming and outgoing sets cannot intermingle, because the disk separates the edges.

Definition 2.3 A *regular projection* of a transverse spatial graph $f\colon G \to S^3$ is a projection that satisfies the following two conditions: (1) For each point p in the

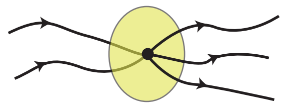

Figure 5: A vertex of a transverse spatial graph shown with the disk

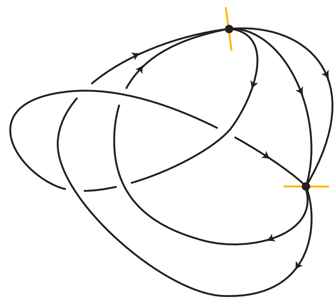

Figure 6: A diagram of a transverse spatial graph; the projections of the disks, transverse to the plane of projection, are indicated by the straight yellow line segments.

image of the underlying graph of G, $f^{-1}(p)$ contains no more than two points and if $f^{-1}(p)$ contains two points then neither is a graph vertex. (2) All of the standard disks of f are perpendicular to the plane of projection. A *diagram* for a transverse spatial graph $f\colon G \to S^3$ is a regular projection of f where all the over and under crossings are indicated.

In Figure 6, a diagram of a transverse spatial graph is shown, where the disks are also shown in the projection. Notice that the incoming edges and outgoing edges are grouped in the projection. We will from here forward not indicate disks in diagrams, because the position of the disk is already clear from the diagram.

The Reidemeister moves for transverse spatial graphs are the same as the Reidemeister moves for graphs shown in Figure 4, with the restriction that RV may *only* be made between pairs of incoming edges or pairs of outgoing edges. For clarity, we will say $\mathrm{R}\overline{\mathrm{V}}$ for this restriction of RV.

Theorem 2.4 *Every transverse spatial graph has a diagram. If two transverse spatial graphs are ambient isotopic, then any two diagrams of them are related by a finite sequence of the Reidemeister moves* RI–RIV, $\mathrm{R}\overline{\mathrm{V}}$ *and planar isotopy.*

Proof We first show that every transverse spatial graph has a diagram. Let $f\colon G \to S^3$ be a transverse spatial graph. A regular projection for the transverse spatial graph is a

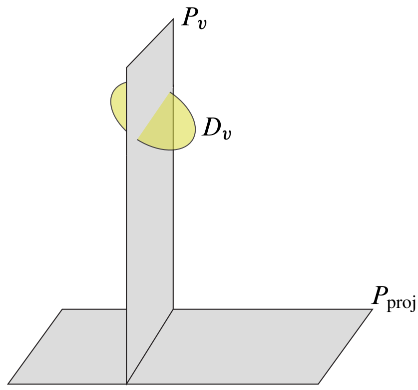

Figure 7: The plane of projection P_{proj} together with a disk D_v of the vertex v and P_v; the plane P_v is the plane perpendicular to P_{proj} meeting D_v in the line parallel to P_{proj}.

projection of $f(G)$ which is a regular projection of the underlying spatial graph of f. That is, the projection only has transverse double points and these are away from the vertices. We will first obtain a regular projection for the transverse spatial graph, and next find a representative in the ambient isotopy class where the disks are transverse to P_{proj}, where P_{proj} denotes the plane of projection for the regular projection. A regular projection for a spatial graph is obtained in the usual way: a point projection of a representative of the ambient isotopy class is obtained via ϵ–perturbations of the graph. To have the disks perpendicular to P_{proj}, a similar process is used. If the disk is transverse to P_{proj}, we will see that there is a unique way to move it via an ambient isotopy of f to a position where it is perpendicular to P_{proj}. So we need only have all of the disks transverse to P_{proj}. For an arbitrary vertex v with disk D_v, let $\mathbf{x}$ be the vector that is perpendicular to D and pointing in the direction of the outgoing edges. If v remains in the same place and the neighborhood around it is allowed to rotate, there is a full sphere of directions in which $\mathbf{x}$ can be pointing. Only two of these directions will result in D_v being parallel to P_{proj}. By dimensionality arguments having the disks transverse to P_{proj} is generic. If any of the disks are not transverse, an ϵ–perturbation is done. For each vertex v, let P_v be the plane that is perpendicular to P_{proj} and meets D_v in the line through v and parallel to P_{proj}. For each disk transverse to P_{proj} there is a unique map via rotation through the acute angle between D_v and P_v, moving D_v into the plane P_v, so that it is perpendicular to P_{proj}; see Figure 7.

For (topological) spatial graphs, any ambient isotopy is made up of elementary moves. Recall that an elementary move of a spatial graph replaces a linear segment of an edge $[e_i, e_j]$ by two new linear segments $[e_i, e_k]$ and $[e_k, e_j]$ that, together with $[e_i, e_j]$, bound a 2–simplex which intersects the original spatial graph only in $[e_i, e_j]$, or is

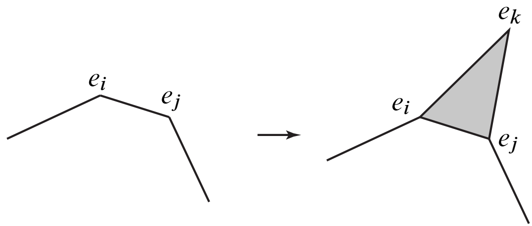

Figure 8: An elementary move

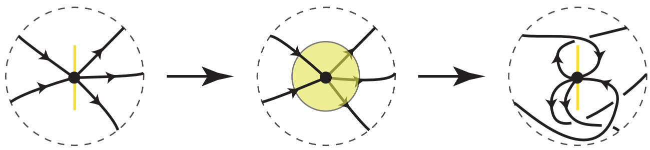

Figure 9: An example of what can happen when a disk flips over, moving through a position parallel to P_{proj}

the reverse of this move; see Figure 8. Kauffman [8] showed that any elementary move for a spatial graph can be obtained by a sequence of the Reidemeister moves shown in Figure 4. Now we consider transverse spatial graphs. An elementary move of a transverse spatial graph $f\colon G \to S^3$ replaces a linear segment of an edge of the underlying graph $[e_i, e_j]$ by two new linear segments $[e_i, e_k]$ and $[e_k, e_j]$ that, together with $[e_i, e_j]$, bound a 2–simplex T which intersects $f(G)$ in $[e_i, e_j]$, or is the reverse of this move; see Figure 8. We note that T must miss all the transverse disks. We will show that any ambient isotopy of transverse spatial graphs is made of elementary moves of a transverse spatial graph. First note that Reidemeister moves RI–RIV preserve the isotopy class of a transverse spatial graph. In addition, one can still interchange a pair of neighboring incoming edges or a pair of neighboring outgoing edges in RV. However, if one tried to interchange an incoming with a neighboring outgoing edge at the vertex v, the disk from the elementary move that would result in RV would intersect the transverse disk D_v. So Reidemeister move RV is restricted to R$\overline{\text{V}}$. Recall that R$\overline{\text{V}}$ is the move RV where only neighboring incoming (respectively outgoing) edges are interchanged.

We claim that one needs only Reidemeister moves RI–RIV and R$\overline{\text{V}}$ to get all ambient isotopies. One might be concerned that this is incomplete because of the danger of a vertex flipping over (ie moving through a position where the disk is parallel to the plane of projection) resulting in a change in the diagram like that shown in Figure 9. However, this move and any move like it can be obtained with the set of Reidemeister moves RI–RIV, and R$\overline{\text{V}}$. To discuss this we will introduce another type of graph. A *flat vertex graph* or *rigid vertex graph* is a spatial graph where the vertices are flat disks or polygons with edges attached along the boundary of the vertex at fixed places. The set of Reidemeister moves for flat vertex graphs is RI–RIV as before, and the move RV* [8]. Reidemeister

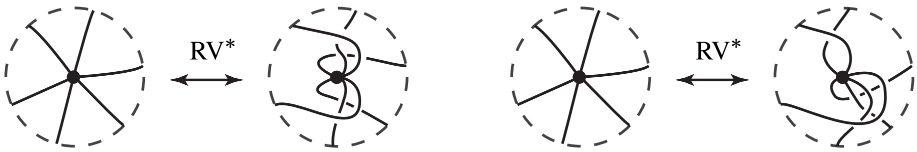

Figure 10: Two different RV* moves for flat vertex graphs

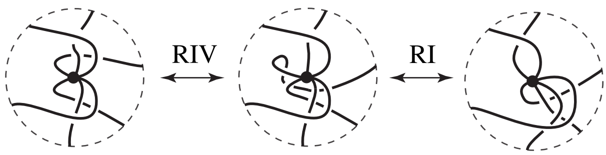

Figure 11: How to move between the two different RV* moves shown in Figure 10; thus only one RV* move is needed for each valence of vertices

Figure 12: This shows how many $R\overline{V}$ moves will give a RV* move

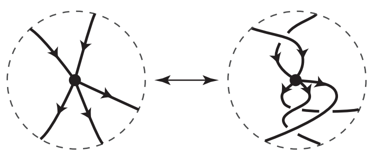

Figure 13: The RV* move that is a result of many $R\overline{V}$ moves

move RV* is the flipping over of a flat vertex by $180°$; see Figure 10. For the move RV* a choice is made of how many edges are on each side when the vertex is flipped, but only one of these moves is need together with Reidemeister moves RI–RIV to do any of the other ones [8]; see Figure 11. In the case of transverse spatial graphs repeated use of $R\overline{V}$ will result in what looks like a RV* move, shown in Figures 12 and 13. Thus flipping the vertex over can be accomplished by Reidemeister moves RI–RIV and $R\overline{V}$. □

3 Graph grid diagrams

In this section, we define the notion of graph grid diagrams and explain their relationship to transverse spatial graphs. To each graph grid diagram we associate a unique transverse spatial graph. On the other hand, we show that every transverse spatial graph can be represented by a nonunique graph grid diagram. As with grid diagrams for knots and links, we define a set of moves on graph grid diagrams (cyclic permutation, commutation$'$, and stabilization$'$) that we call graph grid moves. Finally, we prove that any two graph grid diagrams representing the same transverse spatial graph are related by a sequence of graph grid moves.

3.1 Graph grid diagrams

We will assume that the reader is familiar with grid diagrams for knots and links; see [11; 12]. Recall that a (planar) grid diagram for a link is an $n \times n$ grid of squares in the plane where each square is decorated with an X, an O or nothing, and such that every row (respectively column) contains exactly one X and exactly one O. Here we are using the notation of [12]. To each grid diagram, one can associate a planar link diagram by drawing horizontal line segments from the O's to the X's in each row, and vertical line segments from the X's to the O's in each column with the convention at the crossings that a vertical segment always goes over a horizontal segment. We will define a more general class of grid diagrams that will represent transverse spatial graphs. Before defining a grid diagram we need a technical definition.

Definition 3.1 Suppose D is an n by n grid where each square is decorated with an X, an O or is empty. We let $\mathbb{X}$ be the set of X's and $\mathbb{O}$ be the set of O's. We say that two elements $p, q \in \mathbb{X} \cup \mathbb{O}$ are *related* if p and q share a row or column. Let $\sim$ be the equivalence relation generated by this relation. We define the *connected components* of D to be the equivalence classes of $\sim$.

Definition 3.2 A *graph grid diagram* g is an n by n grid where each square is decorated with an X, O or is empty, a subset of the O's are decorated with $*$, and that satisfies the following conditions. There is exactly one O in each row and column. Each connected component contains at least one O decorated with $*$. If a row or column does not contain exactly one X then the O in that row or column must be decorated with $*$. The total number of rows (equivalently columns) n is called the *grid number* of g. The O's decorated with $*$ are called *vertex* O*'s*, the number of which will be denoted V. We will say that an O is *standard* if the O has exactly one X in its row and exactly one X in its column; otherwise we say it is *nonstandard*. Often, it will be convenient to number the O's and X's by $\{O_i\}_{i=1}^n$ and $\{X_i\}_{i=1}^m$. When numbering, we always assume that $O_1, \dots, O_V$ correspond to the vertex O's.

For convenience, we may sometimes omit the $*$ from a figure when it is clear which O's should have $*$, ie the nonstandard ones. It will also be convenient to think of the grid as the set $[0, n] \times [0, n]$ in the plane, with vertical and horizontal grid lines of the form

$\{i\} \times [0, n]$ and $[0, n] \times \{i\}$, where i is an integer from 0 to n, and the X's and O's are at half-integer coordinates.

As in [11; 12], our chain complex is obtained from a graph grid diagram, with the main difference being the definition of the Alexander grading. To define this, it is sometimes necessary to consider toroidal graph grid diagrams instead of (planar) graph grid diagrams. A *toroidal graph grid diagram* is a graph grid diagram that is considered as being on a torus by identifying the top and bottom edges of the grid and identifying the left and right edges of the grid. We denote the toroidal graph grid diagram by $\mathcal{T}$. We view the torus as being oriented and the orientation being inherited from the plane. When the context is clear, we will just call it a graph grid diagram. In a toroidal graph grid diagram, the horizontal and vertical grid lines, become circles. We denote the horizontal circles by $\alpha_1, \dots, \alpha_n$, the vertical circles $\beta_1, \dots, \beta_n$ and we let $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_n\}$ and $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_n\}$. When the grid is drawn on a plane, by convention, we will order the horizontal (respectively vertical) circles from bottom to top (respectively left to right) so that the leftmost circle is β_1 and the bottommost circle is α_1. Note that to get a (planar) diagram from a toroidal diagram, one takes a *fundamental domain* for the torus and cuts along a horizontal and vertical grid circle and identifies it with $[0, n) \times [0, n)$.

3.2 Graph grid diagram to transverse spatial graphs and their diagrams

Let g be a graph grid diagram. We can associate a transverse spatial f to g as follows. First put a vertex at each of the O's that are decorated with $*$. Let O_i be an O in g lying in row r_i and column c_i. For each X_j in row r_i, connect O_i to X_j with an arc inside of the row (oriented from O_i to X_j) so that it is disjoint from all the X's and O's and so that all the arcs in row r_i are disjoint from one another. We will call these horizontal arcs. Now push the interior of the arcs in row r_i slightly upwards, above the plane. For each X_j in column c_i, connect X_j to O_i with an arc inside of the column (oriented from X_j to O_i) so that it is disjoint from all the X's and O's and so that all the arcs in row c_i are disjoint from one another. We will call these vertical arcs. Now push the interior of the arcs in column c_i slightly downward, below the plane. Put a disk in the squares containing O's decorated with $*$ and the vertex at the disks center. In this case we say that the graph grid diagram g *represents* the spatial graph f. Any choice of arcs gives the same transverse spatial graph.

Note that the aforementioned procedure will actually give us a (nonunique) projection of the transverse spatial graph. However, this will not be a diagram of f since the transverse disks will be parallel to the plane of projection. It will be convenient for us to define a class of grid diagrams that give a well-defined diagram of a transverse spatial graph when following this procedure.

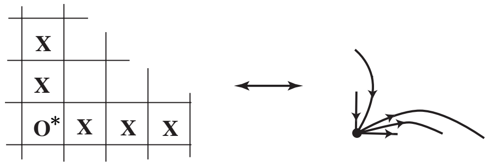

Figure 14: A graph grid diagram of a nonstandard O with the flock in L–formation (left), and the diagram of the associated vertex for this portion of the graph grid diagram, showing the order in which the edges appear around the vertex (right)

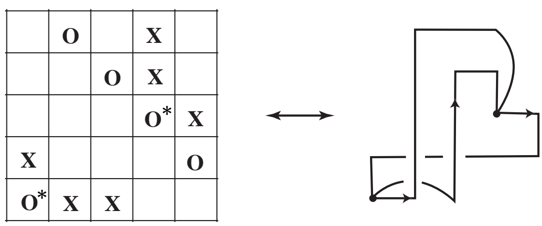

Figure 15: A preferred grid diagram (left), and the diagram of the transverse spatial graph associated with this graph grid diagram (right)

Consider a graph grid diagram. For a nonstandard O, let the set of X's that appear in a row or column with this O be called its *flock*. If the X's in the flock of an O are all adjacent to the O or adjacent to other X's that are adjacent to the O, then the flock is said to be *clustered*. A flock is in *L–formation*, if the X's are all to the right and above the O. It should be noted that the choice of having the X's above and to the right of the O is arbitrary; any pair of below and to the right, above and to the left, or below and to the left will work similarly. A *preferred* graph grid diagram is a graph grid diagram where all nonstandard O's (with more than one X in it's flock) have their flocks in L–formation.

Now suppose that g is a preferred graph grid diagram. We follow the procedure in the first paragraph of this section except that now we put the transverse disks perpendicular to the plane so that they divide the horizontal and vertical arcs. Moreover, to get a unique diagram D, the ordering of the edges around the vertex for nonstandard O's is given by the convention illustrated in Figure 14. In this case we say that the *graph grid diagram g represents the diagram of the transverse spatial graph*, D. We note that the transverse spatial graph associated to this diagram is equivalent to the transverse spatial graph obtained by following the procedure in the first paragraph of this section.

3.3 Graph embedding to preferred graph grid diagram

We have shown that to each graph grid diagram, we can associate a transverse spatial graph. We now show that for each transverse spatial graph, there is a (preferred) graph grid diagram representing it.

Proposition 3.3 *Let $f\colon G \to S^3$ be a transverse spatial graph. Then there is a preferred graph grid diagram g representing f. Moreover, for each diagram of a transverse spatial graph, D, there is a preferred graph grid diagram g representing D.*

Proof Choose a diagram D of the transverse spatial graph f. We construct a graph grid diagram g, representing D, by the following procedure. At the vertices, the edges are partitioned into two sets: incoming edges and outgoing edges, as D is a diagram of a transverse spatial graph. Move the edges around each vertex (and perhaps the disk) by planar isotopy so that all outgoing edges are to the right of the vertex and all incoming edges are above the vertex, as shown in Figure 14. Away from vertices the process is the same as that for knots or links. The arcs of the edges are made "square". All crossings are made so that the horizontal arc goes under the vertical arc; see Figure 16. Then the diagram is moved via a planar isotopy so that no vertical arcs or vertices with their incoming edge arcs are in the same vertical line, and similarly for horizontal arcs and vertices with their outgoing edge arcs. A vertex along with its incoming edge arcs are associated with a single column, and the vertex together with its outgoing edge arcs are associated with a single row; see Figure 14. Each of the vertical and horizontal arcs are also associated with a column and row of the grid, respectively. This will result in an equal number of rows and columns. Each vertex will add a row and a column. Each vertical arc that is not next to a vertex will add a column. Each vertical arc can be paired with the following horizontal arc that is not next to a vertex, which will add a row. A graph grid representation is then given by placing X's and O's on the n by n grid. At each vertex a single O^* is placed, then an X is placed in the same row at the corner of each of the outgoing edges and an X is placed in the same column at the corner for each of the incoming edge, this is done as shown in Figure 14. Next X's and O's are placed along the edges at the corners consistent with the orientation: arcs go from X's to O's in columns and from O's to X's in rows. □

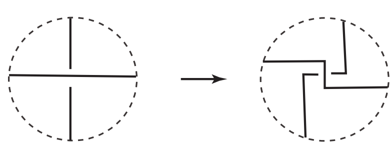

Figure 16: How to change a horizontal over-crossing to a vertical over-crossing without changing the embedded graph

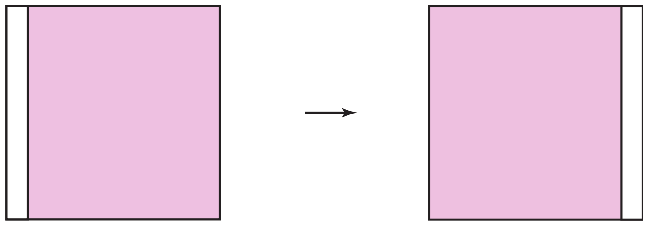

Figure 17: An example of cyclic permutation of columns

3.4 Grid moves

Following Cromwell [2] and Dynnikov [3], any two grid diagrams of the same link are related by a finite sequence of grid moves:

Cyclic permutation The rows and columns can be cyclically permuted; see Figure 17.

Commutation Pairs of adjacent columns (respectively rows) may be exchanged when the following conditions are satisfied. For columns, the four X's and O's in the adjacent columns must lie in distinct rows, and the vertical line segments connecting O and X in each column must be either disjoint or nested (one contained in the other) when projected to a single vertical line. There is an obvious analogous condition for rows; see Figure 18.

Stabilization/destabilization Let g be an $(n-1)\times(n-1)$ graph grid diagram with decorations $\{\mathrm{O}_i\}_{i=1}^{n-1}$ and $\{\mathrm{X}_j\}_{j=1}^{n-1}$. Then $\bar{g}$, an $n\times n$ graph grid diagram, is a stabilization of g if it is obtained from g as follows. Suppose there is a row of g that contains O_i and X_j. In $\bar{g}$, we replace this one row with two new rows and add one new column. We place O_i into one of the new rows (and in the same column as before) and X_j into the other new row (and in the same column as before). We place decorations O_n and X_n into the new column so that O_n occupies the same row as X_j and X_n occupies the same row as O_i. See Figure 19 for an example. There is a similar move with the roles of columns and rows interchanged. A destabilization is the reverse of a stabilization.

For the graph grid moves there are two differences. We will replace the usual commutation with a slightly more general commutation$'$ to include exchanging neighboring columns which have entries in the same row (or rows with entries in the same columns) and to include exchanges of rows and columns that have more than a single X in them (or no X's). We will also restrict the stabilization/destabilization move to only occur along edges (which we explain below).

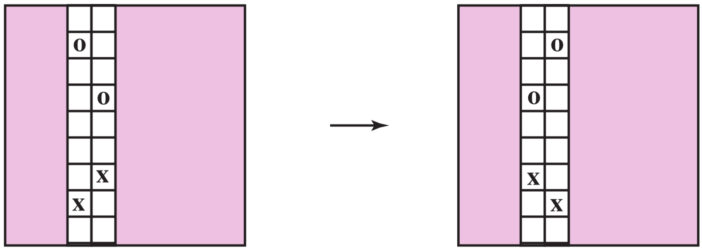

Figure 18: An example of commutation of columns

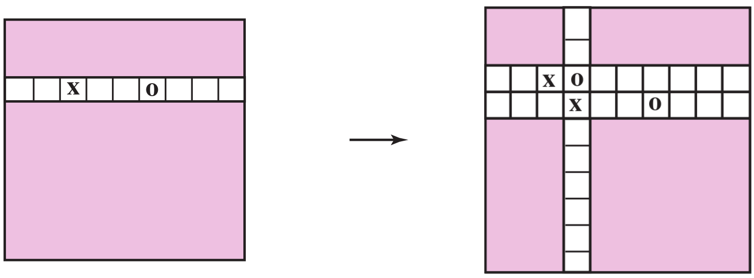

Figure 19: An example of stabilization

The graph grid moves

Cyclic permutation is unchanged.

> **Cyclic permutation** The rows and columns can be cyclically permuted; see Figure 17.

Commutation will be replaced with the more general commutation$'$.

> **Commutation$'$** Pairs of adjacent columns may be exchanged when the following conditions are satisfied. There are vertical line segments LS_1 and LS_2 on the torus such that (1) $\mathrm{LS}_1 \cup \mathrm{LS}_2$ contain all the X's and O's in the two adjacent columns, (2) the projection of $\mathrm{LS}_1 \cup \mathrm{LS}_2$ to a single vertical circle β_i is β_i, and (3) the projection of their endpoints, $\partial(\mathrm{LS}_1) \cup \partial(\mathrm{LS}_2)$, to a single β_i is precisely two points. Here we are thinking of $\mathbb{X}$ and $\mathbb{O}$ as a collection of points in the grid with half-integer coordinates. There is an obvious analogous condition for rows; see Figure 20.

We define a generalization of stabilization, called stabilization$'$. This move will add a jog or a nugatory crossing to the edge of the projection of the associated transverse spatial graph.

> **Stabilization$'$/destabilization$'$** Let g be an $(n-1)\times(n-1)$ graph grid diagram with decorations $\{\mathrm{O}_i\}_{i=1}^{n-1}$ and $\{\mathrm{X}_j\}_{j=1}^{m-1}$. Then $\bar{g}$, an $n\times n$ graph grid diagram,

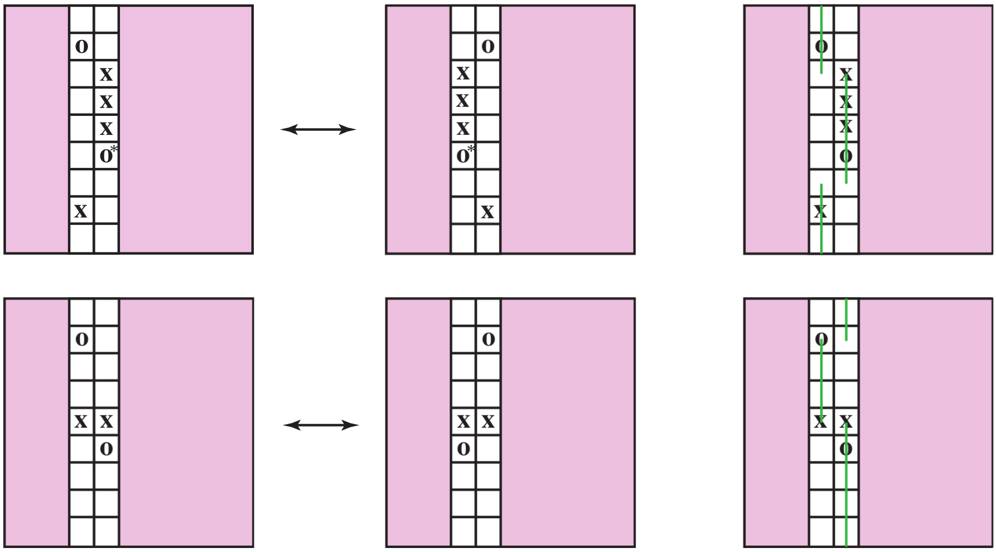

Figure 20: Two examples of commutation′ moves of columns

is a row stabilization′ of g if it is obtained from g as follows. Suppose there is a row of g that contains the decorations $\mathrm{O}_k, \mathrm{X}_{j_1}, \dots \mathrm{X}_{j_l}$ with $l \geq 1$. In $\bar{g}$, we replace this one row with two new rows and add one new column. We place $\mathrm{O}_k, \mathrm{X}_{j_2}, \dots, \mathrm{X}_{j_l}$ into one of the new rows (and in the same column as before) and X_{j_1} into the other new row (and in the same column as before). We place decorations O_n and X_m into the new column so that O_n occupies the same row as X_{j_1} and X_m occupies the same row as O_k. See Figure 21 for an example. A column stabilization′ is a row stabilization′ where one reverses the roles of rows and columns. We say that $\bar{g}$ is obtained from g by a *stabilization′* if it is obtained by a row or column stabilization′. A destabilization′ is the reverse of a stabilization′.

Note that O_n will not be associated to a vertex so will not be decorated with $*$. Also, if any O_i is decorated with $*$ (including O_k) in g then it will also be decorated with $*$ in $\bar{g}$. We do not allow stabilization′ of rows with no X's in them.

Remark 3.4 If $\bar{g}$ is obtained as row stabilization′ on the graph grid diagram g then one can use multiple commutation′ moves to change $\bar{g}$ into a row stabilization′ obtained from g, where $\mathrm{X}_{j_1}, \mathrm{X}_m, \mathrm{O}_n$ share a corner, X_{j_1} is directly to the left of O_n, and O_n is directly above X_m (as in Figure 21). Note that by using only commutation, like in [12], one can only assume that $\mathrm{X}_{j_1}, \mathrm{X}_m, \mathrm{O}_n$ share a corner, which leaves four cases instead of one. This will allow us to simplify the proof of stabilization′. There is a similar statement for column stabilization′.

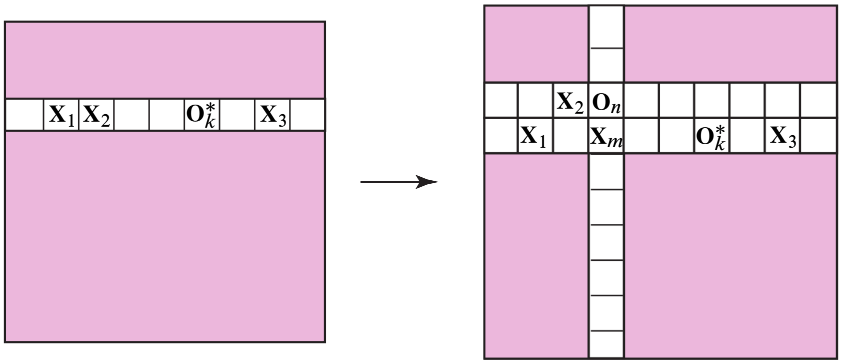

Figure 21: An example of stabilization′

3.5 The graph grid theorem

Before the main theorem of this section we need a lemma. To each diagram of a transverse spatial graph f there are an infinite number of different graph grid diagrams representing f that can be constructed using the procedure described in the proof of Proposition 3.3. This procedure produces a preferred grid diagram. However, doing a graph grid move on a preferred graph grid diagram will result in diagrams that are not necessarily in preferred form. Moreover, if one chooses a random graph grid diagram representing a transverse spatial, it will not necessarily be in preferred form. Indeed, in practice, one can often reduce the size of the grid number by moving it out of preferred form.

Lemma 3.5 *Every graph grid diagram g representing a transverse spatial graph f is related to a preferred graph grid diagram representing f by a finite sequence of graph grid moves.*

Proof Recall that a preferred grid diagram is one in which all of the nonstandard O's have their flocks in L–formation. Given a graph grid diagram, choose a nonstandard O that is not in L–formation. We will explain an algorithm to move this O into L–formation, but first we must separate this flock from any of the other flocks that are in L–formation. If there are any X's in the flock with this O that are also in a flock of another O that is in L–formation, then the other L–formation flock will need to be moved. If our nonstandard O of interest is in a column with an X that is in L–formation with another nonstandard O, we use the following procedure to move the flock out of the way. An example of this is shown in Figure 22. We do a row stabilization′ at said X; the new row is placed below the L–formation flock. Now the nonstandard O's no longer share an X, but if there were any X's to the right of the previously shared X,

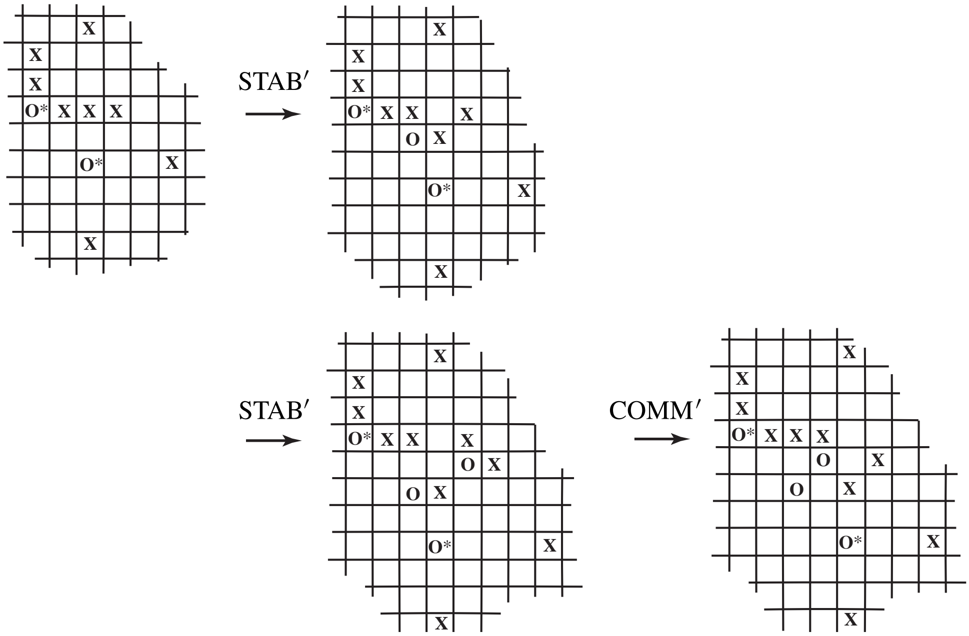

Figure 22: An example of the moves needed to separate the flocks of two O^* and move the upper most flock back into L–formation

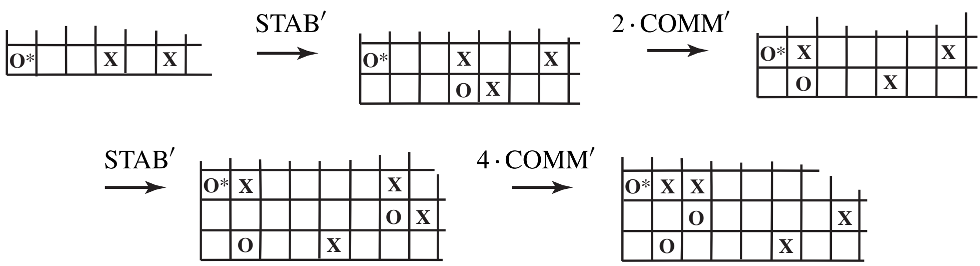

Figure 23: An example of the moves needed to move the X in the row into L–formation

the flock was split by the stabilization′ move and so it is no longer in L–formation. To move the flock back into L–formation a stabilization′ move is done at each of the X's to the right of the split (going from left to right), each time adding a row below the flock. After the stabilization′ moves are done, the columns containing an X in the flock can be moved by commutation′ moves to be next to the other X's in the flock. This is done with all of the X's, so the flock is in L–formation again. If the X shares a row with our O and a column with a different nonstandard O that is in L–formation, a similar procedure is done with the roles of the rows and columns switched.

Suppose there is no X in the flock that is already in L–formation with a different O. Use cyclic permutation to put the O at the lower left corner of the grid. Then, a row stabilization$'$ is done on the rightmost X that is in the row with the O, adding a row to the bottom of the diagram and adding an X and O next to each other in the new column. Commutation$'$ can be used to move the new column next to the nonstandard O, or next to X's that are next to the O; see Figure 23. This is repeated until all of the X's in the row are adjacent to the O. A similar process is done with the X's in the column of the O, bringing the O into L–formation. This process can be repeated until all of the O's are in L–formation.

This will increase the number of nonstandard O's in L–formation, because no other flock is moved out of L–formation. We continue until all flocks are in L–formation. □

For the following proof, we need a few more definitions. If an O^* is associated with a vertex v and is in L–formation, then all of the columns that contain an X in the flock are called *v–columns*, similarly those rows containing the flock are called *v–rows*. We will give a name to certain sequences of the graph grid moves, which will be called (column or row) *vertex stabilization* (and *destabilization*). A column (or row) *vertex stabilization* introduces a stabilization to the left of (or below) all of the X's in the column (or in the row) with a nonstandard O, as shown in Figure 24. The row vertex stabilization is a combination of a number of stabilization$'$s and commutation$'$s. For a nonstandard O, first a row stabilization$'$ is done, where the rightmost X is placed into the lower new row by itself, the new column is placed to the left and the new X and O are added. Next, the second from the right X is moved by commutation$'$ so that it is in the rightmost position. A stabilization$'$ move is done in the same way. Then commutation$'$ moves are done on the rows, moving the newest row directly below the flock, below the rows created in the stabilization$'$s that happened before. Finally, the X in the flock is moved back to the original place in the flock via commutation$'$. Follow the same procedure for all of the X's in the row.

Theorem 3.6 *If g and g' are two graph grid diagrams representing the same transverse spatial graph, then g and g' are related by a finite sequence of graph grid moves.*

Proof First, using Section 3.5 we move both g and g' to preferred graph grid diagrams. We know that the diagrams of two isotopic transverse spatial graphs are related by a finite sequence of the graph Reidemeister moves, RI–RIV and R$\overline{\text{V}}$, shown in Figure 4, together with planar isotopy. So we need only show that preferred graph grid diagrams that result from embeddings that differ by a single Reidemeister move (or planar isotopy) can be related by a finite sequence of graph grid moves.

Due to the work of Cromwell [2] and Dynnikov [3], it is known that any two grid diagrams of the same link are related by a finite sequence of grid moves, cyclic

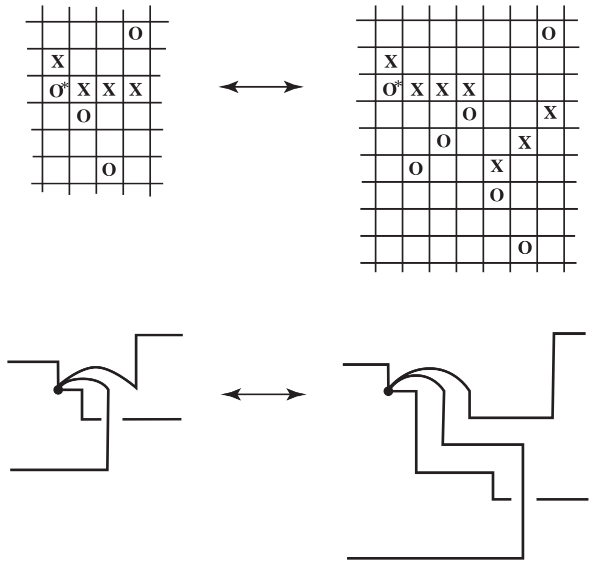

Figure 24: An example of a row vertex stabilization

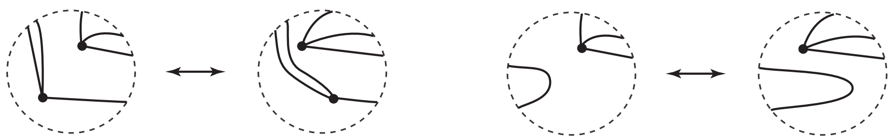

Figure 25: Examples of the two planar isotopies that can occur which contain vertices: two vertices are moved parallel to each other (left), and an arc and a vertex move parallel to each other (right)

permutation, commutation, and stabilization/destabilization. The Reidemeister moves are local moves. In the grid diagram there is a set of columns and or rows that will be moved to accomplish any one of RI–RIII. Because the first three Reidemeister moves do not involve vertices and we are working with preferred diagrams, the rows and columns that are moved will not contain an O^*. It could however contain rows or columns that contain X's that are in a flock with an O^*. In this case, first a vertex stabilization is done, so that the flock is not disrupted and the graph grid stays in preferred formation. Thus we need only show that any two preferred graph grid diagrams that come from the same embedding and differ as a result of a single Reidemeister move or planar isotopy which involves vertices can be related by a finite sequence of graph grid moves.

There are two moves and three planar isotopies with vertices to be considered: RIV, $\mathrm{R}\overline{\mathrm{V}}$, a planar isotopy in which a valence two-vertex is moved along the arc of the edges, a planar isotopy in which two vertices are moved parallel to each other, and a planar

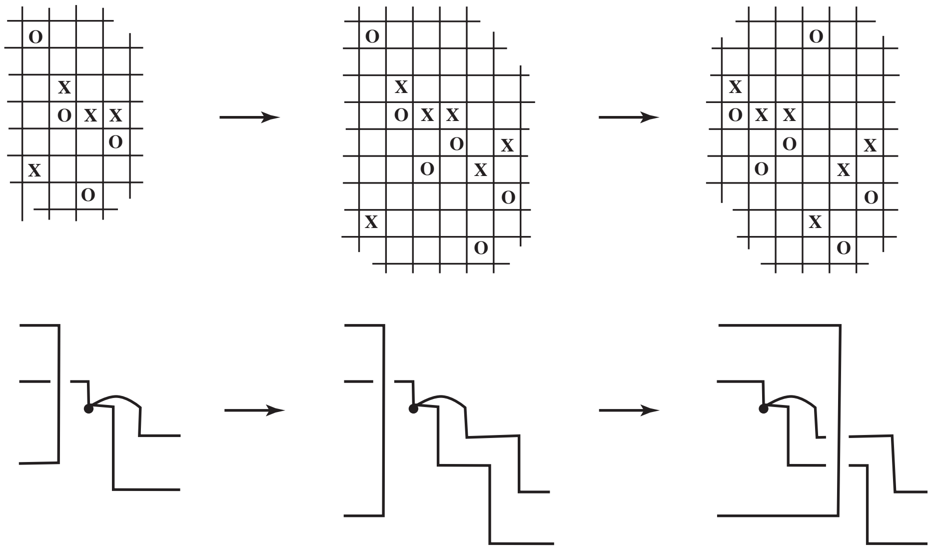

Figure 26: An example of a vertex stabilization move followed by three commutation′ moves, resulting in a RIV move in the associated graph

isotopy in which an arc and a vertex move parallel to each other, shown in Figure 25. For the figures of the graph grid diagrams in this proof we will only place $*$ on an O if it is not obvious from the grid that it is an O^*.

RIV move The move RIV moves an arc from one side of a vertex to the other, either over or under the vertex. Up to planar isotopy we can assume that the edge is next to the vertex that it will pass over (or under). An example of RIV is shown in Figure 26, here a row vertex stabilization is done followed by three column commutation′ moves. In general, RIV can be obtained via the following: first a vertex stabilization move, if needed, then a number of commutation′ moves between the v–columns (resp. v–rows) and other column (resp. row) that is associated with an appropriate arc.

R$\overline{\textbf{V}}$ move The R$\overline{\text{V}}$ move corresponds to switching the order of the edges in the projection next to the vertex, which introduces a crossing between these edges. See the leftmost move in Figure 12 for reference. Since we are working with transverse spatial graphs, such a move can only occur between pairs of incoming edges and outgoing edges. We will look at the graph grid moves needed for an R$\overline{\text{V}}$ move between two outgoing edges. The proof is similar for two incoming edges.

In general, a commutation′ move between columns or rows that contain X's in the same flock will result in a R$\overline{\text{V}}$ move between the two associated edges involved. In order to be able to iterate such moves, we present the follow processes. In an R$\overline{\text{V}}$ move, two edges are switched next to a vertex. Let's call one of them the left edge and one

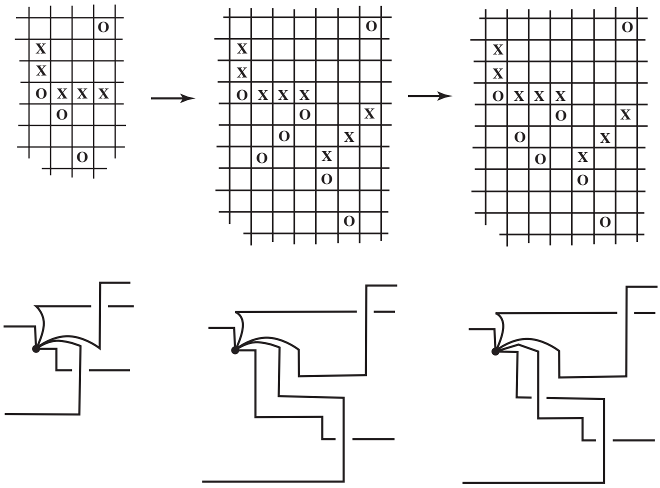

Figure 27: A row vertex stabilization move, followed by a commutation′ move in the grids, producing a R$\overline{\text{V}}$ move in the associated spatial graph

the right edge. There are two possibilities with a R$\overline{\text{V}}$ move: the right edge either goes under or over the left edge.

In Figure 27, we show an example of R$\overline{\text{V}}$ where the right edge goes under the left between the two leftmost outgoing edges. In general, to have the right edge go under the left edge between two outgoing edges, first a row vertex stabilization move is done, followed by a commutation′ move between the columns containing the X's associated with the edges involved.

In Figure 28, we show an example of R$\overline{\text{V}}$ where the right edge goes over the left between the two leftmost outgoing edges. Let X_1 and X_2, from left to right, be the X's in the flock that are associated with the edges that will be interchanged next to the vertex. In general to have the right edge go over the left edge, a row vertex stabilization move is done, if needed. Next a row stabilization′ move is done on the row that contains the standard O that is in the same column as X_2. Call this O O_i. The column that is added in the stabilization′ is placed immediately to the right of the flock. Then a commutation′ move is done to move the row containing O_i below the row containing the standard O that is in the same column as X_1. Finally a commutation′ move is done between the columns containing X_1 and X_2. To do R$\overline{\text{V}}$ for the incoming edges, one needs only switch the role of the row and column.

Movement of a valence-two vertex The movement of a valence-two vertex is equivalent to moving an O^* with a single X in both its row and column to the position of a standard O that is on one of the incident edges. This could be thought of as choosing

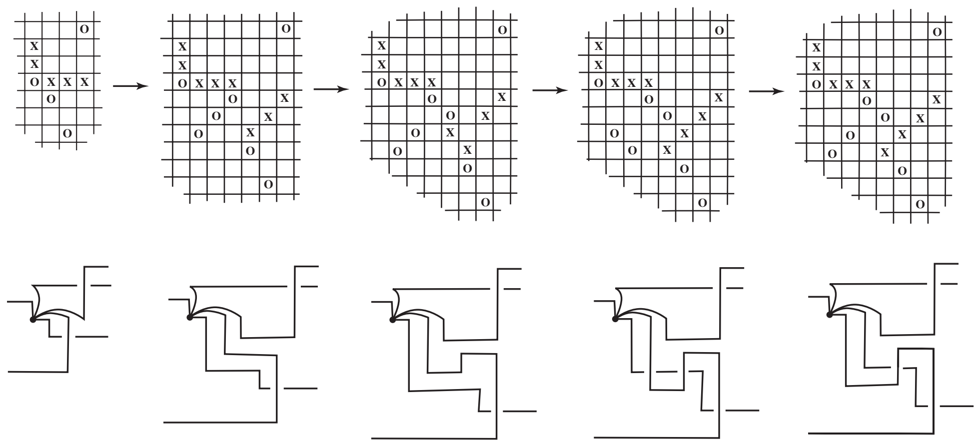

Figure 28: A row vertex stabilization move, followed by a row stabilization′ move, a row commutation′ move and a column commutation′ move in the grids, producing a $\mathrm{R}\overline{\mathrm{V}}$ move in the associated spatial graph

a different O on the edge to be special, and was first addressed for grid diagrams in [12, Lemma 2.12]. A proof of the independence of which O is special is given in [21, Lemma 4.1]. Since this is a local change, the same diagrammatic proof works in the graph case. We outline the proof here.

We will describe in words the moves needed to do this. However, the reader may just choose to look at the moves done in Figure 29. To move a valence-two vertex along an edge, we move the associated O^* to the position of a standard O on an adjacent edge. First a row stabilization′ is done at one of the neighboring X's, between the X and the O^*. The new column containing the new X and O are moved by commutation′ next to O^*, shown in the second image in Figure 29. Then the row containing O^* can be moved by commutation′ moves to the X in the column with the O^*. Then the column containing O^* can be moved by commutation′ to the O in the column with the X that is next to O^*. Now O^* is left and the O and X can be moved in their row by commutation′ to the X that is in the O's column. These X and O can then be removed by a column destabilization′.

Two vertices pass each other The planar isotopy where one vertex v passes another vertex w can be obtained via first vertex stabilization moves if needed, and then a number of commutation′ moves between the v–columns and the w–columns. In Figure 30, we show an example where only a single stabilization′ move is needed before the commutation′ moves, switching the order of the v–columns and the w–columns. To have the vertices move passed each other vertically rather than horizontally, the roles of the rows and columns are interchanged.

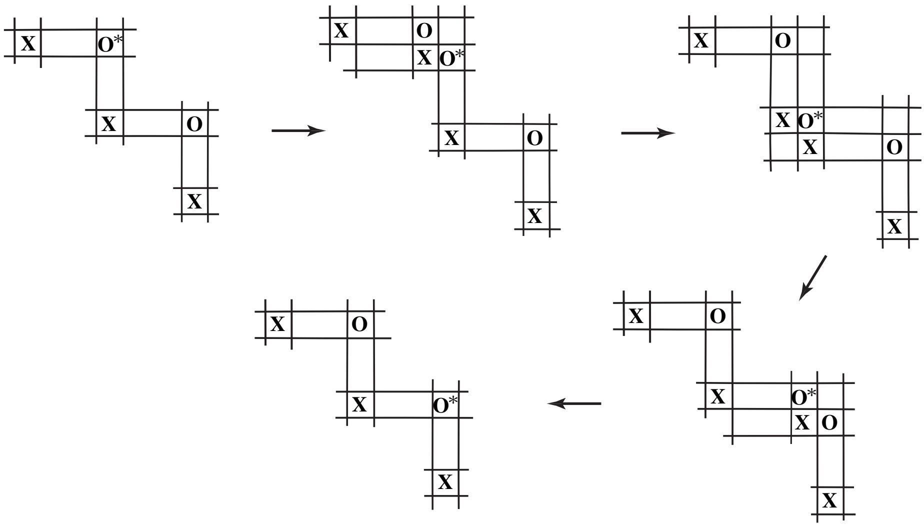

Figure 29: The graph grid moves needed to move a standard O^* to the position of an O

A vertex and arc pass each other The planar isotopy where an arc and a vertex move passed each other can be obtained via first vertex stabilization moves if needed and then commutation′ moves between a set of v–columns (resp. v–rows) and another column (resp. row) that is associated with an appropriate arc; see Figure 31. □

This shows that even though there are numerous different graph grid diagrams that will represent the same transverse spatial graph, all such grids are related by a sequence of the graph grid moves.

4 Graph Floer homology

In this section, we will define the main invariant of this paper, which we call the graph Floer homology of a spatial graph. This will take the form of the homology of a bigraded chain complex that is a module over a polynomial ring (or more generally, the quasi-isomorphism type of the chain complex). One of the gradings is the homological grading (also called the Maslov grading) and the other grading is called the Alexander grading, and will take values in the first homology of the exterior of the transverse spatial graph. Our definitions will generalize those given in [11] and [12] except that we only get a (relatively) bigraded object instead of a filtered object. In particular, when the spatial graph is a knot or link, we recover the associated graded objects from [12] (but with a relative Alexander grading). In this section and throughout the rest of the paper, we assume that the reader is familiar with the material of [12, Sections 1–3].

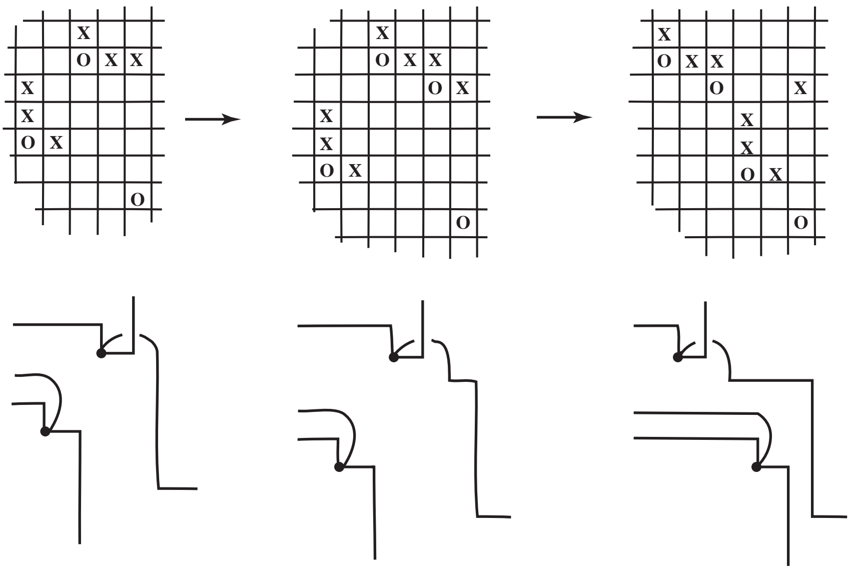

Figure 30: An example illustrating the two steps to construct a planar isotopy in which one vertex passes another vertex in the diagram of the associated transverse spatial graph. First a stabilization′ move was done then a number of the commutation′ moves were done.

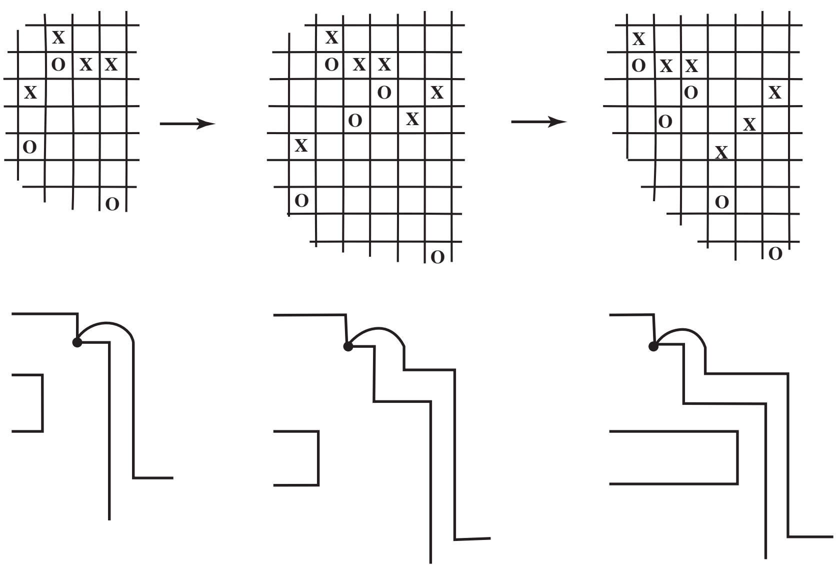

Figure 31: An example illustrating the two steps to construct a planar isotopy where a vertex passes an arc in the diagram of the associated transverse spatial graph. First a row vertex stabilization move was done then a number of the commutation′ moves were done.

4.1 Algebraic terminology

We start with some algebraic preliminaries. The reader can skip this section upon first reading and refer back to it as needed. Many of these definitions are similar to those of [12, Section 2.1].

Let C be a vector space over $\mathbb{F}$, where $\mathbb{F}$ is the field with 2 elements and let $\mathbb{G}$, $\mathbb{G}_1$ and $\mathbb{G}_2$ be abelian groups. Recall that a *$\mathbb{G}$–grading on C* (also called an *absolute $\mathbb{G}$–grading*) is a decomposition $C = \bigoplus_{g\in\mathbb{G}} C_g$, where $C_g \subseteq C$ is a vector subspace of C for each g. In this case we say that C is *graded over* $\mathbb{G}$ or is *graded*. A linear map $\phi\colon C \to C'$ between two graded vector spaces is *a graded map of degree h* if $\phi(C_g) \subseteq C'_{g+h}$ for all $g \in \mathbb{G}$. A *relative $\mathbb{G}$–grading on C* is a $\mathbb{G}$–grading that is well defined up to a shift in $\mathbb{G}$. That is, $C = \bigoplus_{g\in\mathbb{G}} C_g$ and $C = \bigoplus_{g\in\mathbb{G}} C'_g$ give the same relative $\mathbb{G}$–gradings if there exists an $a \in \mathbb{G}$ such that $C_g = C'_{g+a}$ for all $g \in \mathbb{G}$. Thus if $C = \bigoplus_{g\in\mathbb{G}} C_g$ has a well-defined relative grading and $x \in C_{g_1}$ and $y \in C_{g_2}$, then the difference between their gradings $g_1 - g_2$ is well-defined and independent of the choice of direct sum decomposition. In this case, we say that C is *relatively graded over* $\mathbb{G}$ or is *relatively graded.* A linear map $\phi\colon C \to C'$ between two relatively graded vector spaces is a *graded map* if there exists an $h \in \mathbb{G}$ such that $\phi(C_g) \subseteq C'_{g+h}$ for all $g \in \mathbb{G}$. Note that it does not make sense to talk about the degree of this map since we can shift the subgroups and get a different value for h. In this paper, we will be interested in relatively *bigraded* vector spaces over $H_1(E(f))$ and $\mathbb{Z}$, where $E(f)$ is the complement of a transverse spatial graph in S^3.

Definition 4.1 A $(\mathbb{G}_1, \mathbb{G}_2)$*–bigrading* on C is a $\mathbb{G}_1 \oplus \mathbb{G}_2$–grading on C. In this case we say that C is *bigraded* over $\mathbb{G}_1$ and $\mathbb{G}_2$, or is *graded.* We may also refer to a bigrading as an *absolute bigrading* when convenient. A linear map $\phi\colon C \to C'$ between two bigraded vector spaces is a *bigraded map of degree* (h_1, h_2) if $\phi(C_{(g_1,g_2)}) \subseteq C'_{(g_1+h_1,g_2+h_2)}$ for all $(g_1, g_2) \in \mathbb{G}_1 \oplus \mathbb{G}_2$. A *relative bigrading* on C over $\mathbb{G}_1$ and $\mathbb{G}_2$ is a relative $\mathbb{G}_1 \oplus \mathbb{G}_2$–grading on C. In this case we say that C is *relatively bigraded over* $\mathbb{G}_1$ *and* $\mathbb{G}_2$ or is *relatively graded.* A linear map $\phi\colon C \to C'$ between two relatively graded vector spaces is a *bigraded map* if there exists an $(h_1, h_2) \in \mathbb{G}_1 \oplus \mathbb{G}_2$ such that $\phi(C_{(g_1,g_2)}) \subseteq C'_{(g_1+h_1,g_2+h_2)}$ for all $(g_1, g_2) \in \mathbb{G}_1 \oplus \mathbb{G}_2$.

Note that a (relative) bigrading of C over $\mathbb{G}_1$ and $\mathbb{G}_2$ gives a well-defined (relative) grading over $\mathbb{G}_i$ for $i = 1, 2$ in the obvious way:

$$C = \bigoplus_{g_1\in\mathbb{G}_1} C_{(g_1,g_2)}\Bigl(\bigoplus_{g_2\in\mathbb{G}_2} C_{(g_1,g_2)}\Bigr) \quad \text{and} \quad C = \bigoplus_{g_2\in\mathbb{G}_2} C_{(g_1,g_2)}\Bigl(\bigoplus_{g_1\in\mathbb{G}_1} C_{(g_1,g_2)}\Bigr).$$

Our main invariant will turn out to be graded over two groups, one of which will be relatively graded and one of which will be (absolutely) graded. In the following definition, RA stands for relative absolute or relatively absolutely.

Definition 4.2 Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be abelian groups and C be a vector space. An *RA* $(\mathbb{G}_1, \mathbb{G}_2)$*–bigrading* on C is a $\mathbb{G}_1 \oplus \mathbb{G}_2$–grading that is well defined up to a shift in $\mathbb{G}_1 \oplus \{0\}$. That is, $C = \bigoplus_{g \in \mathbb{G}} C_g$ and $C = \bigoplus_{g \in \mathbb{G}} C'_g$ give the same RA $(\mathbb{G}_1, \mathbb{G}_2)$–bigradings if there exists some $(g_1, 0) \in \mathbb{G}$ such that $C_g = C'_{g+(g_1,0)}$ for all $g \in \mathbb{G}$. In this case, we say that C is *RA bigraded* over $\mathbb{G}_1$ and $\mathbb{G}_2$, or, simply, is *RA bigraded*. A linear map $\phi\colon C \to C'$ between two RA bigraded vector spaces is a *bigraded map of degree* $(*, h_2)$ if there exists an $h_1 \in \mathbb{G}_1$ such that $\phi(C_{(g_1,g_2)}) \subseteq C'_{(g_1+h_1, g_2+h_2)}$ for all $(g_1, g_2) \in \mathbb{G}_1 \oplus \mathbb{G}_2$. A linear map $\phi\colon C \to C'$ between two RA bigraded vector spaces is a *bigraded map* if it is a bigraded map of some degree.

Note that an RA $(\mathbb{G}_1, \mathbb{G}_2)$–bigrading of C gives a well-defined relative grading over $\mathbb{G}_1$ and a well-defined (absolute) grading over $\mathbb{G}_2$. We will also need to define bigraded chain complexes and their equivalences.

Definition 4.3 A $(\mathbb{G}, \mathbb{Z})$*–bigraded chain complex* is a $(\mathbb{G}, \mathbb{Z})$–bigraded vector space C and bigraded map $\partial\colon C \to C$ of degree $(0, -1)$ such that $\partial^2 = 0$. For $g \in \mathbb{G}$ and $i \in \mathbb{Z}$, a linear map $\phi\colon C \to C'$ between $\mathbb{G} \oplus \mathbb{Z}$–bigraded chain complexes is a *bigraded chain map of degree* (g, i) if it is a chain map, ie $\partial \circ \phi = \phi \circ \partial$, and it a bigraded map of degree (g, i). We say that ϕ is a *bigraded chain map* if it is a bigraded chain map of some degree.

Note that if $C = \bigoplus_{g \in \mathbb{G}} C_g$ has a relative $\mathbb{G}$–grading then it makes sense to talk about a *graded map* $\phi\colon C \to C$ *of degree* h as one that satisfies $\phi(C_g) \subset C_{g+h}$ for all $g \in \mathbb{G}$. For, suppose $a \in \mathbb{G}$ and $C = \bigoplus_{g \in \mathbb{G}} C'_g$ are such that for all $g \in \mathbb{G}$, $C_g = C'_{g+a}$. Then $\phi(C'_g) = \phi(C_{g-a}) \subset C_{g-a+h} = C'_{g+h}$ for all $g \in \mathbb{G}$. Similarly, we can define a bigraded map of degree (g_1, g_2) between a relatively (or RA) bigraded vector space and itself. We say a linear map (between absolutely, relatively or RA bigraded vector spaces) is a *bigraded* map if it is a bigraded map of some degree.

Definition 4.4 A *relative* (respectively *RA*) $(\mathbb{G}, \mathbb{Z})$*–bigraded chain complex* is a relative (respectively RA) $(\mathbb{G}, \mathbb{Z})$–bigraded vector space C and bigraded map $\partial\colon C \to C$ of degree $(0, -1)$ such that $\partial^2 = 0$. A linear map $\phi\colon C \to C'$ between relative (respectively RA) $\mathbb{G} \oplus \mathbb{Z}$–bigraded chain complexes is a *bigraded chain map* if it is a chain map, ie $\partial \circ \phi = \phi \circ \partial$, and it is a bigraded map.

Similarly, we can define $(\mathbb{G}, \mathbb{Z}_2)$*–bigraded*, *relatively bigraded*, and *RA bigraded chain complexes*. This will be used in the last section of the paper when we compare our invariant to sutured Floer homology.

The primary invariant will be a chain complex that also takes the form of a module over a multivariable polynomial ring. Recall that a $\mathbb{G}$*–graded ring* is a commutative ring R with a direct sum decomposition as abelian groups $R = \bigoplus_{g\in\mathbb{G}} R_g$, where R_g is a subgroup of R and $R_{g_1}R_{g_2} \subset R_{g_1+g_2}$ for all $g_i \in \mathbb{G}$. If R is a $\mathbb{G}$–graded ring, a $\mathbb{G}$*–graded (left)* R*–module* is a (left) R–module with a direct sum decomposition as abelian groups $M = \bigoplus_{g\in\mathbb{G}} M_g$, where M_g is a subgroup of M and $R_{g_1}M_{g_2} \subset M_{g_1+g_2}$ for all $g_i \in \mathbb{G}$. A *graded* R*–module homomorphism of degree* h is a graded map $\phi\colon C \to C'$ (of degree h) between $\mathbb{G}$–graded R–modules that is also an R–module homomorphism. We can similarly define *relatively graded, bigraded, relatively bigraded, and relatively absolutely bigraded* R*–modules* and graded module homomorphisms in these cases.

Definition 4.5 A $(\mathbb{G},\mathbb{Z})$*–bigraded (left)* R*–module chain complex* is a $(\mathbb{G},\mathbb{Z})$–bigraded (left) R–module C and bigraded R–module homomorphism $\partial\colon C \to C$ of degree $(0,-1)$ such that $\partial^2 = 0$. For $g \in \mathbb{G}$ and $i \in \mathbb{Z}$, an R–module homomorphism $\phi\colon C \to C'$ between $\mathbb{G} \oplus \mathbb{Z}$–bigraded R–module chain complexes is a *bigraded* R*–module chain map of degree* (g,i) if $\partial \circ \phi = \phi \circ \partial$ and it is a bigraded map of degree (g,i). We say that $\phi\colon C \to C'$ is a *bigraded* R*–module chain map* if it is a bigraded R–module map of some degree. We can similarly define a relative (respectively RA) $(\mathbb{G},\mathbb{Z})$–bigraded R–module chain complex and a relative bigraded R–module chain map.

We remark that if C is a $(\mathbb{G},\mathbb{Z})$–bigraded chain complex of (left) R–modules then it is not necessarily the case that any of $C_{(g,m)}$, $\bigoplus_{h\in\mathbb{G}} C_{(g,m)}$ or $\bigoplus_{m\in\mathbb{Z}} C_{(g,m)}$ is an R–module.

The primary invariant of this paper associates to each graph grid diagram a bigraded R–module chain complex. However, choosing different graph grid diagrams representing the same transverse spatial graph will lead to different chain complexes. We will show that they are all quasi-isomorphic.

Definition 4.6 A chain map $\phi\colon C \to C'$ of chain complexes is a *quasi-isomorphism* if it induces an isomorphism on homology. We say that two chain complexes C and D are *quasi-isomorphic* if there is a sequence of chain complexes $C_0, \dots, C_r$ and quasi-isomorphisms

$$\begin{tikzcd}
C_0 \arrow[dr, "\phi_1"] & & C_2 \arrow[dl, "\phi_2"'] \arrow[dr, "\phi_3"] & & \cdots & & C_{r-2} \arrow[dl, "\phi_{r-2}"'] \arrow[dr, "\phi_{r-1}"] & & C_r \arrow[dl, "\phi_r"'] \\
& C_1 & & C_3 & & C_{r-3} & & C_{r-1} &
\end{tikzcd}$$

such that $C_0 = C$, $C_r = D$. Suppose C and D are $(\mathbb{G}, \mathbb{Z})$–bigraded R–module chain complexes. We say that $\phi\colon C \to C'$ is a bigraded R–module quasi-isomorphism if ϕ is a quasi-isomorphism and a bigraded R–module homomorphism. We say that C and D are *quasi-isomorphic (as $(\mathbb{G}, \mathbb{Z})$–bigraded R–module chain complexes)* if they are quasi-isomorphic and all the quasi-isomorphisms are bigraded R–module chain maps. We can similarly define quasi-isomorphism for two relative (or RA) $(\mathbb{G}, \mathbb{Z})$–bigraded R–module chain complexes.

Remark 4.7 In the definition of quasi-isomorphism, it suffices to consider sequences of length $r = 2$. To see that these are equivalent, see [16, Proposition A.3.11].

4.2 The chain complex

For technical reasons, we need to restrict our definition to graphs that are both sinkless and sourceless. The graph grid diagrams representing these transverse spatial graphs have at least one X per column and row. We will need this condition to ensure that $\partial^2 = 0$.

Definition 4.8 A graph grid diagram is *saturated* if there is at least one X in each row and each column. A transverse spatial graph $f\colon G \to S^3$ is called *sinkless and sourceless* if its underlying graph G is sinkless and sourceless (ie has no vertices with only incoming edges or only outgoing edges).

Remark 4.9 (1) Suppose that g is a graph grid diagram representing the transverse spatial graph f. Then g is saturated if and only if f is sinkless and sourceless. (2) If one performs a graph grid move on a saturated graph grid diagram, then the resulting graph grid diagram is saturated.

Convention For the rest of this paper, unless otherwise mentioned, we will assume that all transverse spatial graphs are sinkless and sourceless and all graph grid diagrams are saturated.

Let $f\colon G \to S^3$ be a sinkless and sourceless transverse spatial graph, define $E(f) = S^3 \setminus N(f(G))$, where $N(f(G))$ is a regular neighborhood of $f(G)$ in S^3, let g be an $n \times n$ saturated graph grid diagram representing f, and $\mathcal{T}$ be its corresponding toroidal diagram. Now, let

$$\boldsymbol{S} = \{\{x_i\}_{i=1}^n \mid x_i = \alpha_i \cap \beta_{\sigma(i)}, \sigma \in S_n\},$$

where S_n is the symmetric group on n elements, and define $C^-(g)$ to be the free (left) R_n–module generated by $\boldsymbol{S}$, where

$$R_n = \mathbb{F}[U_1, \dots, U_n]$$

and $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$ denotes the field with two elements. When working with generators on a planar grid diagram, we use the convention that places the intersection point on the bottom leftmost grid line (as opposed to the top rightmost grid line). When we want to specify the grid, we may write $\boldsymbol{S}(g)$ instead of $\boldsymbol{S}$.

Using the toroidal grid diagram, we can view the torus $\mathcal{T}$ as a two-dimensional CW complex with n^2 0–cells (intersections of α_i and β_j), $2n^2$ 1–cells (consisting of line segments on α_i and β_j), and n^2 2–cells (squares cut out by α_i and β_j). Note that a generator $\boldsymbol{x} \in \boldsymbol{S}$ can be viewed as a 0–chain. Let $U_{\boldsymbol{\alpha}}$ be the 1–dimensional subcomplex of $\mathcal{T}$ consisting of the union of the horizontal circles. We define paths, domains, and rectangles in the same way as [12]. Given two generators $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\boldsymbol{S}$, a *path* from $\boldsymbol{x}$ to $\boldsymbol{y}$ is a 1–cycle γ such that the boundary of the 1–chain obtained by intersecting γ with $U_{\boldsymbol{\alpha}}$ is $\boldsymbol{y} - \boldsymbol{x}$. A *domain* D from $\boldsymbol{x}$ to $\boldsymbol{y}$ is an 2–chain in $\mathcal{T}$ whose boundary ∂D is a path from $\boldsymbol{x}$ to $\boldsymbol{y}$. The support of D is the union of the closures of the 2–cells appearing in D (with nonzero multiplicity). We denote the set of domains from $\boldsymbol{x}$ to $\boldsymbol{y}$ by $\pi(\boldsymbol{x},\boldsymbol{y})$ and note that there is a composition of domains $*\colon \pi(\boldsymbol{x},\boldsymbol{y}) \times \pi(\boldsymbol{y},\boldsymbol{z}) \to \pi(\boldsymbol{x},\boldsymbol{z})$. A domain from $\boldsymbol{x}$ to $\boldsymbol{y}$ that is an embedded rectangle r is called a *rectangle* that connects $\boldsymbol{x}$ to $\boldsymbol{y}$. Let $\mathrm{Rect}(\boldsymbol{x},\boldsymbol{y})$ be the set of rectangles that connect $\boldsymbol{x}$ to $\boldsymbol{y}$. Notice if $\boldsymbol{x}$ and $\boldsymbol{y}$ agree in all but two intersection points then there are exactly two rectangles in $\mathrm{Rect}(\boldsymbol{x},\boldsymbol{y})$, otherwise $\mathrm{Rect}(\boldsymbol{x},\boldsymbol{y}) = \varnothing$. A rectangle $r \in \mathrm{Rect}(\boldsymbol{x},\boldsymbol{y})$ is *empty* if $\mathrm{Int}(r) \cap \boldsymbol{x} = \varnothing$ where $\mathrm{Int}(r)$ is the interior of the rectangle in $\mathcal{T}$. Let $\mathrm{Rect}^o(\boldsymbol{x},\boldsymbol{y})$ be the set of empty rectangles that connect $\boldsymbol{x}$ to $\boldsymbol{y}$.

We now make $C^-(g)$ into a chain complex $(C^-(g), \partial^-)$ in the usual way: by counting empty rectangles. Note that in [11; 12], the authors consider rectangles that contain both X's and O's. However, because there is no natural filtration of $H_1(E(f))$, we must restrict to rectangles without any X's and thus we get a graded object instead of a filtered object. Let $\mathbb{X}$ and $\mathbb{O}$ be the set of X's and O's in the grid. Put an ordering on each of these sets, $\mathbb{O} = \{\mathrm{O}_i\}_{i=1}^n$ and $\mathbb{X} = \{\mathrm{X}_i\}_{i=1}^m$ so that $\mathrm{O}_1, \dots, \mathrm{O}_V$ are associated to the V vertices of the graph. For a domain $D \in \pi(\boldsymbol{x},\boldsymbol{y})$, let $\mathrm{O}_i(D)$ (respectively $\mathrm{X}_i(D)$) denote the multiplicity with which O_i (respectively X_i) appear in D. More precisely, D is a domain so $D = \sum a_j r_j$, where r_j is a rectangle in $\mathcal{T}$. Thus $\mathrm{O}_i(D) = \sum a_j \mathrm{O}_i(r_j)$, where $\mathrm{O}_i(r_j)$ is 1 if $\mathrm{O}_i \in r_j$ and 0 otherwise (similarly for $\mathrm{X}_i(D)$). We note that if r is a rectangle then $\mathrm{O}_i(r) \geq 0$. Define $\partial^-\colon C^-(g) \to C^-(g)$ as follows. For $\boldsymbol{x} \in \boldsymbol{S}$, let

$$\partial^-(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \boldsymbol{S}} \sum_{\substack{r \in \mathrm{Rect}^o(\boldsymbol{x},\boldsymbol{y}) \\ \mathrm{Int}(r) \cap \mathbb{X} = \varnothing}} U_1^{\mathrm{O}_1(r)} \cdots U_n^{\mathrm{O}_n(r)} \cdot \boldsymbol{y}.$$

Extend ∂^- to all of $C^-(g)$ so that it is an R_n–module homomorphism. When we want to specify the grid, we will write ∂_g^- instead of ∂^-.

Proposition 4.10 *If g is a saturated graph grid diagram then $\partial_g^- \circ \partial_g^- = 0$.*

Proof This proof follows that of [12, Proposition 2.10, page 2349] almost verbatim. The only change is that we only consider regions that do not contain any X's. Briefly, let $\boldsymbol{x} \in \boldsymbol{S}$. Then

$$\partial^- \circ \partial^-(\boldsymbol{x}) = \sum_{\boldsymbol{z} \in \boldsymbol{S}} \sum_{D \in \pi(\boldsymbol{x},\boldsymbol{z})} N(D) \cdot U_1^{O_1(D)} \cdots U_n^{O_n(D)} \cdot \boldsymbol{z},$$

where $N(D)$ is the number of ways one can decompose D as $D = r_1 * r_2$, where $r_1 \in \mathrm{Rect}^o(\boldsymbol{x},\boldsymbol{y})$, $r_2 \in \mathrm{Rect}^o(\boldsymbol{y},\boldsymbol{z})$ and $\mathrm{Int}(r_i) \cap \mathbb{X} = \varnothing$. When $\boldsymbol{x} \neq \boldsymbol{z}$ there are three general cases. The rectangles are either disjoint, overlapping or share a common edge. In each of these cases there are two ways that the region can be decomposed as empty rectangles. Thus the resulting $\boldsymbol{z}$ occurs in the sum twice. The case where $\boldsymbol{x} = \boldsymbol{z}$ is the result of domains $D \in \pi(\boldsymbol{x},\boldsymbol{x})$, which are width-one annuli. Such domains do not occur in the image of $\partial^- \circ \partial^-(\boldsymbol{x})$ because ∂^- only counts rectangles that do not contain X's, and we have assumed that our graph grid diagram is saturated. Thus we see that $\partial^- \circ \partial^-(\boldsymbol{x})$ vanishes. □

4.3 Gradings

We put two gradings on the $(C^-(g), \partial^-)$. The first is the homological grading, also called the Maslov grading. This will be defined exactly the same as in [12]. We quickly review the definition for completeness.

Given two finite sets of points A and B in the plane and a point $q = (q_1, q_2)$ in the plane, define $\mathcal{I}(q, B) = \#\{(b_1, b_2) \in B \mid b_1 > q_1, b_2 > q_2\}$. That is, $\mathcal{I}(q, B)$ is the number of points in B above and to the right of q. Let $\mathcal{I}(A, B) = \sum_{q \in A} \mathcal{I}(q, B)$ and $\mathcal{J}(A, B) = (\mathcal{I}(A, B) + \mathcal{I}(B, A))/2$. It will be useful to note that an equivalent definition of $\mathcal{I}(A, q)$ is the number of points in the set $\{a \in A \mid q_1 > a_1, q_2 > a_2\}$, ie the number of points in A below and to the left q. So $\mathcal{J}(q, A)$ counts with weight one half all the points in A above and to the right of q and down and to the left of q. Slightly abusing notation, we view $\mathbb{O}$ as a set of points in the grid with half-integer coordinates (the points where the O's occur). Similarly, we view $\mathbb{X}$ as a set of points in the grid with half-integer coordinates. We extend $\mathcal{J}$ bilinearly over formal sums and differences of subsets in the plane and define the Maslov grading of $\boldsymbol{x} \in \boldsymbol{S}$ to be

$$M(\boldsymbol{x}) = \mathcal{J}(\boldsymbol{x} - \mathbb{O}, \boldsymbol{x} - \mathbb{O}) + 1.$$

This is consistent with the definition given in [12], and only depends on the set $\mathbb{O}$. Since, like in [12], we have exactly one O per column and row, [12, Lemmas 2.4 and 2.5] also hold for grid diagrams of transverse spatial graphs. Thus, it follows that M is a well-defined function on the toroidal grid diagram [12, Lemma 2.4]. In addition,

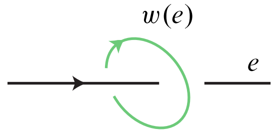

Figure 32: The weight $w(e)$ assigned to an edge $e \in E(G)$

for $x, y \in \boldsymbol{S}$, we can define the relative Maslov grading by $M(\boldsymbol{x}, \boldsymbol{y}) := M(\boldsymbol{x}) - M(\boldsymbol{y})$. By [12, Lemma 2.5],

$$M(\boldsymbol{x}, \boldsymbol{y}) = M(\boldsymbol{x}) - M(\boldsymbol{y}) = 1 - 2n_{\mathbb{O}}(r), \tag{1}$$

for any *empty*[1] rectangle r connecting $\boldsymbol{x}$ to $\boldsymbol{y}$, where $n_{\mathbb{O}}(r)$ is the number of O's in r.

Before defining the Alexander grading we first need to establish a weight system on the edges of G. Let $E(G)$ be the set of edges of the graph G. We define $w\colon E(G) \to H_1(E(f))$ by sending each edge to the meridian of the edge, with orientation given by the right-hand rule, seen as an element of $H_1(E(f))$; see Figure 32. We say that $w(e)$ is the weight of the edge e. We will denote the weight of X by $w(\mathrm{X})$ and the weight of O by $w(\mathrm{O})$, and define them based on the weight of the associated edge. If X or O appear on the interior of an edge $e \in E(G)$ in the associated transverse spatial graph, then $w(\mathrm{X}) := w(e)$ and $w(\mathrm{O}) := w(e)$. If the O is associated to the vertex v in the associated transverse spatial graph, then

$$w(\mathrm{O}) := \sum_{e \in \mathrm{In}(v)} w(e) = \sum_{e \in \mathrm{Out}(v)} w(e),$$

where $\mathrm{In}(v)$ and $\mathrm{Out}(v)$ are the sets of incoming and outgoing edges of v, respectively.

Remark 4.11 Recall that an edge of a graph G is called a *cut edge* if the number of connected components of $G \setminus e$ is greater than the number of connected components of G. Observe that $w(e) = 0$ if and only if e is a cut edge. This will be useful in the proof of Theorem 6.6.

Let the function $\epsilon\colon \mathbb{O} \cup \mathbb{X} \to \{1, -1\}$ be defined by

$$\epsilon(p) = \begin{cases} 1 & \text{if } p \in \mathbb{X}, \\ -1 & \text{if } p \in \mathbb{O}. \end{cases}$$

For a point q in the grid, define

$$A^g(q) = \sum_{p \in \mathbb{O} \cup \mathbb{X}} \mathcal{J}(q, p) w(p) \epsilon(p).$$

[1] If the rectangle r contains m points of $\boldsymbol{x}$ in its interior, then $M(\boldsymbol{x}) = M(\boldsymbol{y}) + 1 + 2(m - n_{\mathbb{O}(r)})$.

We define the *Alexander grading* of $\boldsymbol{x}$ with respect to the grid g to be

$$A^g(\boldsymbol{x}) = \sum_{p \in \mathbb{O} \cup \mathbb{X}} \mathcal{J}(\boldsymbol{x}, p) w(p) \epsilon(p) = \sum_{x_i \in \boldsymbol{x}} A^g(x_i).$$

This value a priori lives in $\frac{1}{2} H_1(E(f))$; however, by Lemma 4.16, $A^g(\boldsymbol{x}) \in H_1(E(f))$. Note that this definition depends on the choice of *planar* grid g, and is not a well-defined function on the toroidal grid.[2] However, the relative Alexander grading is a well-defined function on the toroidal grid diagram.

Definition 4.12 For $\boldsymbol{x}, \boldsymbol{y} \in S$, let $A^{\mathrm{rel}}(\boldsymbol{x}, \boldsymbol{y}) := A^g(\boldsymbol{x}) - A^g(\boldsymbol{y})$ be the *relative Alexander grading* of $\boldsymbol{x}$ and $\boldsymbol{y}$.

When it is clear, we may drop the "rel" or "g" on A. By the following lemma, A^{rel} does not depend on how you cut open the toroidal diagram $\mathcal{T}$ to give a planar grid diagram g. We first need to define some notation. For a rectangle r, we define $w_{\mathbb{O}}(r) = \sum_{q \in \mathbb{O} \cap r} w(q)$ and, similarly, $w_{\mathbb{X}}(r) = \sum_{q \in \mathbb{X} \cap r} w(q)$. If D is a domain then $D = \sum a_i D_i$, where D_i is a rectangle. We extend $w_{\mathbb{O}}$ and $w_{\mathbb{X}}$ linearly to domains, so that $w_{\mathbb{O}}(D) = \sum a_i w_{\mathbb{O}}(D_i)$ and $w_{\mathbb{X}}(D) = \sum a_i w_{\mathbb{X}}(D_i)$.

Note that a path from $\boldsymbol{x}$ to $\boldsymbol{y}$, on the toroidal grid, gives a 1–cycle in $E(f)$. To see this, recall that the transverse spatial graph associated to the graph grid diagram is constructed from vertical arcs going from an X to an O outside the torus (above the plane) and horizontal arcs going from O to X inside the torus (below the plane). Thus, the intersection of the transverse spatial graph and the torus is $\mathbb{X} \cup \mathbb{O}$. Since a path is a 1–cycle on the torus missing $\mathbb{X} \cup \mathbb{O}$, we get an element of $H_1(E(f))$. Since the α_i and β_i bound disks in $E(f)$, this is a well-defined element of $H_1(E(f))$, independent of the choice of path.

Lemma 4.13 *Let $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{S}$. If $D \in \pi(\boldsymbol{x}, \boldsymbol{y})$ is a domain connecting $\boldsymbol{x}$ to $\boldsymbol{y}$ then*

$$A^g(\boldsymbol{x}) - A^g(\boldsymbol{y}) = w_{\mathbb{X}}(D) - w_{\mathbb{O}}(D). \tag{2}$$

If γ is a path connecting $\boldsymbol{x}$ to $\boldsymbol{y}$ then

$$A^g(\boldsymbol{x}) - A^g(\boldsymbol{y}) = [\gamma], \tag{3}$$

where $[\gamma] \in H_1(E(f))$ is the homology class of γ.

We note that the domain (or rectangle) in this lemma does not have to be empty.

[2] One can slightly change this definition to make it well-defined on the toroidal grid diagram. Our invariant will still only be relatively graded in the end however.

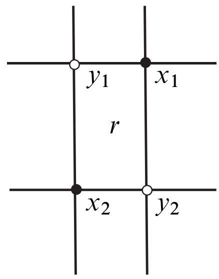

Figure 33: The rectangle r with the intersections labeled

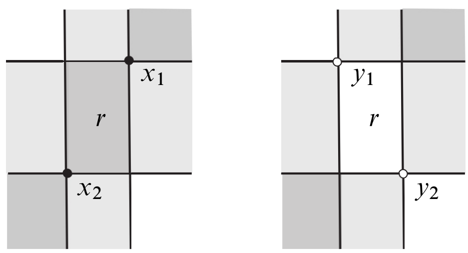

Figure 34: The lightly shaded regions are those that will be counted with weight one half in $\mathcal{J}(x_1,-)$ and $\mathcal{J}(x_2,-)$ (left), or in $\mathcal{J}(y_1,-)$ and $\mathcal{J}(y_2,-)$ (right).

Proof Let r be a rectangle connecting $\boldsymbol{x}$ to $\boldsymbol{y}$. We will first show that (2) holds for r. Consider

$$A^g(\boldsymbol{x}) - A^g(\boldsymbol{y}) = \sum_{q\in\mathbb{O}\cup\mathbb{X}} \mathcal{J}(\boldsymbol{x},q)w(q)\epsilon(q) - \sum_{q\in\mathbb{O}\cup\mathbb{X}} \mathcal{J}(\boldsymbol{y},q)w(q)\epsilon(q).$$

Let x_1, x_2, y_1, and y_2 be the intersection points at the corners of r, as shown in Figure 33. Since the intersection points of $\boldsymbol{x}$ and $\boldsymbol{y}$ only differ at the corners of the rectangle r this difference reduces to

$$\sum_{q\in\mathbb{O}\cup\mathbb{X}} \big[\mathcal{J}(x_1,q)+\mathcal{J}(x_2,q)-\mathcal{J}(y_1,q)-\mathcal{J}(y_2,q)\big]w(q)\epsilon(q).$$

By definition, $\mathcal{J}(x_1,\mathbb{O}\cup\mathbb{X})$ counts with weight one half all those X's and O's above and to the right of x_1 and below and to the left of x_1. In Figure 34 (left), we show which regions will have points counted in $\mathcal{J}(x_1,-)$ and $\mathcal{J}(x_2,-)$. The shading indicates if it will be counted with a weight of a half or one, this depends on whether it is counted in one or both of $\mathcal{J}(x_1,-)$ and $\mathcal{J}(x_2,-)$. Similarly, in Figure 34 (right) we show which regions will have points counted in $\mathcal{J}(y_1,-)$ and $\mathcal{J}(y_2,-)$. These counts differ by the points in r counted with weight one.

So we see that this difference is exactly

$$\sum_{q\in r\cap[\mathbb{O}\cup\mathbb{X}]} w(q)\epsilon(q) = \sum_{q\in\mathbb{X}\cap r} w(q) - \sum_{q\in\mathbb{O}\cap r} w(q).$$

Thus, (2) holds for rectangles.

Now we show that (3) holds. Let γ be a path connecting $\boldsymbol{x}$ to $\boldsymbol{y}$. Using the fact that S_n is generated by transposition, it follows that $\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{S}$ are related by a finite sequence of rectangles in $\mathcal{T}$. That is, there is a sequence $\boldsymbol{x}=\boldsymbol{x}_1,\dots,\boldsymbol{x}_l=\boldsymbol{y}$ of points in S and rectangles r_j connecting $\boldsymbol{x}_j$ to $\boldsymbol{x}_{j+1}$ for $1\le j\le l-1$. Thus,

$$\begin{aligned} A^g(\boldsymbol{x})-A^g(\boldsymbol{y}) &= \sum_j (A^g(\boldsymbol{x}_j)-A^g(\boldsymbol{x}_{j+1})) \\ &= \sum_j (w_{\mathbb{X}}(r_j)-w_{\mathbb{O}}(r_j)) = \sum_j [\partial r_j] = \Big[\sum_j \partial r_j\Big]. \end{aligned}$$

Since $\sum_j \partial r_j$ is also a path connecting $\boldsymbol{x}$ to $\boldsymbol{y}$, $\big[\sum_j \partial r_j\big]=[\gamma]$. Thus, we have proved (3).

Finally, we prove (2) for a general domain. Suppose D is a domain connecting $\boldsymbol{x}$ to $\boldsymbol{y}$. Then $D=\sum a_i D_i$ for some rectangles D_i, and ∂D is a path connecting $\boldsymbol{x}$ to $\boldsymbol{y}$. Thus $A^g(\boldsymbol{x})-A^g(\boldsymbol{y})=[\partial D]=\sum_i a_i[\partial D_i]=\sum_i a_i(w_{\mathbb{X}}(D_i)-w_{\mathbb{O}}(D_i))=w_{\mathbb{X}}(D)-w_{\mathbb{O}}(D)$. □

Corollary 4.14 *The relative grading $A^{\mathrm{rel}}\colon \boldsymbol{S}\times\boldsymbol{S}\to H_1(E(f))$ is a well-defined function on the toroidal graph grid diagram.*

Proof Any two $\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{S}$ are related by a sequence of rectangles and hence there is always a path γ connecting $\boldsymbol{x}$ to $\boldsymbol{y}$. Since the homology class of the path is independent of the choice of path and $A^{\mathrm{rel}}(\boldsymbol{x},\boldsymbol{y})=[\gamma]$, we see that A^{rel} is independent of how you cut open the toroidal graph grid diagram to get a planar graph grid diagram. □

We now provide an easy way to compute A^g for a planar graph grid diagram g. Let $\mathcal{L}$ be the lattice points in the grid, that is, the n^2 intersections between the horizontal and vertical grid lines (ie the set of points that on the torus become $\alpha_i\cap\beta_j$). Define $h\colon\mathcal{L}\to H_1(E(f))$, the *generalized winding number*, of a point $q\in\mathcal{L}$ as follows. Place the planar graph grid diagram g on the Euclidean plane with the lower left corner at the origin (and the upper right corner at the point (n,n)). Now consider the following projection of the associated transverse spatial graph $\mathrm{pr}(f)$. Like in the last section, this is obtained by connecting the X's to O's by arcs in the columns and O's to the X's in the rows. However, we now require that the arcs do not leave the $n\times n$ planar

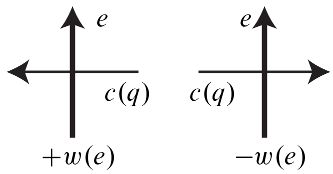

Figure 35: $\sigma(b_i) = \pm w(e)$, where the sign is given by the sign of $e \cdot c(q)$

grid (they cannot go around the torus). We also project this to the plane and ignore crossings. For $q \in \mathcal{L}$, let c be any path along the horizontal and vertical grid lines starting at the origin and ending at q. We also require that c meets $\mathrm{pr}(g)$ transversely. Then c intersects $\mathrm{pr}(g)$ in a finite number of points $b_1, \dots, b_k$, where b_i lives on the interior of some edge of the spatial graph. Suppose b_i lies on edge e. Define $\sigma(b_i)$ to be $\pm w(e)$, where the sign is given by the sign of the intersection of e with $c(q)$, with the usual orientation of the plane; see Figure 35. Using this, we set

$$h(q) = \sum_{i=1}^{k} \sigma(b_i).$$

When it is useful, we may also write h^g to specify that we are computing h in the graph grid diagram g.

Lemma 4.15 *The map h is well defined.*

Proof Consider the transverse spatial graph associated to g whose projection is $\mathrm{pr}(g)$ but which is pushed slightly above the plane. Fix a base point at infinity in S^3. It is easy to see that $h(q)$ is the homology of a loop made up of the path from infinity to the origin, then a path in the plane, from the origin to q, and finally the path from q to the point at infinity. This is independent of the choice of path from the origin to q. Therefore h is well defined. $\square$

Lemma 4.16 *For any point q on the lattice*

$$A^g(q) = -h(q). \tag{4}$$

Proof Let g be a graph grid diagram. First we see that $h(q) = 0$ for any point on the boundary of g by definition. Notice that we have

$$\sum_{p \in \mathrm{col}_i} w(p)\epsilon(p) = 0$$

for each i, where col_i is the set of O's and X's in the i^{th} column. Similarly,

$$\sum_{p \in \mathrm{row}_i} w(p)\epsilon(p) = 0$$

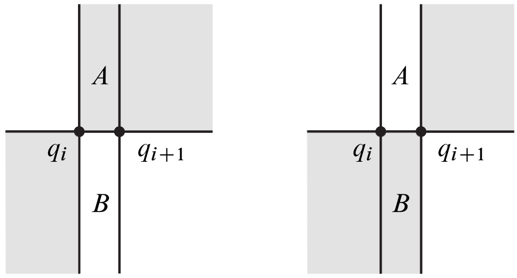

Figure 36: Here we shade the regions that will be counted with weight one half in $\mathcal{J}(q_i, -)$ (left), or $\mathcal{J}(q_{i+1}, -)$ (right).

for each i, where row_i is the set of O's and X's in the i^{th} row. Given these observations, it is immediate that $A^g(q) = \sum_{p\in\mathbb{O}\cup\mathbb{X}} \mathcal{J}(q, p)w(p)\epsilon(p)$ is also zero for any point on the boundary of g.

We will proceed by induction on the vertical grid line on which our lattice point occurs. Suppose the equality in (4) holds for q_i a lattice point on the i^{th} vertical arc, and consider q_{i+1} the point immediately to the right of q_i on the grid. Let

$$\mathrm{RHS} := \sum_{p\in\mathbb{O}\cup\mathbb{X}} \mathcal{J}(q_{i+1}, p)w(p)\epsilon(p) - \sum_{p\in\mathbb{O}\cup\mathbb{X}} \mathcal{J}(q_i, p)w(p)\epsilon(p),$$

and

$$\mathrm{LHS} := -h(q_{i+1}) - (-h(q_i)).$$

Figure 36 shows the regions in which the points of $\mathbb{O}\cup\mathbb{X}$ will be counted with weight one half in $\mathcal{J}(q_i, \mathbb{O}\cup\mathbb{X})$ and $\mathcal{J}(q_{i+1}, \mathbb{O}\cup\mathbb{X})$. These differ only in what is counted in the i^{th} column. So we see that

$$\mathrm{RHS} = \frac{1}{2}\Bigg[\sum_{p\in B\cap(\mathbb{O}\cup\mathbb{X})} w(p)\epsilon(p) - \sum_{p\in A\cap(\mathbb{O}\cup\mathbb{X})} w(p)\epsilon(p)\Bigg].$$

Using the fact $\sum_{p\in\mathrm{col}_i} w(p)\epsilon(p) = 0$ again, we can simplify this to,

$$\mathrm{RHS} = -\sum_{p\in A\cap(\mathbb{O}\cup\mathbb{X})} w(p)\epsilon(p) = \sum_{p\in B\cap(\mathbb{O}\cup\mathbb{X})} w(p)\epsilon(p).$$

We now consider $\mathrm{LHS} = -h(q_i) - [-h(q_{i+1})] = h(q_{i+1}) - h(q_i)$. Suppose the O in the i^{th} column is in A. Then all the vertical arcs of $\mathrm{pr}(g)$ in the i^{th} column that intersect the arc from q_i to q_{i+1} are oriented upwards. Thus $\mathrm{LHS} = \sum_{p\in\mathbb{X}\cap B} w(p) = \sum_{p\in B\cap(\mathbb{O}\cup\mathbb{X})} w(p)\epsilon(p)$. If the O in the i^{th} column is in B, the all the vertical arcs of $\mathrm{pr}(g)$ in the i^{th} column that intersect the arc from q_i to q_{i+1} are oriented downwards. So $\mathrm{LHS} = \sum_{p\in\mathbb{X}\cap A} -w(p) = -\sum_{p\in A\cap(\mathbb{O}\cup\mathbb{X})} w(p)\epsilon(p)$. □

Corollary 4.17 *For all $\boldsymbol{x} \in \boldsymbol{S}(g)$,*

$$A^g(\boldsymbol{x}) = -\sum_{x_i \in \boldsymbol{x}} h(x_i) \in H_1(E(f)).$$

For a given saturated graph grid diagram g, the functions A^g and M make $C^-(g)$ into a well-defined $(H_1(E(f)), \mathbb{Z})$–bigraded R_n–module chain complex once we say how the grading changes when we multiply a generator by U_i. We set

$$A^g(U_i) = -w(\mathrm{O}_i), \quad M(U_i) = -2 \tag{5}$$

and define

$$A^g(U_1^{a_1} \cdots U_n^{a_n} \boldsymbol{x}) = A^g(\boldsymbol{x}) + \sum_{i=1}^{n} a_i A^g(U_i)$$

and

$$M(U_1^{a_1} \cdots U_n^{a_n} \boldsymbol{x}) = M(\boldsymbol{x}) + \sum_{i=1}^{n} a_i M(U_i).$$

We remark that since g is saturated, $A^g(U_i) \neq 0$.

For each $a \in H_1(E(f))$ and $m \in \mathbb{Z}$, let $C^-(g)_{(a,m)}$ be the (vector) subspace of $C^-(g)$ with basis $\{U_1^{a_1} \cdots U_n^{a_n} \boldsymbol{x} \mid A^g(U_1^{a_1} \cdots U_n^{a_n} \boldsymbol{x}) = a,\ M(U_1^{a_1} \cdots U_n^{a_n} \boldsymbol{x}) = m\}$. This gives a bigrading on $C^-(g) = \sum_{(a,m)} C^-(g)_{(a,m)}$.

Proposition 4.18 *The differential ∂^- drops the Maslov grading by one and respects the Alexander grading. That is, $\partial^-\colon C^-(g)_{(a,m)} \to C^-(g)_{(a,m-1)}$ for all $a \in H_1(E(f))$ and $m \in \mathbb{Z}$.*

Proof Consider $U_1^{\mathrm{O}_1(r)} \cdots U_m^{\mathrm{O}_m(r)} \cdot \boldsymbol{y}$ appearing in the boundary of $\boldsymbol{x}$. Since $\boldsymbol{x}$ and $\boldsymbol{y}$ are connected by an empty rectangle r, we see that $M(\boldsymbol{x}) = M(\boldsymbol{y}) + 1 - 2n_{\mathbb{O}(r)}$ by (1). Since each U_i drops the Maslov grading by two, $M(\boldsymbol{x}) = M(U_1^{\mathrm{O}_1(r)} \cdots U_m^{\mathrm{O}_m(r)} \cdot \boldsymbol{y}) + 1$. Thus the differential ∂^- drops the Maslov grading by one.

Next $A^g(\boldsymbol{x}) = A^g(\boldsymbol{y}) + \sum_{\mathbb{X} \cap r} w(\mathrm{X}) - \sum_{\mathbb{O} \cap r} w(\mathrm{O})$, but by definition of the differential $\mathbb{X} \cap r = \varnothing$. So $A^g(\boldsymbol{x}) = A^g(\boldsymbol{y}) - \sum_{\mathbb{O} \cap r} w(\mathrm{O}) = A^g(U_1^{\mathrm{O}_1(r)} \cdots U_m^{\mathrm{O}_m(r)} \cdot \boldsymbol{y})$. Thus the differential ∂^- respects the Alexander grading. □

We are interested in viewing $(C^-(g), \partial^-)$ as a module instead of just a vector space. Using the definition in (5), $\mathbb{F}[U_1, \ldots, U_n]$ becomes an $(H_1(E(f)), \mathbb{Z})$–bigraded ring (ie $H_1(E(f)) \oplus \mathbb{Z}$–graded), and with this grading, $(C^-(g), \partial^-)$ is a $(H_1(E(f)), \mathbb{Z})$–bigraded R_n–module chain complex. We would like to define an invariant of the graph grid diagram that is unchanged under any graph grid moves, giving an invariant of the

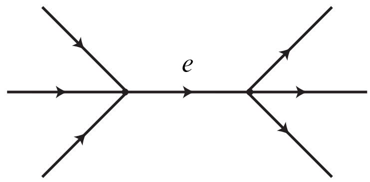

Figure 37: A singular edge e

transverse spatial graph. Since $\mathbb{F}[U_1,\dots,U_n]$ depends on the size of the grid, we need to view $C^-(g)$ as a module over a smaller ring. One choice would be to view $C^-(g)$ as a module over $\mathbb{F}[U_1,\dots,U_{V+E}]$, where $U_1,\dots,U_V$ correspond to the vertices of the graph and $U_{V+1},\dots U_{V+E}$ each correspond to a choice of O on a distinct edge of the graph. However, by Proposition 4.21, multiplication by a U_i corresponding to an edge is chain homotopic either to 0 or to multiplication by a U_j corresponding to a vertex, where $1 \le j \le V$. Thus it makes sense to view $C^-(g)$ as a module over $\mathbb{F}[U_1,\dots,U_V]$ (recall that we ordered $\mathbb{O}$ such that $\mathrm{O}_1,\dots,\mathrm{O}_V$ are vertex O's).

Let $I\colon \mathbb{F}[U_1,\dots,U_V] \to \mathbb{F}[U_1,\dots,U_n]$ be the natural inclusion of rings defined by setting $I(U_i)=U_i$ for $1\le i\le V$. Using I, any module over $\mathbb{F}[U_1,\dots,U_n]$ naturally becomes an $\mathbb{F}[U_1,\dots,U_V]$–module. Thus, we will view $C^-(g)$ as an R_V–module, where $R_V=\mathbb{F}[U_1,\dots,U_V]$. Note that ∂^- preserves the homology, hence the homology of $(C^-(g),\partial^-)$ inherits the structure of a $(H_1(E(f)),\mathbb{Z})$–bigraded R_V–module.

Definition 4.19 Let g be a saturated graph grid diagram representing the transverse spatial graph $f\colon G\to S^3$. The *graph Floer chain complex* of g is the RA $(H_1(E(f)),\mathbb{Z})$–bigraded R_V–module chain complex $(C^-(g),\partial^-)$. The *graph Floer homology* of g, denoted $\mathrm{HFG}^-(g)$, is the homology of $(C^-(g),\partial^-)$ viewed as an RA $(H_1(E(f)),\mathbb{Z})$–bigraded R_V–module.

Before stating Proposition 4.21, we need some terminology.

Definition 4.20 We say an edge of a graph is *singular* if at each of its endpoints it is the only outgoing edge or the only incoming edge. See Figure 37 for an example.

We note that if a component of the graph is a simple closed curve, then every edge in that component is singular.

Proposition 4.21 (1) *If* O_i *and* O_j *are on the interior of the same edge, then multiplication by* U_i *is chain homotopic to multiplication by* U_j. (2) *If* O_i *is associated to a vertex with a single outgoing or incoming edge* e *and* O_j *is on the interior of* e, *then multiplication by* U_i *is chain homotopic to multiplication by* U_j. (3) *If* O_i *is on* *the interior of an edge that is not singular, then multiplication by U_i is null homotopic. Moreover, each of the chain homotopies is a bigraded R_n–module homomorphism of degree $(-w(\mathrm{O}_i), -1)$.*

Note that this implies that the chain homotopies are also bigraded R_V–module homomorphisms.

Proof Let $\mathrm{X}_k \in \mathbb{X}$. We define $H_k\colon C^-(g) \to C^-(g)$ by counting rectangles that contain X_k but do not contain any other X_s. Specifically,

$$H_k(\boldsymbol{x}) := \sum_{\boldsymbol{y}\in \boldsymbol{S}} \sum_{\substack{r\in \mathrm{Rect}^o(\boldsymbol{x},\boldsymbol{y}),\, \mathrm{X}_k\in r\\ \mathrm{X}_s\notin r\ \forall\, \mathrm{X}_s\in\mathbb{X}\setminus\{\mathrm{X}_k\}}} U_1^{\mathrm{O}_1(r)}\cdots U_n^{\mathrm{O}_n(r)}\cdot \boldsymbol{y}.$$

Note that $H_k\colon C^-(g)_{(a,m)} \to C^-(g)_{(a-w(\mathrm{O}_i),m-1)}$. Suppose that X_k shares a row with O_i and shares a column with O_j. There are three cases we need to consider:

(i) If X_k is the only element of $\mathbb{X}$ in its row and column then, like in the proof of [12, Lemma 2.8], we have that

$$\partial^-\circ H_k + H_k\circ\partial^- = U_i + U_j.$$

(ii) If X_k is the only element of $\mathbb{X}$ in its row but it shares its column with other elements of $\mathbb{X}$, then

$$\partial^-\circ H_k + H_k\circ\partial^- = U_i.$$

The difference here is that the vertical annulus containing X_k also includes another X_s for $s\neq k$. Thus it does not contribute to $\partial^-\circ H_k + H_k\circ\partial^-$.

(iii) Similarly, if X_k is the only element of $\mathbb{X}$ in its column but it shares its row with other elements of $\mathbb{X}$, then

$$\partial^-\circ H_k + H_k\circ\partial^- = U_j.$$

Note that we need not consider the case where X_k shares both its row and column with other elements of $\mathbb{X}$. We use the fact that $w(\mathrm{O}_i) = w(\mathrm{O}_j)$ in parts (1) and (2) and the fact that you can add two chain homotopies to get another chain homotopy to complete the proof. □

In Section 5, we show that $(C^-(g), \partial^-)$, viewed as an RA $(H_1(E(f)), \mathbb{Z})$–bigraded R_V–module chain complex, changes by a quasi-isomorphism under graph grid moves. Thus, its homology is an invariant of the spatial graph and not just the grid representative.

Theorem 4.22 *If g_1 and g_2 are saturated graph grid diagrams representing the same transverse spatial graph $f\colon G \to S^3$ then $(C^-(g_1), \partial^-)$ is quasi-isomorphic to $(C^-(g_2), \partial^-)$ as RA $(H_1(E(f)), \mathbb{Z})$–bigraded R_V–modules. In particular,* $\mathrm{HFG}^-(g_1)$ *is isomorphic to* $\mathrm{HFG}^-(g_2)$ *as RA $(H_1(E(f)), \mathbb{Z})$–bigraded R_V–modules.*

Proof Suppose g_1 and g_2 are saturated graph grid diagrams that are related by a cyclic permutation move. Then they have the same toroidal grid $\mathcal{T}$. Note that there is a natural identification of $\boldsymbol{S}$ for both graph grid diagrams so that $C(g_1) = C(g_2)$ as abelian groups. Let $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{S}$. By Corollary 4.14, $A^{g_1}(\boldsymbol{x}) - A^{g_1}(\boldsymbol{y}) = A^{g_2}(\boldsymbol{x}) - A^{g_2}(\boldsymbol{y})$. Hence $A^{g_1}(\boldsymbol{x}) - A^{g_2}(\boldsymbol{x}) = A^{g_1}(\boldsymbol{y}) - A^{g_2}(\boldsymbol{y})$ is a constant a that is independent of element of $\boldsymbol{S}$. By [12, Lemma 2.4], $M(\boldsymbol{x})$ gives the same value for both g_1 and g_2. Thus the identity map id: $C^-(g_1) \to C^-(g_2)$ is an $(H_1(E(f)), \mathbb{Z})$–bigraded R_V–module chain map of degree $(a, 0)$. Thus $(C^-(g_1), \partial^-)$ is quasi-isomorphic to $(C^-(g_1), \partial^-)$ as RA $(H_1(E(f)), \mathbb{Z})$–bigraded R_V–modules. If g_1 and g_2 are saturated graph grid diagrams that are related by a commutation$'$ or stabilization$'$ moves then by Propositions 5.1 and 5.5, $(C^-(g_1), \partial^-)$ is quasi-isomorphic to $(C^-(g_1), \partial^-)$ as RA $(H_1(E(f)), \mathbb{Z})$–bigraded R_V–modules. By Theorem 3.6, g_1 and g_2 are related by a finite sequence of graph grid moves, which completes the proof. □

By Theorem 4.22, the following definitions of $\mathrm{QI}^-(f)$ and $\mathrm{HFG}^-(f)$ are well defined and independent of choice of grid diagram.

Definition 4.23 Let $f\colon G \to S^3$ be a sinkless and sourceless transverse spatial graph. We define $\mathrm{QI}^-(f)$ to be the quasi-isomorphism class of the RA $(H_1(E(f)), \mathbb{Z})$–bigraded R_V–module chain complex $(C^-(g), \partial^-)$, for any saturated graph grid diagram g representing f. The *graph Floer homology* of f, denoted $\mathrm{HFG}^-(f)$, is the homology of $(C^-(g), \partial^-)$ viewed as an RA $(H_1(E(f)), \mathbb{Z})$–bigraded R_V–module, for any saturated graph grid diagram g representing f.

We note that $C^-(g)$ is a finitely generated R_n–module. However, as an R_V–module, it is not finitely generated, but, using Proposition 4.21, we can show that $\mathrm{HFG}^-(f)$ is.

Proposition 4.24 $\mathrm{HFG}^-(f)$ *is a finitely generated R_V–module for any sinkless and sourceless transverse spatial graph $f\colon G \to S^3$.*

Proof The proof is similar to [12, Lemma 2.13]. Let g be a saturated graph grid diagram representing f. We first note that $C^-(g)$ is a finitely generated R_n–module, so $H_*(C^-(g))$ is finitely generated as an R_n–module. Let $[z_1], \dots, [z_r]$ be the generators of $H_*(C^-(g))$ as a finitely generated R_n–module. Let $[c] \in H_*(C^-(g))$. Then we can write $[c] = \sum p_i[z_i]$ for some $p_i \in R_n$. Let $V + 1 \le j \le n$ and $[b] \in \mathrm{HFG}^-(f)$. Then by Proposition 4.21, $U_j[b]$ is either 0 or equal to $U_i[b]$ for some $1 \le i \le V$. Using this repeatedly, it follows that $p_i[z_i] = q_i[z_i]$ for some $q_i \in R_V$. □

4.4 Tilde and hat variants

For a saturated graph grid diagram g, we can define two other variants of $(C^-(g), \partial^-)$. First we define the hat theory. Let $\mathcal{U}_V$ be the $\mathbb{F}$–vector subspace of $C^-(g)$ spanned by $U_1 C^-(g) \cup \cdots \cup U_V C^-(g)$. Define $\widehat{C}(g)$ to be the quotient $C^-(g)/\mathcal{U}_V$. Since $\partial^-(\mathcal{U}_V) \subset \mathcal{U}_V$, it follows that ∂^- descends to an R_V–module homomorphism

$$\widehat{\partial}\colon \widehat{C}(g) \to \widehat{C}(g).$$

Since $C^-(g)$ has a basis of homogeneous elements $\{b_i\}_{i \in I}$ as an $\mathbb{F}$–vector space, with respect to the $(H_1(E(f)), \mathbb{Z})$–grading on $C^-(g)$, and $\mathcal{U}_V$ has a basis that is a subbasis of $\{b_i\}_{i \in I}$, the $(H_1(E(f)), \mathbb{Z})$–bigrading on $C^-(g)$ descends to a well-defined $(H_1(E(f)), \mathbb{Z})$–bigrading on $\widehat{C}(g)$.

Definition 4.25 Let g be saturated graph grid diagram representing the sinkless and sourceless transverse spatial graph $f\colon G \to S^3$. The *graph Floer hat chain complex* of g is the $(H_1(E(f)), \mathbb{Z})$–bigraded chain complex $(\widehat{C}(g), \widehat{\partial})$. The *graph Floer hat homology* of g, denoted $\widehat{\mathrm{HFG}}(g)$, is the homology of $(\widehat{C}(g), \widehat{\partial})$ viewed as an $(H_1(E(f)), \mathbb{Z})$–bigraded vector space over $\mathbb{F}$.

For a given sinkless and sourceless transverse spatial graph f, we can use Theorem 4.22 to show that the quasi-isomorphism class (and hence homology) of $(\widehat{C}(g), \widehat{\partial})$ does not depend on the choice of graph grid diagram representing f. The following lemma is well-known but we include it for completeness.

Lemma 4.26 *Let C and D be $\mathbb{F}[U_1, \dots, U_V]$–module chain complexes and let $\phi\colon C \to D$ be an $\mathbb{F}[U_1, \dots, U_V]$–module quasi-isomorphism. Then ϕ descends to a quasi-isomorphism $\widehat{\phi}\colon \widehat{C} \to \widehat{D}$ of $\mathbb{F}$–vector spaces, where $\widehat{C} = C/\mathcal{U}_V$ and $\widehat{D} = D/\mathcal{U}_V$.*

Proof We prove this by induction on V. Suppose $V = 1$. Then ϕ is an $\mathbb{F}[U_1]$–module homomorphism, and ϕ descends to a well-defined chain map $\widehat{\phi}\colon \widehat{C} \to \widehat{D}$. The following diagram commutes and the horizontal sequences are exact:

$$\begin{tikzcd}
0 \arrow[r] & C \arrow[r, "U_1"] \arrow[d, "\phi"'] & C \arrow[r, "q"] \arrow[d, "\phi"'] & \widehat{C} \arrow[r] \arrow[d, "\widehat{\phi}"'] & 0 \\
0 \arrow[r] & D \arrow[r, "U_1"] & D \arrow[r, "q"] & \widehat{D} \arrow[r] & 0
\end{tikzcd}$$

Here U_1 indicates the map that is multiplication by U_1 and q is the quotient map. Thus on homology, we get the following commutative diagram with horizontal long exact sequences:

$$\begin{tikzcd}
H_*(C) \arrow[r, "(U_1)_*"] \arrow[d, "\phi_*"'] & H_*(C) \arrow[r, "q_*"] \arrow[d, "\phi_*"'] & H_*(\widehat{C}) \arrow[r] \arrow[d, "\widehat{\phi}_*"'] & H_*(C) \arrow[r, "(U_1)_*"] \arrow[d, "\phi_*"'] & H_*(C) \arrow[d, "\phi_*"'] \\
H_*(D) \arrow[r, "(U_1)_*"] & H_*(D) \arrow[r, "q_*"] & H_*(\widehat{D}) \arrow[r] & H_*(D) \arrow[r, "(U_1)_*"] & H_*(D)
\end{tikzcd}$$

Since ϕ_* is an isomorphism, by the five lemma, so is $\widehat{\phi}_*$.

Now suppose the lemma is true for V and let C and D be $\mathbb{F}[U_1,\dots,U_{V+1}]$–module chain complexes and $\phi\colon C\to D$ be an $\mathbb{F}[U_1,\dots,U_{V+1}]$–module quasi-isomorphism. Consider C and D as $\mathbb{F}[U_1,\dots,U_V]$–modules. Then by the inductive hypothesis, $\phi'\colon C/\mathcal{U}_V\to D/\mathcal{U}_V$ is a quasi-isomorphism, where ϕ' is induced from ϕ (which we called $\widehat{\phi}$ before). Moreover, note that ϕ' is an $\mathbb{F}[U_{V+1}]$–module homomorphism. Let $C'=C/\mathcal{U}_V$ and $D'=D/\mathcal{U}_V$. Using the proof from the case when $V=1$, we can show that the induced map $\widehat{\phi}'\colon C'/U_{V+1}C'\to D'/U_{V+1}D'$ is a quasi-isomorphism. It is straightforward to show that the natural map $C/\mathcal{U}_{V+1}\to C'/U_{V+1}C'$ is a chain isomorphism (similarly for D), which completes the proof. □

Corollary 4.27 *If g_1 and g_2 are saturated graph grid diagrams that represent the same transverse spatial graph $f\colon G\to S^3$, then $(\widehat{C}(g_1),\widehat{\partial})$ and $(\widehat{C}(g_2),\widehat{\partial})$ are quasi-isomorphic as RA $(H_1(E(f)),\mathbb{Z})$–bigraded vector spaces. In particular, $\widehat{\mathrm{HFG}}(g_1)$ is isomorphic to $\widehat{\mathrm{HFG}}(g_2)$ as RA $(H_1(E(f)),\mathbb{Z})$–bigraded vector spaces.*

As a result, the following definitions of $\widehat{\mathrm{QI}}(f)$ and $\widehat{\mathrm{HFG}}(f)$ are well-defined and independent of choice of graph grid diagram.

Definition 4.28 Let $f\colon G\to S^3$ be a sinkless and sourceless transverse spatial graph. We define $\widehat{\mathrm{QI}}(f)$ to be the quasi-isomorphism class of the RA $(H_1(E(f)),\mathbb{Z})$–bigraded chain complex $(\widehat{C}(g),\widehat{\partial})$, for any saturated graph grid diagram g representing f. The *graph Floer hat homology* of f, denoted $\widehat{\mathrm{HFG}}(f)$, is the homology of $(\widehat{C}(g),\widehat{\partial})$ viewed as an RA $(H_1(E(f)),\mathbb{Z})$–bigraded vector space over $\mathbb{F}$, for any saturated graph grid diagram g representing f.

Note that $(\widehat{C}(g),\widehat{\partial})$ is an infinitely generated vector space, but in the same way as Proposition 4.24 one can show that its homology is finitely generated.

Proposition 4.29 *$\widehat{\mathrm{HFG}}(f)$ is a finitely generated vector space over $\mathbb{F}$ for any sinkless and sourceless transverse spatial graph $f\colon G\to S^3$.*

Proof Choose a saturated graph grid diagram g representing f. Let j be such that $V+1\le j\le n$. Then by Proposition 4.21, there is a chain homotopy $H\colon C^-(g)\to C^-(g)$

(H is H_k for some k) that is an R_V–module homomorphism and satisfies one of the following conditions: (i) $\partial^- \circ H + H \circ \partial^- = U_j$ or (ii) there exists an $1 \le i \le V$ such that $\partial^- \circ H + H \circ \partial^- = U_i + U_j$. Since H is an R_V–module homomorphism, H descends to a well-defined $\widehat{H}\colon \widehat{C}(g) \to \widehat{C}(g)$ satisfying

$$\widehat{\partial} \circ \widehat{H} + \widehat{H} \circ \widehat{\partial} = U_j.$$

Therefore $U_j[b] = 0$ for all $[b] \in \widehat{\mathrm{HFG}}(g)$. The rest of the proof is similar to the proof of Proposition 4.24. □

We now define the tilde theory. The tilde theory will be the easiest theory to compute. However, it narrowly fails to be an invariant of the spatial graph since it will depend on the grid size. On the other hand, one can recover the hat theory from it, which makes it quite useful. It will also be easier to compute the bigraded Euler characteristic (Alexander polynomial) of the hat theory using tilde theory; for more details, see Section 6.

Let $\mathcal{U}_n$ be the $\mathbb{F}$–vector subspace of $C^-(g)$ spanned by $U_1 C^-(g) \cup \cdots \cup U_n C^-(g)$. Define $\widetilde{C}(g)$ to be the quotient $C^-(g)/\mathcal{U}_n$. Since $\partial^-(\mathcal{U}_n) \subset \mathcal{U}_n$, it follows that ∂^- descends to a linear map

$$\widetilde{\partial}\colon \widetilde{C}(g) \to \widetilde{C}(g)$$

of vector spaces over $\mathbb{F}$. Since $C^-(g)$ has a basis of homogeneous elements $\{b_i\}_{i \in I}$ as an $\mathbb{F}$–vector space, with respect to the $(H_1(E(f)), \mathbb{Z})$ grading on $C^-(g)$, and $\mathcal{U}_n$ has a basis that is a subbasis of $\{b_i\}_{i \in I}$, the $(H_1(E(f)), \mathbb{Z})$–bigrading on $C^-(g)$ descends to a well-defined $(H_1(E(f)), \mathbb{Z})$–bigrading on $\widetilde{C}(g)$. Thus $(\widetilde{C}(g), \widetilde{\partial})$ is a $(H_1(E(f)), \mathbb{Z})$–bigraded chain complex.

Definition 4.30 Let g be a saturated graph grid diagram representing the sinkless and sourceless transverse spatial graph $f\colon G \to S^3$. The *graph Floer tilde chain complex* of g is the $(H_1(E(f)), \mathbb{Z})$–bigraded chain complex $(\widetilde{C}(g), \widetilde{\partial})$. The *graph Floer tilde homology* of g, denoted $\widetilde{\mathrm{HFG}}(g)$, is the homology of $(\widetilde{C}(g), \widetilde{\partial})$ viewed as an $(H_1(E(f)), \mathbb{Z})$–bigraded vector space over $\mathbb{F}$.

We will relate $\widetilde{\mathrm{HFG}}(g)$ and $\widehat{\mathrm{HFG}}(g)$ for a given graph grid diagram g. First, we recall the (bigraded) mapping cone which we will use in the next lemma.

Let (A, ∂^A) and (B, ∂^B) be $(\mathbb{G}, \mathbb{Z})$–bigraded chain complexes and let $\phi\colon A \to B$ be a bigraded chain map of degree (g, m) for some $g \in \mathbb{G}$ and $m \in \mathbb{Z}$. Define the *(bigraded) mapping cone complex* of ϕ, denoted $(\mathrm{cone}(\phi), \partial)$, as follows:

$$\mathrm{cone}(\phi) = \bigoplus_{(h,n) \in \mathbb{G} \oplus \mathbb{Z}} \mathrm{cone}(\phi)_{(h,n)},$$

where

$$\mathrm{cone}(\phi)_{(h,n)} = A_{(h-g,n-m-1)} \oplus B_{(h,n)}$$

and the boundary map is defined as

$$\partial(a,b) = (-\partial^A(a), -\phi(a) + \partial^B(b))$$

for all $a \in A$ and $b \in B$. Checking the definitions, we see that ∂ is a bigraded map of degree $(0,-1)$. We also note that if A and B are $(\mathbb{G},\mathbb{Z})$–bigraded R–module chain complexes and ϕ is a bigraded R–module chain map then $(\mathrm{cone}(\phi),\partial)$ is an $(\mathbb{G},\mathbb{Z})$–bigraded R–module chain complex.

The following lemma is similar to [12, Lemma 2.14] except now we have bigraded chain complexes instead of filtered graded chain complexes. For $g \in \mathbb{G}$ and $m \in \mathbb{Z}$, let $W(-g,-m+1)$ be the two dimensional $(\mathbb{G},\mathbb{Z})$–bigraded vector space over $\mathbb{F}$ spanned by one generator in degree $(0,0)$ and the other in degree $(-g,-m+1)$. If (C,∂) is any bigraded $(\mathbb{G},\mathbb{Z})$–chain complex over $\mathbb{F}$, then $C \otimes W(-g,-m+1)$ becomes a bigraded chain complex with boundary $\partial \otimes \mathrm{id}$ in the usual way. That is,

$$(C \otimes W(-g,-m+1))_{(h,l)} = \bigoplus_{(h,l)=(h_1+h_2,l_1+l_2)} C_{(h_1,l_1)} \otimes W(-g,-m+1)_{(h_2,l_2)}.$$

Lemma 4.31 *Let (C,∂) be a $(\mathbb{G},\mathbb{Z})$–bigraded $\mathbb{F}[U_1,\dots,U_s]$–module chain complex and $g \in \mathbb{G}$ and $m \in \mathbb{Z}$ be fixed group elements. Suppose that for each $i \geq 2$, multiplication by U_i (which we denote by U_i) is a bigraded $\mathbb{F}[U_1,\dots,U_s]$–module chain map of degree $(-g,-m)$ and that*

(1) *U_i is chain homotopic to U_1 or*
(2) *U_i is null-homotopic (where the chain homotopy is an $\mathbb{F}[U_1,\dots,U_s]$–module homomorphism).*

Then $(C/\mathcal{U}_s,\partial)$ is quasi-isomorphic to $(C/\mathcal{U}_1 \otimes W(-g,-m+1)^{\otimes s-1}, \partial \otimes \mathrm{id})$ and hence

$$H_*(C/\mathcal{U}_s) \cong H_*(C/\mathcal{U}_1) \otimes W(-g,-m+1)^{\otimes s-1}.$$

Note that by $H_*(C/\mathcal{U}_1)$ (respectively $H_*(C/\mathcal{U}_s)$), we mean the homology of the chain complex whose chain group is $C/\mathcal{U}_1$ (respectively $C/\mathcal{U}_s$) and whose boundary map is induced by ∂.

Proof Let $D = C/\mathcal{U}_1 = C/U_1C$ and $\partial^D\colon D \to D$ be induced by ∂. Consider multiplication by U_2 on D, $\widehat{U}_2\colon D \to D$. Since U_1 and U_2 commute, this is a well-defined bigraded map. There is a long exact sequence

$$0 \to D \xrightarrow{\widehat{U}_2} D \xrightarrow{\mathrm{pr}} D/\widehat{U}_2 D \to 0,$$

where pr is the natural projection to the quotient. Let $F\colon \mathrm{cone}(\widehat{U}_2) \to D/\widehat{U}_2 D$ be defined by $F(c_1, c_2) = \mathrm{pr}(c_2)$. By [23, Section 1.5.8], the map F is a quasi-isomorphism. Moreover, F is a bigraded map of degree $(0,0)$.

If U_2 is chain homotopic to U_1 via a chain homotopy H that is an $\mathbb{F}[U_1, \dots, U_s]$–module homomorphism, then H induces a well-defined map $\widehat{H}\colon D \to D$ such that

$$\partial^D \circ \widehat{H} + \widehat{H} \circ \partial^D = \widehat{U}_2.$$

This also holds if U_2 is null-homotopic. Since $\widehat{U}_2$ is null-homotopic, there is a bigraded chain isomorphism from $\mathrm{cone}(\widehat{U}_2)$ to $\mathrm{cone}(0\colon D \to D)$ of degree $(0,0)$. Hence $\mathrm{cone}(\widehat{U}_2)$ is isomorphic to $D \oplus D[g, m-1]$ as bigraded chain complexes, where $D[g, m-1]$ is the bigraded vector space defined by $D[g, m-1]_{(h,n)} = D_{h+g,n+m-1}$ and the boundary map on $D \oplus D[g, m-1]$ is $\partial^D \oplus \partial^D$. Moreover, this is isomorphic, as bigraded chain complexes, to $D \otimes W(-g, -m+1)$. The proof for $n = 2$ is complete after noting that the obvious map $C/\mathcal{U}_2$ to $D/\widehat{U}_2 D$ is a bigraded chain isomorphism. To complete this proof, continue this type of argument, one by one for each U_i. □

We can use this to relate the tilde and hat chain complexes.

Proposition 4.32 *Let g be a saturated graph grid diagram representing the sinkless and sourceless transverse spatial graph $f\colon G \to S^3$. Then*

$$\widetilde{\mathrm{HFG}}(g) \cong \widehat{\mathrm{HFG}}(g) \otimes \bigotimes_{e \in E(G)} W(-w(e), -1)^{\otimes n_e}$$

as $(H_1(E(f)), \mathbb{Z})$–bigraded $\mathbb{F}$–vector spaces, where n_e is the number of O's in g associated to the interior of e (not including the vertices).

Proof We note that if O_i is on the interior of edge e then multiplication by U_i is a graded map of degree $(-w(e), -2)$. In addition, U_i is either null homotopic or homotopic to some U_j with $j \leq V$, where O_j is a vertex. Use Lemma 4.31 repeatedly to complete the proof. □

5 Invariance of $\mathrm{HFG}^-(f)$

In this section we will complete the necessary steps to prove that $\mathrm{HFG}^-(f)$ is an invariant of a sinkless and sourceless transverse spatial graph; see Theorem 4.22 and its proof for more details. That is, we will show that $\mathrm{HFG}^-(g)$ is invariant under each of the graph grid moves: commutation$'$ and stabilization$'$. The proofs will be similar to the proofs found in [12]. There are two major differences though. The first is that we are working with more general commutation and stabilization moves, commutation$'$ and stabilization$'$. The second difference is the Alexander grading.

5.1 Commutation$'$ invariance

Proposition 5.1 *Suppose g and $\bar{g}$ are saturated graph grid diagrams that differ by a commutation$'$ move. Let $f\colon G \to S^3$ be the transverse graph associated to g and $\bar{g}$, and V be the number of vertices of G. Then there is an $(H_1(E(f)), \mathbb{Z})$–bigraded R_V–module quasi-isomorphism $(C(g), \partial_g^-) \to (C(\bar{g}), \partial_{\bar{g}}^-)$ of degree $(\delta(g,\bar{g}), 0)$ for some $\delta(g,\bar{g}) \in H_1(E(f))$.*

We remark that the quasi-isomorphism above will be a bigraded map for some degree (that depends on g and $\bar{g}$), but will not necessarily be of degree $(0,0)$.

The proof of Proposition 5.1 will take up the rest of this subsection. We will prove the case when $\bar{g}$ is obtained from g by a commutation$'$ move of columns. The case where you exchange rows is similar.

As in [12], we draw both graph grid diagrams on a single $n \times n$ grid (respectively torus when the sides are identified), which we will call the combined grid diagram, as follows. Let the vertical line segment (respectively circle) between the columns that are exchanged be labeled β in g and γ in $\bar{g}$ and call the other vertical circles $\beta_1, \dots, \beta_{n-1}$, where n is the size of the grid for g. Let γ be a simple closed curve on the graph grid diagram g such that the following conditions are held: (1) γ is homotopic to β, (2) γ hits each of the horizontal curves, α_i, precisely once, (3) γ does not intersect β_i for $i \le n-1$, (4) after removing the β curve, one obtains $\bar{g}$, (5) γ and β intersect transversely exactly twice, and (6) the intersections of γ and β do not lie on the horizontal curves. It is easy to use the line segments LS_1 and LS_2 in the definition of commutation$'$ to see that such a curve exists. First, note that we can assume the endpoints of the line segments do not lie on α curves by slightly changing them. Now, take pushoffs of the line segments LS_1 and LS_2 to the left or right as needed, and connect them up so that they satisfy the requirements above. Let a and b be the intersections of β and γ. See Figures 38 and 39 for examples. We still let $\mathcal{T}$ be the torus of the combined grid diagram obtained by gluing the top/bottom and sides.

We will define a chain map $\Phi_{\beta\gamma}\colon C^-(g) \to C^-(\bar{g})$ and show that it is a chain homotopy equivalence. This will show that $\Phi_{\beta\gamma}$ is a quasi-isomorphism. For $\boldsymbol{x} \in \boldsymbol{S}(g)$ and $\boldsymbol{y} \in \boldsymbol{S}(\bar{g})$, we let $\mathrm{Pent}_{\beta\gamma}(\boldsymbol{x}, \boldsymbol{y})$ be the set of embedded pentagons with the following

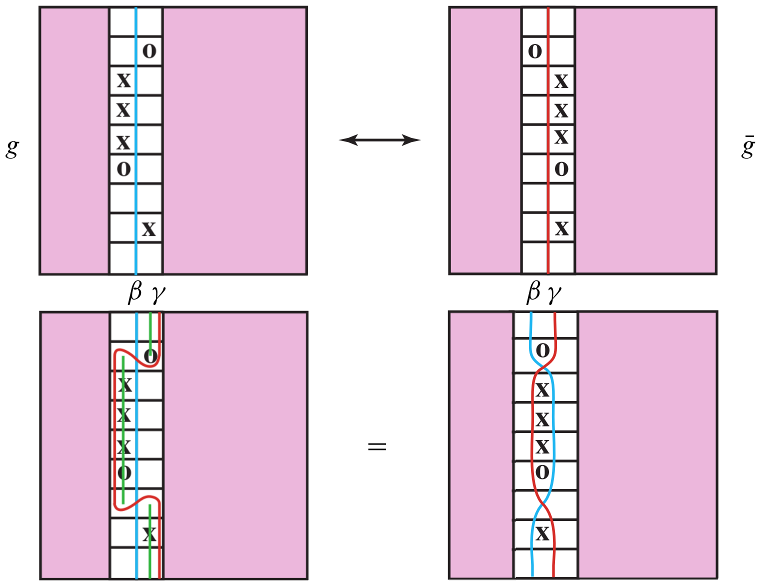

Figure 38: Commutation$'$ move between g and $\bar{g}$ (top), and the corresponding combined grid diagram, where we show the line segments in the definition of commutation$'$ on the left (bottom)

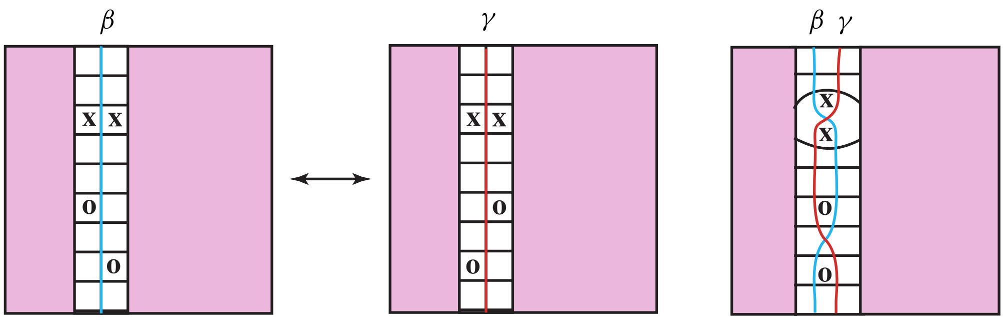

Figure 39: Another example of a commutation$'$ move, in which the X's appear in the same row; from left to right we have g, $\bar{g}$ and the combined grid diagram

properties. If $\boldsymbol{x}$ and $\boldsymbol{y}$ do not coincide at $n-2$ points, then we let $\mathrm{Pent}_{\beta\gamma}(\boldsymbol{x},\boldsymbol{y})=\varnothing$. Suppose that $\boldsymbol{x}$ and $\boldsymbol{y}$ coincide at $n-2$ points (say $x_3=y_3,\dots,x_n=y_n$). Without loss of generality, let $x_2=\boldsymbol{x}\cap\beta$ and $y_2=\boldsymbol{y}\cap\gamma$. An element $p\in\mathrm{Pent}_{\beta\gamma}(\boldsymbol{x},\boldsymbol{y})$ is an embedded disk in $\mathcal{T}$, whose boundary consists of five arcs, each of which are contained in the circles β_i, α_i, β or γ and satisfies the following conditions. The intersections of the arcs lie on the points x_1,x_2,y_1,y_2 and a. The point a is in $\beta\cap\gamma$ and locally looks like the top intersection in Figure 40 (b is the one that locally looks the bottom

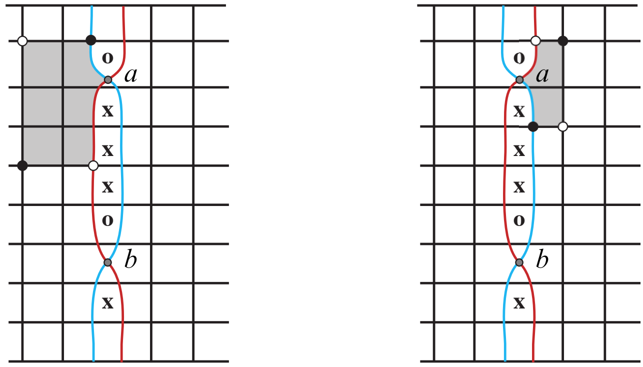

Figure 40: Examples of pentagons in $\mathrm{Pent}^o_{\beta\gamma}(\boldsymbol{x},\boldsymbol{y})$

intersection point in $\beta\cap\gamma$). Moreover, start at the point in x_2 and transverse the boundary of p, using the orientation given by p. The condition to be in $\mathrm{Pent}_{\beta\gamma}(\boldsymbol{x},\boldsymbol{y})$ is that you will first travel along a horizontal circle, meet y_1, proceed along a vertical circle β_i, meet x_1, continue along another horizontal circle, meet y_2, proceed though an arc in γ until you meet a, and finally traverse an arc in β until arriving back at x_2. Finally, all angles are required to be less than straight.

The set of empty pentagons, $\mathrm{Pent}^o_{\beta\gamma}(\boldsymbol{x},\boldsymbol{y})$, are those pentagons $p\in\mathrm{Pent}_{\beta\gamma}(\boldsymbol{x},\boldsymbol{y})$ such that $\boldsymbol{x}\cap\mathrm{Int}(p)=\varnothing$. The map $\Phi_{\beta\gamma}\colon C^-(g)\to C^-(\bar{g})$ is defined by counting empty pentagons, that do not contain X's in the combined grid diagram as follows. For $\boldsymbol{x}\in\boldsymbol{S}(g)$, define

$$\Phi_{\beta\gamma}(\boldsymbol{x})=\sum_{\boldsymbol{y}\in\boldsymbol{S}(\bar{g})}\sum_{\substack{p\in\mathrm{Pent}^o_{\beta\gamma}(\boldsymbol{x},\boldsymbol{y})\\ \mathrm{Int}(p)\cap\mathbb{X}=\varnothing}} U_1^{O_1(p)}\cdots U_n^{O_n(p)}\cdot\boldsymbol{y}\in C^-(\bar{g}).$$

Extend $\Phi_{\beta\gamma}$ to $C^-(g)$ so that it is an R_n–module homomorphism. In particular, it is also an R_V–module homomorphism.

Lemma 5.2 *$\Phi_{\beta\gamma}$ is an $(H_1(E(f)),\mathbb{Z})$–bigraded R_V–module chain map of some degree.*

Proof Since $\Phi_{\beta\gamma}$ is an R_V–module homomorphism, the proof is broken into three parts: checking that each grading is preserved, and showing that

$$\partial^-\circ\Phi_{\beta\gamma}=\Phi_{\beta\gamma}\circ\partial^-.$$

The map $\Phi_{\beta\gamma}$ preserves the Maslov grading Since the definition of the Maslov grading only depends on the set $\mathbb{O}$, and we consider a subset of the pentagons considered

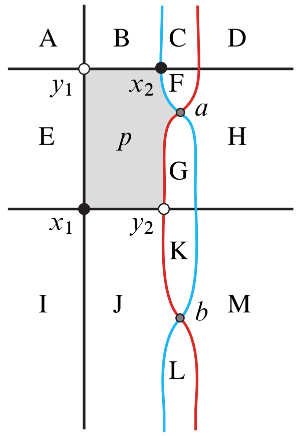

Figure 41: The combined grid diagram, with regions $\mathrm{A}, \dots, \mathrm{M}$ labeled and the pentagon p shaded

in the proof of commutation in [12, Section 3.1], this technically follows from [12, Lemma 3.1]. However, since they do not include a proof that the Maslov grading is preserved (this is left to the reader), we will include a sketch of the proof here.

We will go though the details of this computation for the case pictured in Figure 41, other cases follow similarly. Consider a $U_1^{O_1(p)} \cdots U_n^{O_n(p)} \cdot \boldsymbol{y}$ in the sum of $\Phi_{\beta\gamma}(\boldsymbol{x})$. Recall that

$$M(\boldsymbol{x}) = \mathcal{J}(\boldsymbol{x}, \boldsymbol{x}) - 2\mathcal{J}(\boldsymbol{x}, \mathbb{O}) + \mathcal{J}(\mathbb{O}, \mathbb{O}) + 1.$$

To compare the Maslov grading we interpret each of these terms for $\boldsymbol{x}$ in the grid g in relation to $\boldsymbol{y}$ in the grid $\bar{g}$. Let the intersection points of $\boldsymbol{x}$ be $x_1, \dots, x_n$ and the intersection points of $\boldsymbol{y}$ be $y_1, \dots, y_n$, with the same subscript where they coincide. Label the intersection points where $\boldsymbol{x}$ and $\boldsymbol{y}$ differ as x_1, x_2, y_1 and y_2 and break the combined grid diagram into 14 regions labeled $\mathrm{A}, \dots, \mathrm{M}$, and p, as shown in Figure 41.

Notice that the count for x_i is the same as y_i for $i \neq 1, 2$. The number of points in $\mathbb{O}$ up and to the right, and down and to the left are not changed, since this could only be changed for an intersection between the commuted edges (ie x_2 or y_2). So $\mathcal{J}(x_i, \mathbb{O}) = \mathcal{J}(y_i, \mathbb{O})$. The number of points in $\boldsymbol{x}$ and $\boldsymbol{y}$ up and to the right and down and to the left are the same, this can be checked region by region. If an intersection point is in region E then x_2 will be counted in the points up and to the right; this is replaced by the point y_1 which is also up and to the right. Similarly, for all of the regions, since $x_i = y_i \in \{\mathrm{A}, \mathrm{B}, \dots, \mathrm{M}\}$, we have $\mathcal{J}(x_i, \boldsymbol{x}) = \mathcal{J}(y_i, \boldsymbol{y})$ for all $i \neq 1, 2$.

Now for the points where $\boldsymbol{x}$ and $\boldsymbol{y}$ differ, Figure 42 (left) shows the regions that are counted for x_1 and x_2, and Figure 42 (right) shows the regions that are counted for

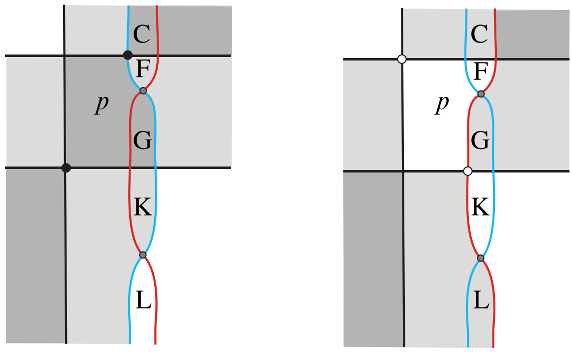

Figure 42: The lightly shaded regions are those that will be counted with weight one half in $\mathcal{J}(x_1,-)$ and $\mathcal{J}(x_2,-)$ (left), or in $\mathcal{J}(y_1,-)$ and $\mathcal{J}(y_2,-)$ (right).

y_1 and y_2. So we see that the regions C, F, G and K are counted with weight $\frac{1}{2}$ more for x_1 and x_2, and the region p is counted with weight 1 more for x_1 and x_2 and, lastly, region L is counted with weight $\frac{1}{2}$ less for x_1 and x_2. So,

$$\mathcal{J}(\boldsymbol{x},\boldsymbol{x}) = \mathcal{J}(\boldsymbol{y},\boldsymbol{y}) + 1,$$

because x_1 will count x_2 with weight $\frac{1}{2}$ and vice versa, but y_1 and y_2 do not count each other. Next, $\boldsymbol{x}$ will count all of the points in $\mathbb{O}$ that $\boldsymbol{y}$ will count and, additionally, will count those O's in the region p with weight 1, those O's in the regions C, F, G, and K with weight $\frac{1}{2}$ and those O's in the region L with weight $-\frac{1}{2}$. Notice that the region made up of G and K must contain exactly one O. Thus,

$$\mathcal{J}(\boldsymbol{x},\mathbb{O}) = \mathcal{J}(\boldsymbol{y},\mathbb{O}) + \mathrm{O}_1(p) + \cdots + \mathrm{O}_n(p) + \tfrac{1}{2}(\mathrm{O}(C) + \mathrm{O}(F) + 1) - \tfrac{1}{2}\mathrm{O}(L),$$

where $\mathrm{O}(C)$, $\mathrm{O}(F)$, and $\mathrm{O}(L)$ are the number of O's in the respective regions.

Lastly, if we look at what happens for the different diagrams with the sets of $\mathbb{O}$, the only difference is for the O's in the columns that are changed. Again we know that there is exactly one O in the regions G and K. So we have

$$\mathcal{J}(\mathbb{O},\mathbb{O})_g = \mathcal{J}(\mathbb{O},\mathbb{O})_{\bar{g}} + \mathrm{O}(C) + \mathrm{O}(F) - \mathrm{O}(L).$$

Putting this all together, we have

$$\begin{aligned} M(\boldsymbol{x}) = {} & [\mathcal{J}(\boldsymbol{y},\boldsymbol{y}) + 1] \\ & \quad - 2[\mathcal{J}(\boldsymbol{y},\mathbb{O}) + \mathrm{O}_1(p) + \cdots + \mathrm{O}_n(p) + \tfrac{1}{2}(\mathrm{O}(C) + \mathrm{O}(F) + 1) - \tfrac{1}{2}\mathrm{O}(L)] \\ & \quad + [\mathcal{J}(\mathbb{O},\mathbb{O})_{\bar{g}} + \mathrm{O}(C) + \mathrm{O}(F) - \mathrm{O}(L)] + 1 \\ = {} & \mathcal{J}(\boldsymbol{y},\boldsymbol{y}) - 2\mathcal{J}(\boldsymbol{y},\mathbb{O}) + \mathcal{J}(\mathbb{O},\mathbb{O})_{\bar{g}} - 2[\mathrm{O}_1(p) + \cdots + \mathrm{O}_n(p)] + 1 \\ = {} & \mathcal{J}(\boldsymbol{y} - \mathbb{O}, \boldsymbol{y} - \mathbb{O}) + 1 - 2[\mathrm{O}_1(p) + \cdots + \mathrm{O}_n(p)]. \end{aligned}$$

Thus we see that the Maslov grading is unchanged.

The map $\Phi_{\beta\gamma}$ preserves the Alexander grading up to a shift Now, consider the $H_1(E(f))$–grading on $C^-(g)$, $C^-(g)=\bigoplus_{a\in H_1(E(f))}C^-(g)_a$ (similarly for $C^-(\bar{g})$). We will show that there is an element $\delta(g,\bar{g})$ that only depends on g and $\bar{g}$ (not on $\boldsymbol{x}$) such that $\Phi_{\beta\gamma}(C^-(g)_a)\subset C^-(\bar{g})_{a+\delta(g,\bar{g})}$. We will work with the second definition of the Alexander grading, $A^g(\boldsymbol{x})=\sum_{x_i\in\boldsymbol{x}}[-h^g(x_i)]$ to prove this.

Let $\boldsymbol{x}\in\boldsymbol{S}(g)$ and $p\in\mathrm{Pent}^o_{\beta\gamma}(\boldsymbol{x},\boldsymbol{y})$ such that $p\cap\mathbb{X}=\varnothing$ and $\mathrm{Pent}^o_{\beta\gamma}(\boldsymbol{x},\boldsymbol{y})\neq\varnothing$. Then $U_1^{\mathrm{O}_1(p)}\cdots U_n^{\mathrm{O}_n(p)}\cdot\boldsymbol{y}$ is a term in $\Phi_{\beta\gamma}(\boldsymbol{x})$. We use the convention as before that $x_i=y_i$ for $i\geq 3$, $x_2\in\beta$, and $y_2\in\gamma$. We note that $h^g(x_i)=h^{\bar{g}}(y_i)$ for $i\geq 3$. So we need to show that

$$h^g(x_1)+h^g(x_2)-h^{\bar{g}}(y_1)-h^{\bar{g}}(y_2)-\sum_{i=1}^{n}w(\mathrm{O}_i)\mathrm{O}_i(p)=\delta(g,\bar{g}) \tag{6}$$

for some fixed $\delta(g,\bar{g})\in H_1(E(f))$.

We will prove the case when a is the topmost intersection of β and γ and p is a pentagon lying to the left of a. See two examples of these pentagons in Figure 43. Note that the boundary of the pentagon can contain b, and p can contain an O that lies between β and γ. The other three cases are similar. Let β_{n-1} be the vertical line segment/circle in g directly to the left of β, and β_1 be the vertical line segment/circle in g directly to the right of β. We will label the α_i in the usual way so that α_i is height $i-1$. Let α_l be the horizontal line segment/circle directly below b, α_{l+1} be the horizontal line segment/circle directly above b, α_k be the horizontal circle directly below a, and α_{k+1} be the horizontal line segment/circle directly above a. Finally, let u_1 be the point on β_{n-1} that is at the same height as x_1 and let u_2 be the point on β_{n-1} that is at the same height as x_2. See Figures 43 and 44 for our conventions.

We will say that the pentagon p is narrow if $y_1\in\beta_{n-1}$. If p is not narrow, then there is a rectangle in g that is contained in p. Let r be the largest such rectangle. Then p decomposes into r and a narrow pentagon p'. Note that $r\in\mathrm{Rect}^o(\{x_1,u_2\},\{u_1,y_1\})$ and $p'\in\mathrm{Pent}^o(\{x_2,u_1\},\{y_2,u_2\})$. Moreover, $r\cap\mathbb{X}=p\cap\mathbb{X}=\varnothing$. Since the boundary map on $C^-(g)$ preserves the Alexander grading, we see that

$$h^g(x_1)+h^g(u_2)=\sum_{i=1}^{n}w(\mathrm{O}_i)\mathrm{O}_i(r)+h^g(y_1)+h^g(u_1).$$

We consider $h^g(x_2)-h^g(u_2)$ and $h^{\bar{g}}(y_2)-h^{\bar{g}}(u_1)=h^{\bar{g}}(y_2)-h^g(u_1)$.

In order to compute h^g or $h^{\bar{g}}$, draw the transverse spatial graph for g and g' so that the horizontal and vertical arcs connecting the X's and O's are inside the grid, and

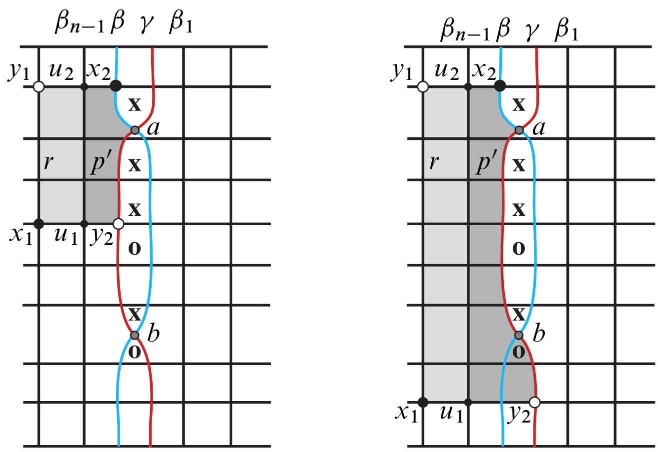

Figure 43: Decomposing p into a narrow pentagon p' and a rectangle r

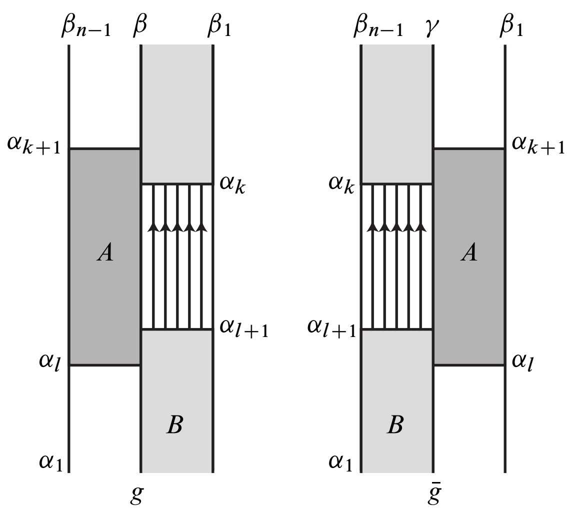

Figure 44: A commutation$'$ move; regions A and B contain X's and O's

consider their projections $\mathrm{pr}(g)$ and $\mathrm{pr}(g')$. We will think of column $n-1$ as the column with β_{n-1} on the left and column n as the column with β or γ on the left. Assume that LS_1 lies inside column $n-1$ and LS_2 lies in column n. Let A be the union of rectangles containing LS_1 and B be the union of rectangles containing LS_2 (recall, we are assuming the ends of LS_i do not lie on an α curve). Then the X's and O's in columns $n-1$ and n are contained in $A \cup B$ and projections of A and B intersect in the two rows that contain a and b. Since there are no X's or O's in $\mathrm{col}_n \backslash B$, the vertical arcs in $\mathrm{pr}(g) \cap (\mathrm{col}_n \setminus B)$ form a collection of parallel arcs starting and stopping at α_k and α_{k+1} which are all oriented in the same direction (all upwards or all downwards). In addition, $\mathrm{pr}(g) \cap (\mathrm{col}_{n-1} \setminus A)$ is empty. See Figure 44.

Let O' be the O in column $n-1$ of g and let O'' be the O in column $n-1$ of $\bar{g}$. Note that p' either contains O' or O'' or both or is empty (in particular, it never contains an element of $\mathbb{X}$). We first consider $h^g(x_2) - h^g(u_2)$. There are three cases. First suppose $x_2 \in \alpha_i$ for $i \geq k+1$ or $i \leq l$. Then $h^g(x_2) - h^g(u_2) = 0$ since the region above and below A in g contains no vertical arcs in $\mathrm{pr}(g)$. In addition, we see that p' cannot contain O' so that $h^g(x_2) - h^g(u_2) = 0 = \mathrm{O}'(p')w(\mathrm{O}')$ where by $\mathrm{O}'(p')$, we mean the number of O's in p'. Now suppose $x_2 \in \alpha_i$ for $l+1 \leq i \leq k$. If p' does not contain O' then $h^g(x_2) - h^g(u_2) = 0 = \mathrm{O}'(p')w(\mathrm{O}')$. If p' contains O' then the arc going from u_2 to x_2 crosses all the vertical strands emanating from this O, all oriented downwards, since p' does not contain any X's. Thus, $h^g(x_2) - h^g(u_2) = \mathrm{O}'(p')w(\mathrm{O}')$. Thus, in all cases, we see that $h^g(x_2) - h^g(u_2) = \mathrm{O}'(p')w(\mathrm{O}')$.

Now consider $h^{\bar{g}}(y_2) - h^{\bar{g}}(u_1)$. We again have three cases to consider. First suppose that $y_2 \in \alpha_j$ for $l+1 \leq j \leq k$. Then the arc from u_1 to y_2 crosses m vertical arcs, all oriented in the same direction so $h^{\bar{g}}(y_2) - h^{\bar{g}}(u_1) = \eta$ for some $\eta \in H_1(E(f))$. In addition, $\mathrm{O}''(p) = 0$ so that $h^{\bar{g}}(y_2) - h^{\bar{g}}(u_1) = \eta - \mathrm{O}''(p')w(p')$. Now suppose that $x_2 \in \alpha_i$ for $i \geq k+1$ or $i \leq l$. If p' does not contain O'' then $h^{\bar{g}}(y_2) - h^{\bar{g}}(1_2) = \eta = \eta - \mathrm{O}''(p')w(p')$. If p' contains O'', $h^{\bar{g}}(y_2) - h^{\bar{g}}(u_1) = \eta - w(\mathrm{O}'')$. Thus, in all cases, $h^{\bar{g}}(y_2) - h^{\bar{g}}(u_1) = \eta - \mathrm{O}''(p')w(p')$.

Putting this together and using that $h^{\bar{g}}(y_1) = h^g(y_1)$ and $h^{\bar{g}}(u_1) = h^g(u_1)$, we have

$$\begin{aligned} h^g(x_1) + h^g(x_2) - h^{\bar{g}}(y_1) - h^{\bar{g}}(y_2) &= (h^g(x_1) + h^g(u_2)) - (h^g(y_1) + h^g(u_1)) \\ &\qquad + (h^g(x_2) - h^g(u_2)) - (h^{\bar{g}}(y_2) - h^{\bar{g}}(u_1)) \\ &= \sum_{i=1}^n w(\mathrm{O}_i)\mathrm{O}_i(r) + \mathrm{O}'(p')w(\mathrm{O}') - (\eta - \mathrm{O}''(p')w(p')) \\ &= \sum_{i=1}^n w(\mathrm{O}_i)\mathrm{O}_i(p) - \eta. \end{aligned}$$

Thus, (6) holds with $\delta(g, \bar{g}) = -\eta$ which completes the proof that $\Phi_{\beta\gamma}$ is a graded map (with respect to the Alexander grading).

$\Phi_{\beta\gamma}$ is a chain map The remaining portion of the proof that $\Phi_{\beta\gamma}$ is a chain map follows almost immediately from the proof of [12, Lemma 3.1]. However, our pentagons and rectangles cannot count X's so we need to be a little more careful.

For $\boldsymbol{x} \in \boldsymbol{S}(g)$, there is a unique element $c(\boldsymbol{x})$ of $\boldsymbol{S}(\bar{g})$, called the canonical closest generator of $\boldsymbol{x}$, defined as follows. Let t be such that $\boldsymbol{x} \cap \beta \in \alpha_t$ and let x' be the point in $\alpha_t \cap \gamma$. Define

$$c(\boldsymbol{x}) := \{x_i \in \boldsymbol{x} \mid x_i \notin \beta\} \cup \{x'\} \in \boldsymbol{S}(\bar{g}).$$

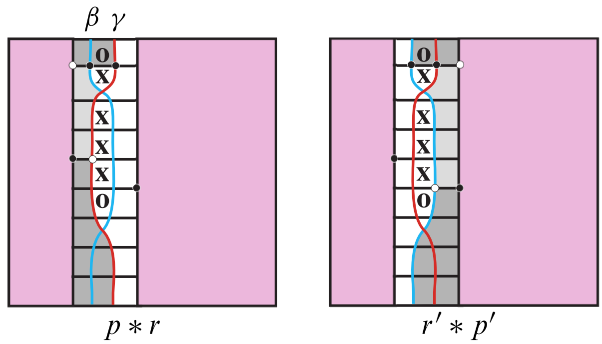

Figure 45: An example of two domains connecting $\boldsymbol{x}$ to $c(\boldsymbol{x})$. The generators $\boldsymbol{x}$ and $c(\boldsymbol{x})$ are both in black circles; they only disagree on one row.

Suppose D is a domain of the form $p * r$ representing a term in $\partial^- \circ \Phi_{\beta,\gamma}(\boldsymbol{x})$ and D connects $\boldsymbol{x}$ to $\boldsymbol{y}$ with $\boldsymbol{y} \neq c(\boldsymbol{x})$. By this, we mean to consider the juxtaposition of the pentagon p connecting $\boldsymbol{x}$ to $\boldsymbol{z}$ and the rectangle r connecting $\boldsymbol{z}$ to $\boldsymbol{y}$, in the combined grid diagram. Note that the domain does not contain any element of $\mathbb{X}$. Then there is exactly one other empty rectangle r' (in g or $\bar{g}$) and empty pentagon p' such that $r' * p'$ or $p' * r'$ gives a decomposition of D. Note that most of the time the other decomposition is of the form $r' * p'$. To see this, one just needs to draw every possible domain of the form $p * r$ and $r * p$, where r is an empty rectangle (in g or $\bar{g}$) and p is an empty pentagon. In addition since p' and r' will be contained in D, they will not contain any element of $\mathbb{X}$ so $r' * p'$ or $p' * r'$ will represent an element of $\Phi_{\beta,\gamma} \circ \partial^-$ or $\partial^- \circ \Phi_{\beta,\gamma}$. The same statement is true if you start with a domain D of the form $r * p$ representing a term in $\partial^- \circ \Phi_{\beta,\gamma}(\boldsymbol{x})$ as long as D connects $\boldsymbol{x}$ to $\boldsymbol{y}$ with $\boldsymbol{y} \neq c(\boldsymbol{x})$.

Suppose D is a domain of the form $p * r$ representing a term in $\partial^- \circ \Phi_{\beta,\gamma}(\boldsymbol{x})$ and D connects $\boldsymbol{x}$ to $c(\boldsymbol{x})$. Then D consists of two regions C and E, where C is the region that lies to the left of both γ and β and to the right of β_{n-1}, and E is a subset of the region that lies to the right of β and to the left of γ. Note that the $C \cap \mathbb{O} = C \cap \mathbb{X} = E \cap \mathbb{X} = \varnothing$. There is exactly one other domain D' connecting $\boldsymbol{x}$ to $c(\boldsymbol{x})$ of the form $p' * r'$ or $r' * p'$. See Figure 45 for an example. This domain consists of a region C' and E, where C' is the region that lies to the right of both γ and β and to the left of β_1. Note that $C' \cap \mathbb{X} = C' \cap \mathbb{X} = \varnothing$ and so $\mathrm{O}_i(D) = \mathrm{O}_i(D')$ and $D' \cap \mathbb{X} = \varnothing$. Thus D' represents an element of $\Phi_{\beta,\gamma} \circ \partial^-(\boldsymbol{x})$ or $\partial^- \circ \Phi_{\beta,\gamma}(\boldsymbol{x})$. A similar statement holds for domain of the form $p * r$ representing a term in $\partial^- \circ \Phi_{\beta,\gamma}(\boldsymbol{x})$ and connecting $\boldsymbol{x}$ to $c(\boldsymbol{x})$. Thus, every term is canceled by another. So $\partial^- \circ \Phi_{\beta\gamma}(\boldsymbol{x}) = \Phi_{\beta\gamma} \circ \partial^-(\boldsymbol{x})$. ◻

In order to prove that $\Phi_{\beta\gamma}$ is a chain homotopy equivalence we define a similar map $H_{\beta\gamma\beta}$ which counts hexagons in the combined grid diagram. These hexagons are just like the ones in [12] except they don't contain elements of $\mathbb{X}$. We recall the definition. For $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{S}(g)$, let $\mathrm{Hex}_{\beta\gamma\beta}(\boldsymbol{x}, \boldsymbol{y})$ be the set of embedded hexagons defined as follows. If $\boldsymbol{x}$ and $\boldsymbol{y}$ don't coincide at $n-2$ points then $\mathrm{Hex}_{\beta\gamma\beta}(\boldsymbol{x}, \boldsymbol{y})$ is the empty set. Suppose that $\boldsymbol{x}$ and $\boldsymbol{y}$ coincide at $n-2$ points (say $x_3 = y_3, \cdots, x_n = y_n$). Without loss of generality, let $x_2 = \boldsymbol{x} \cap \beta$ and $y_2 = \boldsymbol{y} \cap \gamma$. An element $\mathcal{H} \in \mathrm{Hex}_{\beta\gamma\beta}(\boldsymbol{x}, \boldsymbol{y})$ is an embedded disk in the combined grid diagram, whose boundary consists of six arcs, each of which are contained in the circles β_i, α_i, β or γ and satisfies the following conditions. The intersections of the arcs lie on the points of x_1, x_2, y_1, y_2, a and b. Moreover, start at the point in x_2 and transverse the boundary of $\mathcal{H}$, using the orientation given by $\mathcal{H}$. The condition to be in $\mathrm{Hex}_{\beta\gamma\beta}(\boldsymbol{x}, \boldsymbol{y})$ is that you will first travel along a horizontal circle, meet y_1, proceed along a vertical circle β_i, meet x_1, continue along another horizontal circle, meet y_2, proceed though an arc in β until you meet b, then travel along an arc in γ until you hit a and finally travel along an arc in β until arriving back at x_2. Finally, all angles are required to be less than straight. For $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{S}(\bar{g})$, there is a corresponding set of hexagons $\mathrm{Hex}_{\gamma\beta\gamma}(\boldsymbol{x}, \boldsymbol{y})$. The set of empty pentagons $\mathrm{Hex}^{o}_{\beta\gamma\beta}$ are those hexagons $q \in \mathrm{Hex}_{\beta\gamma\beta}(\boldsymbol{x}, \boldsymbol{y})$ where $\boldsymbol{x} \cap \mathrm{Int}(q) = \varnothing$. Define $H_{\beta\gamma\beta}\colon C^-(g) \to C^-(g)$ by

$$H_{\beta\gamma\beta}(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \boldsymbol{S}(g)} \sum_{\substack{q \in \mathrm{Hex}^{o}_{\beta\gamma\beta}(\boldsymbol{x}, \boldsymbol{y}) \\ \mathrm{Int}(q) \cap \mathbb{X} = \varnothing}} U_1^{O_1(q)} \cdots U_n^{O_n(q)} \cdot \boldsymbol{y}.$$

Proposition 5.3 *The map $\Phi_{\beta\gamma}\colon C^-(g) \to C^-(\bar{g})$ is a chain homotopy equivalence.*

Proof To prove this, we show that

$$\mathrm{Id} + \Phi_{\gamma\beta} \circ \Phi_{\beta\gamma} + \partial^- \circ H_{\beta\gamma\beta} + H_{\beta\gamma\beta} \circ \partial^- = 0$$

and

$$\mathrm{Id} + \Phi_{\beta\gamma} \circ \Phi_{\gamma\beta} + \partial^- \circ H_{\gamma\beta\gamma} + H_{\gamma\beta\gamma} \circ \partial^- = 0.$$

The proof is similar to the proof of [12, Proposition 3.2]. Let $\boldsymbol{x} \in S(g)$. Typically, every domain that arises as the composition of two empty pentagons or an empty hexagon and an empty rectangle representing terms from $\Phi_{\gamma\beta} \circ \Phi_{\beta\gamma}(\boldsymbol{x})$, $\partial^- \circ H_{\beta\gamma\beta}(\boldsymbol{x})$ or $H_{\beta\gamma\beta} \circ \partial^-(\boldsymbol{x})$ can be decomposed in exactly two ways representing terms from $\Phi_{\gamma\beta} \circ \Phi_{\beta\gamma}(\boldsymbol{x})$, $\partial^- \circ H_{\beta\gamma\beta}(\boldsymbol{x})$ or $H_{\beta\gamma\beta} \circ \partial^-(\boldsymbol{x})$. The only case when this does not happen is when the domain connects $\boldsymbol{x}$ to $\boldsymbol{x}$. In this case, the domain consists of the region that is either (1) to the left of both γ and β and to the right of β_{n-1} or (2) to the right of both γ and β and to the left of β_1. Such a domain can be decomposed in three ways representing terms in $\Phi_{\gamma\beta} \circ \Phi_{\beta\gamma}(\boldsymbol{x})$, $\partial^- \circ H_{\beta\gamma\beta}(\boldsymbol{x})$ or $H_{\beta\gamma\beta} \circ \partial^-(\boldsymbol{x})$. The other case follows similarly. □

Remark 5.4 $H_{\beta\gamma\beta}$ is a $(H_1(E(f)),\mathbb{Z})$–bigraded R_V–module homomorphism of degree $(0,1)$. Therefore $(C(g),\partial)$ and $(C(\bar{g}),\partial)$ are chain homotopy equivalent as bigraded R_V–module chain complexes.

5.2 Stabilization′ invariance

Proposition 5.5 *Suppose g and $\bar{g}$ are saturated graph grid diagrams that differ by a stabilization′ move. Let $f\colon G\to S^3$ be the transverse graph associated to g and $\bar{g}$ and V be the number of vertices of G. Then there is an $(H_1(E(f)),\mathbb{Z})$–bigraded R_V–module quasi-isomorphism $(C(g),\partial_g^-)\to(C(\bar{g}),\partial_{\bar{g}}^-)$ of degree $(\delta(g,\bar{g}),0)$ for some $\delta(g,\bar{g})\in H_1(E(f))$.*

The proof of Proposition 5.5 will take up the rest of this section. The proof of stabilization′ is similar to the proof of stabilization in [12]. However, because of the fact that we only have a graded theory instead of a filtered theory, the proof becomes drastically simplified. We also fill in some of the details and clarify some of the arguments in the proof of [12]. We will only prove the case for row stabilization′. The proof of column stabilization′ is similar.

Let g be a graph grid diagram and $\bar{g}$ be obtained from g by a row stabilization′ move. An example of a row stabilization′ move is shown in Figure 46. To get $\bar{g}$, we take some row $\boldsymbol{R}$ in g with l X's, delete it and replace it with two new rows and then add a new column. We place $\mathrm{O}_k,\mathrm{X}_{j_2},\dots,\mathrm{X}_{j_l}$ into one of the new rows (and in the same columns as before) and X_{j_1} into the other new row (and in the same column as before). We place decorations O_n and X_m into the new column so that O_n occupies the same row as X_{j_1}, and X_m occupies the same row as O_k. By Remark 3.4, we may assume that X_{j_1}, X_m and O_n share a corner, called $\star$, where X_{j_1} is directly to the left of O_n, and O_n is directly above X_m; see Figure 46. Let β_n be the vertical grid circle directly to the left of O_n and let β_1 be the vertical grid circle directly to the right of O_n. Let α_n be the horizontal grid circle to between O_n and X_m.

Let $(B,\partial_B)=(C^-(g),\partial_g^-)$ and $(C,\partial_C)=(C^-(\bar{g}),\partial_{\bar{g}}^-)$. Let $(B[U_n],\partial_B)$ be the chain complex obtained as follows. $B[U_n]$ is the free (left) R_n–module generated by $\boldsymbol{S}(g)$, and ∂_B is the unique extension of ∂_g to $B[U_n]$ so that ∂_B is an R_n–module homomorphism. We note that $(B[U_n],\partial_B)$ is isomorphic to the chain complex whose group is $B\otimes_{R_{n-1}}R_n$ and whose boundary map is $\partial_B\otimes\mathrm{id}$. $(B[U_n],\partial_B)$ becomes an $(H_1(E(f)),\mathbb{Z})$–bigraded R_n–module chain complex by setting the $H_1(E(f))$–grading of U_n to be $w(\mathrm{X}_{j_1})$ and the $\mathbb{Z}$–grading of U_n to be -2.

Definition 5.6 Let

$$\zeta=\begin{cases}U_n+U_k & \text{if } l=1,\\ U_n & \text{if } l\geq 2,\end{cases}$$

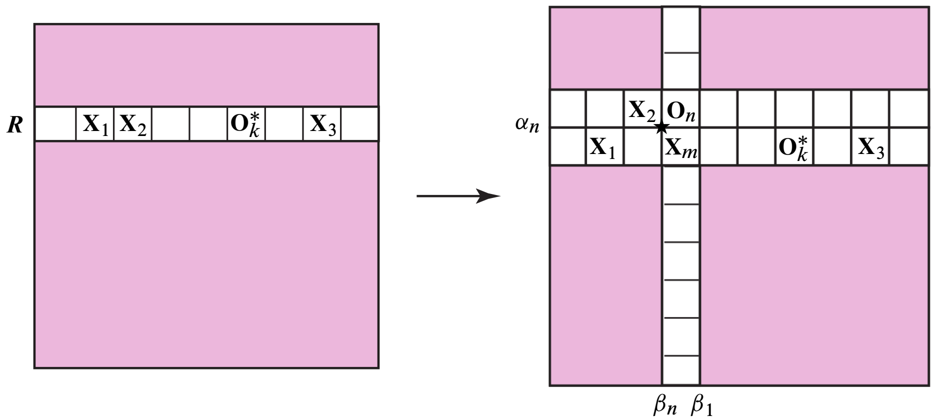

Figure 46: An example of stabilization′

so that ζ: $B[U_n] \to B[U_n]$ is a bigraded R_n–module chain map of degree $(-w(\mathrm{O}_n), -2)$. Let (C', ∂') be the mapping cone complex of ζ. Since $(B[U_n], \partial_B)$ is a $(H_1(E(f)), \mathbb{Z})$–bigraded R_n–module chain complexes, the $(\mathrm{cone}(\zeta), \partial')$ is a $(H_1(E(f)), \mathbb{Z})$–bigraded R_n–module chain complex. See page 1490 for the definition of the mapping cone and its grading.

We will first show that C' is quasi-isomorphic to B and then we will show that C is quasi-isomorphic to C'. The first step follows from basic facts from homological algebra. Consider the cokernel of ζ, $B[U_n]/\mathrm{im}(\zeta)$. The $(H_1(E(f)), \mathbb{Z})$–bigrading on $C^-(g)$ descends to a well-defined $(H_1(E(f)), \mathbb{Z})$–bigrading on $B[U_n]/\mathrm{im}(\zeta)$ and so $(B[U_n]/\mathrm{im}(\zeta), \partial_B)$ is an $(H_1(E(f)), \mathbb{Z})$–bigraded R_n–module chain complex. In addition, using the inclusion $R_{n-1} \subset R_n$, it is also naturally an $(H_1(E(f)), \mathbb{Z})$–bigraded R_{n-1}–module chain complex. Moreover the map

$$B \xrightarrow{\cong} B[U_n]/\mathrm{im}(\zeta),$$

which sends $b \in B$ to the equivalence class of itself, is a bigraded R_{n-1}–module chain isomorphism of degree $(0,0)$. Thus, we just need to show that C' is quasi-isomorphic to $B/\mathrm{im}(\zeta)$.

Lemma 5.7 *Let* pr: $B[U_n] \to B[U_n]/\mathrm{im}(\zeta)$ *be the quotient map. The map from* C' *to* $B[U_n]/\mathrm{im}(\zeta)$ *that sends* (a,b) *to* $\mathrm{pr}(b)$ *is a bigraded* R_n*–module quasi-isomorphism of degree* $(0,0)$.

Proof There is a short exact sequence of chain complexes

$$0 \to B[U_n] \xrightarrow{\zeta} B[U_n] \xrightarrow{\mathrm{pr}} B[U_n]/\mathrm{im}(\zeta) \to 0.$$

Therefore, by [23, Section 1.5.8], the map $\mathrm{cone}(\zeta) \to B[U_n]/\mathrm{im}(\zeta)$ sending (a, b) to $\mathrm{pr}(b)$ is a quasi-isomorphism. This map is bigraded of degree $(0, 0)$ and an R_n–module homomorphism. $\square$

We now define a quasi-isomorphism $F\colon C \to C'$ similar to the one in [12]. However, since we are only considering a bigraded chain complex instead of a filtered chain complex and we only need to consider one of the four stabilization′s, our map becomes very simple. We first consider some notation.

Let $\boldsymbol{I} \subset \boldsymbol{S}(\bar{g})$ be the set of $\boldsymbol{x} \in \boldsymbol{S}(\bar{g})$ that contain $\star$: the intersection of the new grid lines/circles α_n and β_n. There is a natural one-to-one correspondence between $\boldsymbol{I}$ and $\boldsymbol{S}(g)$. For $\boldsymbol{x} \in \boldsymbol{S}(g)$ let $\psi(\boldsymbol{x})$ be the point in $\boldsymbol{I}$ defined by $\boldsymbol{x} \cup \star$. Note that if $\boldsymbol{x} \in \boldsymbol{I}$ then $\boldsymbol{x} = \{x_1, x_2, \dots, x_{n-1}, \star\}$ so $\psi^{-1}(\boldsymbol{x}) = \{x_1, \dots, x_{n-1}\}$ is the generator of B obtained by removing $\star$. The gradings of $\boldsymbol{x} \in \boldsymbol{S}(g)$ and $\psi(\boldsymbol{x})$ are related as follows:

$$M^g(\boldsymbol{x}) = M^{C'}(\boldsymbol{x}, 0) + 1 = M^{C'}(0, \boldsymbol{x}) = M^{\bar{g}}(\psi(\boldsymbol{x})) + 1, \tag{7}$$

$$A^g(\boldsymbol{x}) = A^{C'}(\boldsymbol{x}, 0) + w(\mathrm{O}_n) = A^{C'}(0, \boldsymbol{x}) = A^{\bar{g}}(\psi(\boldsymbol{x})) - A^{\bar{g}}(\star). \tag{8}$$

Define $F_L\colon C \to B[U_n]$ by

$$F_L(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \boldsymbol{x} \notin \boldsymbol{I} \\ \psi^{-1}(\boldsymbol{x}) & \text{if } \boldsymbol{x} \in \boldsymbol{I} \end{cases}$$

and extend it to C so that it is an R_n–module homomorphism. The reason for this choice of map is that the trivial region is the only type-L region in [12] that doesn't contain X_{j_1} when X_{j_1} is directly to the left of O_n. For the definition of an type-L region, see [12, Definition 3.4 and Figure 13].

There is one type-R region that does not contain X_{j_1}: the rectangle with $\star$ in its upper left corner. For $\boldsymbol{x} \in S(\bar{g})$ and $\boldsymbol{y} \in S(g)$, let $\pi_R(\boldsymbol{x}, \psi(\boldsymbol{y}))$ be the set of $p \in \mathrm{Rect}^o(\boldsymbol{x}, \psi(\boldsymbol{y}))$ whose upper left corner is $\star$; see Figure 47. We will call such a domain an *R–domain*. Define $F_R\colon C \to B[U_n]$ by

$$F_R(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in S(g)} \sum_{\substack{p \in \pi_R(\boldsymbol{x}, \psi(\boldsymbol{y})) \\ (\mathbb{X} \setminus \mathrm{X}_m) \cap \mathrm{Int}(p) = \varnothing}} U_1^{\mathrm{O}_1(p)} \cdots U_{n-1}^{\mathrm{O}_{n-1}(p)} \boldsymbol{y}.$$

Note that we are counting domains in $\bar{g}$ that cannot contain $\mathrm{X}_1, \dots, \mathrm{X}_{m-1}$ but can contain X_m. Also, there is no factor of U_n in the terms of $F_R(\boldsymbol{x})$. Using these, we define $F\colon C \to C'$ by

$$F(\boldsymbol{x}) = (F_L(\boldsymbol{x}), F_R(\boldsymbol{x})).$$

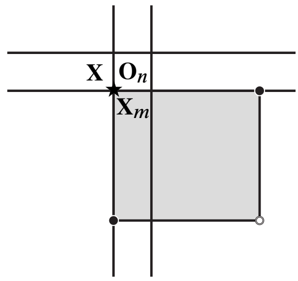

Figure 47: A domain in $\pi_R(\boldsymbol{x}, \psi(\boldsymbol{y}))$; the points of $\boldsymbol{x}$ are in black, the points of $\boldsymbol{y}$ are in white, the domain is shaded

Lemma 5.8 *$F\colon C \to C'$ is a $(H_1(E(f)), \mathbb{Z})$–bigraded R_n–module chain map of degree $(-w(\mathrm{X}_m) - A^{\bar{g}}(\star), 0)$.*

Proof F is an R_n–module homomorphism by definition. It follows from [12, Lemma 3.5] that the Maslov grading is preserved.

***F* is an $H_1(E(f))$–graded map of degree $-A^{\bar{g}}(\star) - w(\mathrm{O}_n)$** Let $\boldsymbol{x} \in \boldsymbol{S}(\bar{g})$. We first consider the grading of $(F_L(\boldsymbol{x}), 0)$. $F_L(\boldsymbol{x})$ is either 0 or is $\psi^{-1}(\boldsymbol{x})$. Suppose $F_L(\boldsymbol{x}) \neq 0$. Then $F_L(\boldsymbol{x}) = \psi^{-1}(\boldsymbol{x})$, so by (8),

$$A^{C'}(F_L(\boldsymbol{x}), 0) = A^g(\psi^{-1}(\boldsymbol{x})) - w(\mathrm{O}_n) = A^{\bar{g}}(\boldsymbol{x}) - A^{\bar{g}}(\star) - w(\mathrm{O}_n).$$

We now consider the grading of $(0, F_R(\boldsymbol{x}))$. Let $U_1^{\mathrm{O}_1(p)} \cdots U_{n-1}^{\mathrm{O}_{n-1}(p)}\boldsymbol{y}$ be a term in $F_R(\boldsymbol{x})$. Then p is a rectangle connecting $\boldsymbol{x}$ to $\psi(y)$ that doesn't contain any elements of $\mathbb{X}$ except X_m. Moreover, p does not contain O_n and contains X_m exactly once. So by Lemma 4.13,

$$A^{\bar{g}}(\boldsymbol{x}) - A^{\bar{g}}(\psi(\boldsymbol{y})) = w_{\mathbb{X}}(p) - w_{\mathbb{O}}(p) = w(\mathrm{X}_m) - \sum_{i=1}^{n-1} \mathrm{O}_i(p) w(\mathrm{O}_i).$$

Therefore, by (8), we have

$$\begin{aligned} A^{C'}(0, U_1^{\mathrm{O}_1(p)} \cdots U_{n-1}^{\mathrm{O}_{n-1}(p)}\boldsymbol{y}) &= A^{C'}(0, \boldsymbol{y}) - \sum_{i=1}^{n-1} \mathrm{O}_i(p) w(\mathrm{O}_i) \\ &= A^{\bar{g}}(\psi(\boldsymbol{y})) - A^{\bar{g}}(\star) - \sum_{i=1}^{n-1} \mathrm{O}_i(p) w(\mathrm{O}_i) \\ &= A^{\bar{g}}(\boldsymbol{x}) - w(\mathrm{X}_m) - A^{\bar{g}}(\star) \\ &= A^{\bar{g}}(\boldsymbol{x}) - w(\mathrm{O}_n) - A^{\bar{g}}(\star). \end{aligned}$$

***F* is a chain map** Let $\boldsymbol{x} \in \boldsymbol{S}(\bar{g})$. Recall that

$$\partial'(F(\boldsymbol{x})) = (\partial_B(F_L(\boldsymbol{x})), 0) + (0, \zeta(\boldsymbol{x})) + (0, \partial_B(F_R(\boldsymbol{x}))$$

and

$$F(\partial_C(\boldsymbol{x})) = (F_L(\partial_C(\boldsymbol{x})), 0) + (0, F_{\boldsymbol{R}}(\partial_C(\boldsymbol{x})).$$

The proof that F is a chain map is similar to the proof that the commutation$'$ map $\Phi_{\beta,\gamma}$ is a chain map.

For this proof, whenever we have an empty rectangle r in g (contributing to a term in $\partial_{\boldsymbol{B}}$), we will view it as living in $\bar{g}$ in the obvious way. Note that such a rectangle cannot have a boundary on α_n or β_n and hence cannot have $\star$ on its boundary. Usually (in a filtered theory), such a rectangle could contain the point $\star$ in its interior. However, since r cannot contain X_{j_1}, it also cannot contain $\star$ in its interior.

We first show that $\partial_{\boldsymbol{B}}(F_L(\boldsymbol{x})) = F_L(\partial_C(\boldsymbol{x}))$. Suppose D is a domain of the form $p * r$ representing a nontrivial term in $\partial_{\boldsymbol{B}} \circ F_L(\boldsymbol{x})$. Then p is a trivial domain connecting a point in $\boldsymbol{I}$ to itself and r is an empty rectangle in g. We will think of p as a point at $\star$. Since $\star$ cannot be on the corner of r or in its interior, r and p must be disjoint. Suppose D is a domain of the form $r' * p'$ representing a term in $F_L \circ \partial_C(\boldsymbol{x})$. Then p' is a trivial domain connecting a point in $\boldsymbol{I}$ to itself and r is an empty rectangle in $\bar{g}$. Since r' is empty, $\star$ is not in the interior of p'. Since r' cannot contain X_m or X_{j_1}, it cannot have $\star$ as one of its corners. Thus r' and p' are disjoint. Therefore the terms in $(\partial_{\boldsymbol{B}}(F_L(\boldsymbol{x})), 0)$ and $(F_L(\partial_C(\boldsymbol{x})), 0)$ cancel each other out.

We now wish to show that $\partial_{\boldsymbol{B}}(F_{\boldsymbol{R}}(\boldsymbol{x})) + \zeta(\boldsymbol{x}) = F_{\boldsymbol{R}}(\partial_C(\boldsymbol{x})$. Suppose D is a domain of the form $p * r$ representing a nontrivial term in $\partial_{\boldsymbol{B}} \circ F_{\boldsymbol{R}}(\boldsymbol{x})$. Then p is an empty rectangle in $\bar{g}$ with $\star$ on its upper left corner and r is an empty rectangle in g. If p and r share zero corners or share one corner then there is exactly one empty rectangle r' in $\bar{g}$ and R–domain p such that $r' * p'$ gives a decomposition of D (p' and r' will also share zero corners or share one corner). These represent terms in $F_{\boldsymbol{R}} \circ \partial_C(\boldsymbol{x})$. Note that p and r cannot share two or three corners. Conversely, suppose D is a domain of the form $p' * r'$ representing a nontrivial term in $F_{\boldsymbol{R}} \circ \partial_C(\boldsymbol{x})$. Then p' is an empty rectangle in $\bar{g}$ with $\star$ on its upper left corner and r' is an empty rectangle in $\bar{g}$. If p' and r' share no corners or share one corner then there is exactly one other empty rectangle r'' (in g or $\bar{g}$) and R–domain p'' such that $r'' * p''$ or $p'' * r''$ gives a decomposition of D. Here p'' and r'' will also share zero corners or share one corner. These will represent terms in $\partial_{\boldsymbol{B}} \circ F_{\boldsymbol{R}}(\boldsymbol{x})$ or $F_{\boldsymbol{R}} \circ \partial_C(\boldsymbol{x})$. Note that p' and r' cannot share two or three corners. Thus, these terms cancel one another out.

First note that if D is a domain of the form $p * r$ representing a nontrivial term in $\partial_{\boldsymbol{B}} \circ F_{\boldsymbol{R}}(\boldsymbol{x})$, then p and r cannot share more than one corner (hence cannot share four corners). Suppose D is a domain of the form $p' * r'$ representing a nontrivial term in $F_{\boldsymbol{R}} \circ \partial_C(\boldsymbol{x})$ where p' and r' share four corners. Then D is either the width-one horizontal annulus containing X_m or the width-one vertical annulus containing O_n.

The width-one vertical annulus always contributes $U_n\boldsymbol{x}$. If there is more than one element of $\mathbb{X}$ in row $\boldsymbol{R}$ $(l \geq 2)$ then the width-one horizontal annulus contains an element of $\mathbb{X}$ that is not X_m, so is not counted. If there is exactly one element of $\mathbb{X}$ in row $\boldsymbol{R}$ $(l = 1)$ then the width-one horizontal annulus containing X_m contributes $U_k\boldsymbol{x}$. Thus these terms cancel with $\zeta(\boldsymbol{x})$. □

To show that F is a quasi-isomorphism, we will first show that $\widetilde{F}\colon \widetilde{C} \to \widetilde{C}'$ is a quasi-isomorphism, where $\widetilde{C}$ is the quotient $C/\mathcal{U}_n$ defined in Section 4.4. To do this, we introduce a filtration on $\widetilde{C}$ and $\widetilde{C}'$ so that $\widetilde{F}$ is a filtered map, and show $\widetilde{F}$ induces a quasi-isomorphism on its associated graded object. The rest of the proof will follow from the following well-known lemma.

Lemma 5.9 [14, Theorem 3.2] *Suppose that $F\colon C \to C'$ is a filtered chain map that induces an isomorphism on the homology of the associated graded object. Then F is a filtered quasi-isomorphism.*

The definition of the filtration and the proof that $\widetilde{F}$ induces a quasi-isomorphism on its associated graded object is essentially the same as in [12]. However, there is a small mistake in their definition of the filtration which we fix. We also give more details which we believe clarifies their proof.

For any $\mathbb{F}[U_1, \dots, U_n]$–module chain complex (C, ∂), one can define the chain complex $(\widetilde{C}, \widetilde{\partial})$ like we did in Section 4.4. Let $\mathcal{U}_V$ be the $\mathbb{F}$–vector subspace of C spanned by $U_1C \cup \cdots \cup U_nC$. Define $\widetilde{C}$ to be the quotient $C/\mathcal{U}_n$. Since $\partial^-(\mathcal{U}_n) \subset \mathcal{U}_n$, it follows that ∂ descends to a linear map

$$\widetilde{\partial}\colon \widetilde{C} \to \widetilde{C}$$

of vector spaces over $\mathbb{F}$.

Define the Q–filtration $\mathcal{F}_k^Q(C)$ on (C, ∂_C), where $C = C^-(g)$, as follows. Let Q be the collection of $(n-1)^2$ dots in $\bar{g}$, with one dot placed in each square which does not appear in the row or column containing O_n. For a domain $p \in \pi(\boldsymbol{x}, \boldsymbol{y})$, let $\mathbb{O}(p)$ be the total number of O's in p counted with sign. That is, $\mathbb{O}(p) = \sum_{i=1}^n \mathrm{O}_i(p)$. Similarly, we define $Q(p)$ to be the total number of dots in p counted with sign. Here, we are viewing the points in $\mathbb{O}$ and Q as having positive orientation. Note with this convention, if r is a rectangle connecting $\boldsymbol{x}$ to $\boldsymbol{y}$, then $Q(r) \geq 0$ and $\mathbb{O}(r) \geq 0$.

Lemma 5.10 *Let p and p' be domains in $\bar{g}$ connecting $\boldsymbol{x}$ to $\boldsymbol{y}$ such that*

$$\mathrm{O}_n(p) = \mathrm{O}_n(p') = 0 = \mathbb{O}(p) = \mathbb{O}(p').$$

Then $Q(p) = Q(p')$.

Proof For $1 \leq i \leq n$, let $\boldsymbol{R}_i \in \pi(\boldsymbol{x},\boldsymbol{x})$ (respectively $\mathbb{C}_i$) be the domain that is the positively oriented row (respectively column) contain O_i. Note that $\mathrm{O}_k(\boldsymbol{R}_i) = \mathrm{O}_k(\boldsymbol{C}_i) = \delta_{ik}$. Suppose $p, p' \in \pi(\boldsymbol{x},\boldsymbol{y})$ then p and p' differ by a domain in $\pi(\boldsymbol{x},\boldsymbol{x})$ so that

$$p' = p + \sum_{i=1}^{n} (a_i \boldsymbol{R}_i + b_i \boldsymbol{C}_i). \tag{9}$$

This follows from the fact that the space of domains on the torus of the form $\pi(\boldsymbol{x},\boldsymbol{x})$ is generated by $\boldsymbol{R}_i$ and $\boldsymbol{C}_i$ with the relation that $\sum_{i=1}^{n}(\boldsymbol{R}_i - \boldsymbol{C}_i) = 0$.

Now, we note that $Q(\boldsymbol{R}_i) = Q(\boldsymbol{C}_i) = n-1$ for $i \neq n$ and $Q(\boldsymbol{R}_n) = Q(\boldsymbol{C}_n) = 0$. Thus

$$Q(p') = Q(p) + (n-1)\sum_{i=1}^{n-1}(a_i + b_i).$$

Since $\mathrm{O}_n(p) = \mathrm{O}_n(p') = 0$ by hypothesis, using (9) we get

$$0 = \mathrm{O}_n(p') = \mathrm{O}_n(p) + \mathrm{O}_n\biggl(\sum_{i=1}^{n}(a_i \boldsymbol{R}_i + b_i \boldsymbol{C}_i)\biggr) = a_n + b_n.$$

Similarly since $\mathbb{O}(p) = \mathbb{O}(p') = 0$,

$$0 = \mathbb{O}\biggl(\sum_{i=1}^{n}(a_i \boldsymbol{R}_i + b_i \boldsymbol{C}_i)\biggr) = \sum_{i=1}^{n}(a_i + b_i).$$

Using these three equalities, we have that $Q(p) = Q(p')$. □

Lemma 5.11 *Let $\boldsymbol{x},\boldsymbol{y} \in \boldsymbol{S}(\bar{g})$, then there is a domain $p \in \pi(\boldsymbol{x},\boldsymbol{y})$ with $\mathrm{O}_n(p) = \mathbb{O}(p) = 0$.*

Proof Let $\boldsymbol{x},\boldsymbol{y} \in \boldsymbol{S}(\bar{g})$. First we note that there is domain connecting $\boldsymbol{x}$ to $\boldsymbol{y}$. Since S_n is generated by transpositions, there is a sequence of rectangles connecting $\boldsymbol{x}$ to $\boldsymbol{y}$ (not necessarily empty). The sum of these rectangles is a domain p_0 connecting $\boldsymbol{x}$ to $\boldsymbol{y}$.

Let $m_0 = \mathrm{O}_n(p)$, which is not necessarily zero. We replace each rectangle containing O_n with the other rectangle connecting its corners as follows. Let

$$p_1 = p_0 - m_0\biggl(\boldsymbol{R}_n + \sum_{i=1}^{n-1} C_i\biggr).$$

Then p_1 connects $\boldsymbol{x}$ to $\boldsymbol{y}$ since $\boldsymbol{R}_n + \sum_{i=1}^{n-1} \boldsymbol{C}_i$ is periodic. Moreover, we have $\mathrm{O}_n(\boldsymbol{R}_n + \sum_{i=1}^{n-1} \boldsymbol{C}_i) = 1$ so that $\mathrm{O}_n(p_1) = 0$. (We could also let $p_1 = p_0 - m_0 \boldsymbol{R}_n$ or $p_1 = p_0 - m_0 \boldsymbol{C}_n$.) Now let $m_1 = \mathbb{O}(p_1)$ and define $p_2 = p_1 - m_1 \boldsymbol{R}_1$. Then p_2 connects $\boldsymbol{x}$ to $\boldsymbol{y}$ and as $\mathrm{O}_n(\boldsymbol{R}_1) = 0$ and $\mathbb{O}(\boldsymbol{R}_1) = 1$, $\mathrm{O}_n(p_2) = \mathbb{O}(p_2) = 0$. □

We use this to define a function $\boldsymbol{F}^Q\colon \boldsymbol{S}(\bar{g}) \to \mathbb{Z}$ by first defining $\boldsymbol{F}^Q(\boldsymbol{x}_0) = 0$, where $\boldsymbol{x}_0$ is the lower left corner of the O's. (It doesn't really matter what value we choose.) Then for $\boldsymbol{x} \in \boldsymbol{S}(\bar{g})$, use Lemma 5.11 to pick a domain $p_{\boldsymbol{x}}$ connecting $\boldsymbol{x}$ to $\boldsymbol{x}_0$ with $\mathrm{O}_n(p_{\boldsymbol{x}}) = \mathbb{O}(p_{\boldsymbol{x}}) = 0$. Define

$$\boldsymbol{F}^Q(\boldsymbol{x}) = \boldsymbol{F}^Q(x_0) + Q(p_{\boldsymbol{x}}) = Q(p_{\boldsymbol{x}}).$$

This is well defined by Lemma 5.10.

Lemma 5.12 *Suppose that* $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{S}(\bar{g})$ *and* p *is any domain connecting* $\boldsymbol{x}$ *to* $\boldsymbol{y}$ *with* $\mathrm{O}_n(p) = \mathbb{O}(p) = 0$. *Then*

$$\boldsymbol{F}^Q(\boldsymbol{x}) - \boldsymbol{F}^Q(\boldsymbol{y}) = Q(p).$$

Proof Let $p_{\boldsymbol{x}}$ be a domain connecting $\boldsymbol{x}$ to $\boldsymbol{x}_0$ with $\mathrm{O}_n(p_{\boldsymbol{x}}) = \mathbb{O}(p_{\boldsymbol{x}}) = 0$. Define $p_{\boldsymbol{y}}$ similarly. Then $p_{\boldsymbol{x}} - p_{\boldsymbol{y}}$ is a domain connecting $\boldsymbol{x}$ to $\boldsymbol{y}$ with $\mathrm{O}_n(p_{\boldsymbol{x}} - p_{\boldsymbol{y}}) = \mathbb{O}(p_{\boldsymbol{x}} - p_{\boldsymbol{y}}) = 0$. So by Lemma 5.10, $\boldsymbol{F}^Q(\boldsymbol{x}) - \boldsymbol{F}^Q(\boldsymbol{y}) = Q(p_{\boldsymbol{x}}) - Q(p_{\boldsymbol{y}}) = Q(p_{\boldsymbol{x}} - p_{\boldsymbol{y}}) = Q(p)$. □

We now use $\boldsymbol{F}^Q$ to define the Q–filtration on $\widetilde{C}$ by

$$\mathcal{F}_p^Q(\widetilde{C}) = \Big\{ \sum b(\boldsymbol{x})\boldsymbol{x} \in \widetilde{C} \;\Big|\; \boldsymbol{F}^Q(\boldsymbol{x}) \le p \text{ whenever } b(\boldsymbol{x}) \ne 0 \Big\}.$$

By Lemma 5.12, $\widetilde{\partial}_C\colon \mathcal{F}_p^Q(\widetilde{C}) \to \mathcal{F}_p^Q(\widetilde{C})$ so that $(\widetilde{C}, \widetilde{\partial}_C)$ becomes a $\mathbb{Z}$–filtered chain complex.

Note that we have already shown that F is a $(H_1(E(f)), \mathbb{Z})$–bigraded R_n–module chain map. Thus, the following proposition will complete the proof that two grids that differ by a stabilization$'$ move are quasi-isomorphic as $(H_1(E(f)), \mathbb{Z})$–bigraded R_n–module chain complexes.

Proposition 5.13 $F\colon C \to C'$ *is a quasi-isomorphism.*

Proof Since C' is an R_n–module chain complex, we can consider the chain complex $(\widetilde{C}', \widetilde{\partial}')$ obtained by setting all the U_i equal to zero as explained above, $\widetilde{C}' = C'/\mathcal{U}_n C'$. Since F is an R_n–module homomorphism and a chain map, it descends to a well-defined chain map $\widetilde{F}\colon \widetilde{C} \to \widetilde{C}'$.

Now we can define a Q–filtration on $\widetilde{C}'$ using the Q–filtration on $\widetilde{C}$. Specifically, we define $\boldsymbol{F}^Q\colon \widetilde{B} \to \mathbb{Z}$ by $\boldsymbol{F}^Q(\boldsymbol{x}) = \boldsymbol{F}^Q(\psi(\boldsymbol{x}))$ for all $\boldsymbol{x} \in \boldsymbol{S}(g)$. From this we have

$$\mathcal{F}^Q_p(\widetilde{B}) = \Big\{\sum b(\boldsymbol{x})\boldsymbol{x} \in \widetilde{B} \;\Big|\; \boldsymbol{F}^Q(\boldsymbol{x}) \leq p \text{ whenever } b(\boldsymbol{x}) \neq 0\Big\}.$$

One can identify $(\widetilde{C}', \widetilde{\partial}')$ with the mapping cone of $\widetilde{\zeta}\colon \widetilde{B} \to \widetilde{B}$ (this is the direct sum because $\widetilde{\zeta}$ is the zero map). Then we can define the Q–filtration on C' by $\mathcal{F}^Q_p(\widetilde{C}') = \mathcal{F}^Q_p(\widetilde{B}) \oplus \mathcal{F}^Q_p(\widetilde{B})$. It is easy to check that $\widetilde{F}$ preserves this filtration.

Consider the map induced on their associated graded complexes

$$\widetilde{F}_Q\colon \widetilde{C}_Q \to \widetilde{C}'_Q.$$

The first chain complex $\widetilde{C}_Q$ is the $\mathbb{F}$ vector space generated by $\boldsymbol{S}(\bar{g})$ whose boundary map counts rectangles supported in the column and row through O_n that do not contain O_n, X_m or X_{j_1}. Since the column and row through O_n look exactly the same as the column and row through O_1 in [12] (after renumbering), this chain complex is the same as it is for links. In particular, [12, Lemma 3.7] holds in our case. Similarly, like in [12], $\widetilde{C}'_Q$ is the chain complex whose underlying group is $\widetilde{B} \oplus \widetilde{B}$ and whose boundary maps are trivial. Moreover, the map $\widetilde{F}_Q$ is exactly the same as in [12]. Thus, their proof holds in our case to show that $\widetilde{F}_Q$ is a quasi-isomorphism. Now by Lemma 5.9, $\widetilde{F}$ is a quasi-isomorphism.

We remark that one can define a filtration on C so that $(\widetilde{C}, \widetilde{\partial}_C)$ is its associated graded object. Define

$$\mathcal{F}^U_p(C) = \Big\{\sum b(\boldsymbol{x})U_1^{a_1(\boldsymbol{x})} \cdots U_n^{a_n(\boldsymbol{x})}\boldsymbol{x} \in C \;\Big|\; \sum a_i(\boldsymbol{x}) \geq p \text{ whenever } b(\boldsymbol{x}) \neq 0\Big\}.$$

The boundary preserves the filtration making (C, ∂_C) into a filtered chain complex. We can do the same with B, making (B, ∂_B) into a filtered chain complex. Since ζ is a filtered map, we can define a filtration on the mapping cone as before: $\mathcal{F}^U_p(C') = \mathcal{F}^U_p(B) \oplus \mathcal{F}^U_p(B)$. It is easy to see that F is a filtered map. Moreover, the map on the associated graded objects of C and C' induced by F, $F_U\colon C_U \to C'_U$, can be identified with $\widetilde{F}\colon \widetilde{C} \to \widetilde{C}'$. Since $\widetilde{F}$ is a quasi-isomorphism, so is F_U. Therefore, F is a quasi-isomorphism by Lemma 5.9. □

6 The Alexander polynomial and sutured Floer homology

In this section, we will define the Alexander polynomial of a transverse spatial graph $f\colon G \to S^3$ as a torsion invariant of a balanced sutured manifold associated to f and show that it agrees with the graded Euler characteristic of $\widehat{\mathrm{HFG}}(f)$ (when f is sourceless and sinkless and G has no cut edges). In addition, we will relate the sutured

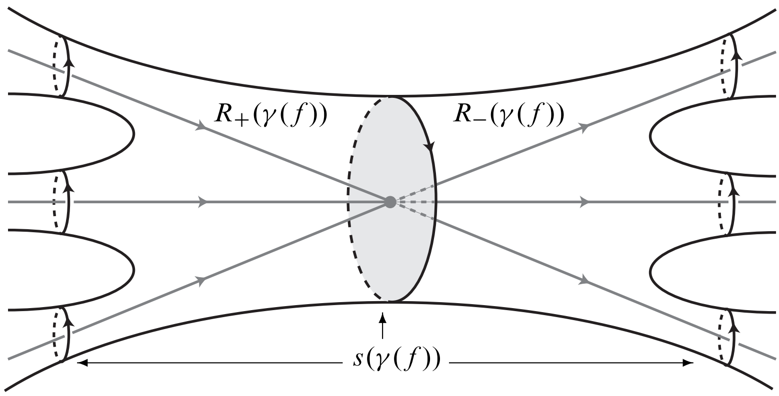

Figure 48: The sutured manifold $(E(f), \gamma(f))$

Floer homology of this balanced sutured manifold to $\widehat{\mathrm{HFG}}(f)$ (when f is sourceless and sinkless and G has no cut edges).

6.1 The Alexander polynomial of a spatial graph

Let $f\colon G \to S^3$ be a transverse spatial graph, and $E(f) = S^3 \setminus N(f(G))$, where $N(f(G))$ is a regular neighborhood of $f(G)$ in S^3. Then $E(f)$ has the structure of a (strongly) balanced sutured manifold $(E(f), \gamma(f))$ which is defined as follows. There will be one suture per edge and one suture per vertex. The suture associated to a vertex is the boundary of the transverse disk at that vertex and the suture associated to an edge is the boundary of a disk transverse to that edge of f. The sutures, denoted by $s(\gamma(f))$, are oriented as shown in Figure 48. This $\gamma(f)$ is a collection of annuli that are small neighborhoods of the sutures in $\partial E(f)$. Recall that $R(\gamma) = \partial E(f) \setminus \operatorname{int}(\gamma)$ is the oriented surface where the orientation of $R(\gamma)$ is such that the induced orientation on each component of $\partial R(\gamma)$ agrees with the orientation of the corresponding suture. Then $R_+(\gamma)$ (respectively $R_-(\gamma)$) is the set of components of $R(\gamma)$ whose normal vectors point out of (respectively into) $E(f)$. Note that $R_+(\gamma)$ is the set of components of $\partial E(f)$ that have the same orientation as $\partial E(f)$. It is easy to check that for each component Σ of $\partial E(f)$, $\chi(\Sigma \cap R_-(\gamma(f))) = \chi(\Sigma \cap R_+(\gamma(f)))$ so that $(E(f), \gamma(f))$ is strongly balanced (in particular, it is balanced). In addition, we note that $\gamma(f)$ contains no toroidal components. Note that we do not need f to be sourceless and sinkless to define this. See [7, Sections 2–3] for the definition of a balanced and strongly balanced sutured manifold.

In [4], S Friedl, A Juhász, and J Rasmussen assign to each balanced sutured manifold (M, γ), a torsion invariant $T(M, \gamma) \in \mathbb{Z}[H_1(M)]$ that is well defined up to $\pm h$ for $h \in H_1(M)$. This invariant is essentially the maximal abelian torsion for the pair

$(M, R_-(\gamma))$. See [4, Sections 3–4] for details. Using their torsion invariant and the sutured manifold associated to a transverse spatial graph defined above, we can define our Alexander polynomial.

Definition 6.1 Let $f\colon G \to S^3$ be a transverse spatial graph. The *(refined) Alexander polynomial* of f, denoted by Δ_f, is defined to be $T(E(f), \gamma(f))$ considered as an element of $\mathbb{Z}[H_1(E(f))]$ modulo units.

Remark 6.2 Others have considered Alexander polynomials of spatial graphs in the past.

(1) The first place this seems to appear is in a 1958 paper by Kinoshita [9]. In this paper, Kinoshita defines the Alexander polynomial of a spatial graph as the Alexander polynomial of its exterior. However, this can be computed using $\pi_1(S^3 \setminus f(G))$ and cannot differentiate between graphs with the same exterior. Our definition depends on more than just the exterior, so it gives more information about the graph than the polynomial defined by Kinoshita.

(2) In 1989, Litherland defined an Alexander polynomial for an embedding of a generalized theta graph that is not determined by its exterior [10]; see also [13]. He considers the Alexander polynomial associated to the torsion free abelian cover of the pair $(S^3 \setminus f(G), R_-)$, where R_- is half of the boundary obtained by cutting open $\partial(S^3 \setminus f(G))$ along the meridians of the edges and throwing away one of the components. We note that, for theta graphs, Litherland's definition and ours are very similar; the main difference is how we decompose $\partial(S^3 \setminus f(G))$. In our case the sutures depend on the orientation of the edges. We note that if all the edges are oriented in the same direction, Δ_f will be zero since $R_-(\gamma)$ will contain a disjoint disk.

(3) If G contains a vertex all of whose edges are incoming or outgoing, then for any transverse spatial graph $f\colon G \to S^3$, $R_-(\gamma)$ will contain a disjoint disk and hence $\Delta_f = 0$.

(4) Like in Litherland's paper, instead of just studying the Alexander polynomial of a transverse spatial graph, one could study the entire Alexander module $\mathbb{Z}[H_1(E(f))]$–module $H_1(E(f), R_-(\gamma(f)))$.

6.2 The sutured Floer homology of a spatial graph

In the preceding subsection, we defined a balanced sutured manifold $(E(f), \gamma(f))$, associated to a transverse spatial graph f. Instead of just considering the torsion of this sutured manifold, we can consider the sutured Floer homology of it. We will describe this in more detail in this subsection. We will also show that this homology theory coincides with our hat theory. We begin with more definitions and background.

X	O
O	X

Figure 49: A graph grid diagram g for the unknot

Associated to each $n \times n$ graph grid diagram g representing f, there is another sutured manifold $(E(f), \gamma(g))$ which is defined as follows. For each X and O on the torus, remove an (open) disk. Then one obtains an oriented torus with $n+r$ disks removed, where n is the size of the grid and r is the number of X's in the grid; call this surface $\Sigma(g)$. Note that the orientation of the torus comes from the standard counterclockwise orientation of the plane. Recall that the horizontal circles of g are called α_i, the vertical circles are called β_j, $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_n\}$ and $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_n\}$. Thus $(\Sigma(g), \boldsymbol{\alpha}, \boldsymbol{\beta})$ gives a sutured Heegaard decomposition associated to g whose underlying manifold is $E(f)$. Let $(E(f), \gamma(g))$ be the sutured manifold associated to $(\Sigma(g), \boldsymbol{\alpha}, \boldsymbol{\beta})$. Recall that $E(f)$ is obtained by from $(\Sigma(g), \boldsymbol{\alpha}, \boldsymbol{\beta})$ by attaching 3–dimensional 2–handles to $\Sigma(g) \times I$ along the curves $\alpha_i \times \{0\}$ and $\beta_j \times \{1\}$ for $i, j \in \{1, \dots, n\}$. The sutures are defined by taking $s(\gamma(g)) = \partial\Sigma(g) \times \left\{\frac{1}{2}\right\}$ and $\gamma(g) = \partial\Sigma(g) \times I$. Here, we are using the outward normal first convention for the induced orientation on the boundary and we are viewing I with the usual orientation (oriented from 0 to 1). Thus, the induced orientation on the boundary would give $\Sigma(g) \times \{1\}$ the same orientation as $\Sigma(g)$, and $\Sigma(g) \times \{0\}$ the opposite. Note that $(E(f), \gamma(g))$ is a strongly balanced sutured manifold with one suture for each X and O. An example for the trivial knot is shown in Figures 49 and 50. Here, we are viewing one of the O's as being associated with a vertex. In Figure 50, one needs to attach 2–handles to $\alpha_i \times \{0\}$ and $\beta_i \times \{1\}$ to obtain $E(f)$.

The sutured manifold $(E(f), \gamma(g))$ is similar to the sutured manifold $(E(f), \gamma(f))$ except that there are an extra $2n_e$ sutures per edge (all parallel and alternating in orientation), where n_e is the number of O's associated to the edge e (this does not count the O's associated to the vertices at the boundary of the edge). Since $(E(f), \gamma(f))$ and $(E(f), \gamma(g))$ are balanced sutured manifolds, we can consider their sutured Floer homologies

$$\mathrm{SFH}(E(f), \gamma(f)) \quad \text{and} \quad \mathrm{SFH}(E(f), \gamma(g)),$$

respectively. See [6] for the definition of sutured Floer homology. We note that the former group is an invariant of the spatial graph while the latter group $\mathrm{SFH}(E(f), \gamma(g))$ depends on the grid g. Each of these groups has two (relative) gradings, an $H_1(M)$ (or equivalently Spin^c) and a homological grading, which we discuss below.

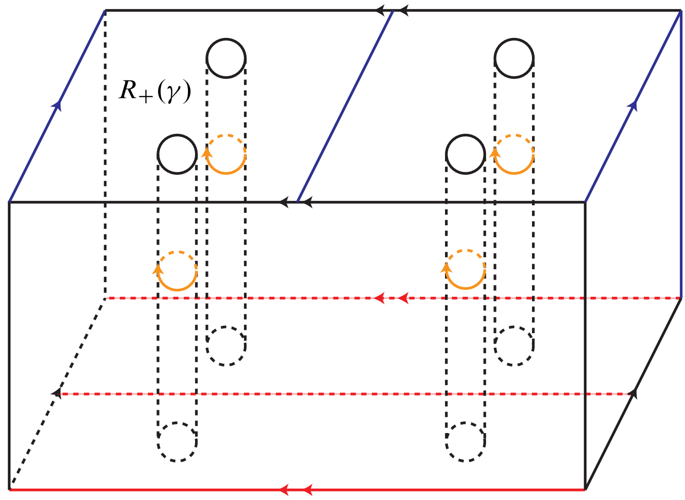

Figure 50: The sutured manifold $(E(f), \gamma(g))$ corresponding to the graph grid diagram g in Figure 49

Let (M,γ) be a balanced sutured manifold. We first discuss the homological grading on $\mathrm{SFH}(M,\gamma)$. Let $(\Sigma,\boldsymbol{\alpha},\boldsymbol{\beta})$ be a balanced and admissible diagram for (M,γ), where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ each contain d disjointly embedded curves. Admissible means that every nontrivial periodic domain has both positive and negative coefficients. Recall that $\mathrm{SFH}(M,\gamma)$ is defined as the homology of a chain complex $\mathrm{CFH}(\Sigma,\boldsymbol{\alpha},\boldsymbol{\beta})$ which is an $\mathbb{F}$–vector space and is roughly defined as follows. Consider $\mathbb{T}_{\boldsymbol{\alpha}} = \alpha_1 \times \cdots \times \alpha_d$ and $\mathbb{T}_{\boldsymbol{\beta}} = \beta_1 \times \cdots \times \beta_d$, the d–dimensional tori in $\mathrm{Sym}^d(\Sigma)$. The set of generators of $\mathrm{CFH}(\Sigma,\boldsymbol{\alpha},\boldsymbol{\beta})$ is $\boldsymbol{x} \in \mathbb{T}_{\boldsymbol{\alpha}} \cap \mathbb{T}_{\boldsymbol{\beta}}$ and the differential is defined by counting rigid holomorphic disks in $\mathrm{Sym}^d(\Sigma)$ connecting two points in $\mathbb{T}_{\boldsymbol{\alpha}} \cap \mathbb{T}_{\boldsymbol{\beta}}$. Choose orientations on $\mathbb{T}_{\boldsymbol{\alpha}}$, $\mathbb{T}_{\boldsymbol{\beta}}$ and $\mathrm{Sym}^d(\Sigma)$, and define $\boldsymbol{m}(\boldsymbol{x})$ to be the intersection sign of $\mathbb{T}_{\boldsymbol{\alpha}}$ and $\mathbb{T}_{\boldsymbol{\beta}}$ in $\mathrm{Sym}^d(\Sigma)$. This depends on the choice of orientations but the difference $\boldsymbol{m}(\boldsymbol{x})\boldsymbol{m}(\boldsymbol{y})^{-1}$ between $\boldsymbol{m}(\boldsymbol{x})$ and $\boldsymbol{m}(\boldsymbol{y})$ is independent of the choice of orientations. So $\boldsymbol{m}$ gives a well-defined relative $\{\pm 1\}$–grading on $\mathrm{CFH}(\Sigma,\boldsymbol{\alpha},\boldsymbol{\beta})$. Let $\mathbb{Z}_2 = \{0,1\}$ be the group with 2 elements and let $\exp\colon \mathbb{Z}_2 \to \{\pm 1\}$ be the isomorphism sending l to $(-1)^l$. Using $e\boldsymbol{m} := \exp^{-1} \circ \boldsymbol{m}$ (instead of just $\boldsymbol{m}$), we get a relative $\mathbb{Z}_2$–grading on $\mathrm{CFH}(\Sigma,\boldsymbol{\alpha},\boldsymbol{\beta})$. Since the parity of the Maslov index of a holomorphic disk connecting $\boldsymbol{x}$ to $\boldsymbol{y}$ is equal to $\exp^{-1}(\boldsymbol{m}(\boldsymbol{x})\boldsymbol{m}(\boldsymbol{y})^{-1})$, the differential reduces the homological grading by $1 \pmod 2$.

We now briefly review the Spin^c–grading. See [7, Section 3] for details. Recall that for any balanced sutured manifold (M,γ), one can define $\mathrm{Spin}^c(M,\gamma)$, the set of Spin^c structures of (M,γ), as the set of homology classes of nowhere-zero vector fields on M that restrict to a special vector field ν_0 on ∂M. The vector field ν_0 depends on the sutures. One can use obstruction theory to see that $\mathrm{Spin}^c(M,\gamma)$

forms an affine space over $H^2(M,\partial M)$ which we will identify with $H_1(M)$ via the Poincaré duality isomorphism (PD: $H^2(M,\partial M)\to H_1(M)$). Thus, once we pick a fixed $\mathfrak{s}_0\in \mathrm{Spin}^c(M,\gamma)$, there is a unique bijective correspondence $\xi_{\mathfrak{s}_0}\colon H_1(M)\to \mathrm{Spin}^c(M,\gamma)$, $v\mapsto v+\mathfrak{s}_0$, making $\mathrm{Spin}^c(M,\gamma)$ into an abelian group. Moreover, for any $\mathfrak{s},\mathfrak{t}\in \mathrm{Spin}^c(M,\gamma)$, there is a well-defined difference, denoted $\mathfrak{s}-\mathfrak{t}\in H_1(M)$, that is defined as the unique element in $H_1(M)$ such that $(\mathfrak{s}-\mathfrak{t})+\mathfrak{t}=\mathfrak{s}$. Note that the difference does not depend on any choices. Indeed, it is easy to see that for any $\mathfrak{s}_0\in \mathrm{Spin}^c(M,\gamma)$ and $v,w\in H_1(M)$, $(v+\mathfrak{s}_0)-(w+\mathfrak{s}_0)=v-w$.

To each $\boldsymbol{x}\in \mathbb{T}_{\boldsymbol{\alpha}}\cap\mathbb{T}_{\boldsymbol{\beta}}$, one can assign an element of $\mathrm{Spin}^c(M,\gamma)$, denoted $\mathfrak{s}(\boldsymbol{x})$, making $\mathrm{CFH}(\Sigma,\boldsymbol{\alpha},\boldsymbol{\beta})$ into a graded vector space over $\mathrm{Spin}^c(M,\gamma)$. The differential on $\mathrm{CFH}(\Sigma,\boldsymbol{\alpha},\boldsymbol{\beta})$ preserves the $\mathrm{Spin}^c(M,\gamma)$–grading, giving a $\mathrm{Spin}^c(M,\gamma)$–grading to $\mathrm{SFH}(M,\gamma)$. Using the bijection $\xi_{\mathfrak{s}_0}\colon H_1(M)\to \mathrm{Spin}^c(M,\gamma)$ as described above, $\mathrm{SFH}(M,\gamma)$ becomes a relatively graded vector space over $H_1(M)$. We can easily compute the relative grading as follows. Let $\boldsymbol{x},\boldsymbol{y}\in \mathbb{T}_{\boldsymbol{\alpha}}\cap\mathbb{T}_{\boldsymbol{\beta}}$ and choose paths $a\colon I\to\mathbb{T}_{\boldsymbol{\alpha}}$ and $b\colon I\to\mathbb{T}_{\boldsymbol{\beta}}$ with $\partial a=\partial b=\boldsymbol{y}-\boldsymbol{x}$. Then $a-b$ can be viewed as a 1–cycle in Σ. Using the inclusion map of Σ into M, we can view this as a cycle in M. Let $\varepsilon(\boldsymbol{y},\boldsymbol{x})\in H_1(M)$ be the homology class of $a-b$. By [6, Lemma 4.7], $\varepsilon(\boldsymbol{y},\boldsymbol{x})$ is the relative grading in $H_1(M)$ associated to the $\mathrm{Spin}^c(M)$–grading on (M,γ). That is,

$$\mathfrak{s}(\boldsymbol{y})-\mathfrak{s}(\boldsymbol{x})=\varepsilon(\boldsymbol{y},\boldsymbol{x}). \tag{10}$$

Note that [6] actually says that $\mathrm{PD}(\mathfrak{s}(\boldsymbol{y})-\mathfrak{s}(\boldsymbol{x}))=\varepsilon(\boldsymbol{y},\boldsymbol{x})$, but we have already identified $\mathfrak{s}(\boldsymbol{y})-\mathfrak{s}(\boldsymbol{x})$ with an element of $H_1(M)$ using Poincaré duality. Using the map $(\xi_{\mathfrak{s}_0}^{-1}\circ\mathfrak{s})\oplus m\colon \mathbb{T}_{\boldsymbol{\alpha}}\cap\mathbb{T}_{\boldsymbol{\beta}}\to H_1(M)\oplus\mathbb{Z}_2$, we see that $\mathrm{CFH}(\Sigma,\boldsymbol{\alpha},\boldsymbol{\beta})$ is a well-defined relatively $(H_1(M),\mathbb{Z}_2)$–bigraded chain complex and $\mathrm{SFH}(M,\gamma)$ is a well-defined relatively $(H_1(M),\mathbb{Z}_2)$–bigraded $\mathbb{F}$–vector space. Even though $\mathrm{CFH}(\Sigma,\boldsymbol{\alpha},\boldsymbol{\beta})$ depends on the choice of Heegaard diagram for (M,γ), by [6], the isomorphism class of $\mathrm{SFH}(M,\gamma)$ as a relatively $(H_1(M),\mathbb{Z}_2)$–bigraded vector space only depends on the sutured manifold.

Definition 6.3 For a transverse spatial graph $f\colon G\to S^3$, the *sutured graph Floer homology* is $\mathrm{SFH}(E(f),\gamma(f))$ considered as a relatively $(H_1(E(f)),\mathbb{Z}_2)$–bigraded $\mathbb{F}$–vector space.

Unless otherwise stated, when we refer to $\mathrm{SFH}(E(f),\gamma(f))$ and $\mathrm{SFH}(E(f),\gamma(g))$, we will be considering them as relatively bigraded vector spaces.

6.3 Relating $\widetilde{\mathrm{HFG}}(f)$ and $\mathrm{SFH}(E(f),\gamma(f))$

The main result of this section is that the graph Floer homology of a sinkless and sourceless transverse spatial graph is isomorphic to its sutured Floer homology (after a slight change of the Alexander grading). To prove this, we first notice that the sutured Floer homology of a grid is the same as $\widetilde{\mathrm{HFG}}(g)$. First we need to discuss how the grading is changed.

Let $\mathbb{G}$ be an abelian group and C be a $(\mathbb{G},\mathbb{Z}_2)$–bigraded chain complex or $\mathbb{F}$–vector space with bigrading $C=\bigoplus_{(g,m)\in\mathbb{G}\oplus\mathbb{Z}_2} C_{(g,m)}$. Let $(\mathrm{r}C)_{(g,m)}=C_{(-g,m)}$ for each $g\in\mathbb{G}$ and $m\in\mathbb{Z}_2$. Then C has a $(\mathbb{G},\mathbb{Z}_2)$–bigrading given by $C=\bigoplus_{(g,m)\in\mathbb{G}\oplus\mathbb{Z}_2}(\mathrm{r}C)_{(g,m)}$, which we call the *reverse bigrading* on C. We denote by $\mathrm{r}C$ the underlying vector space C with its reverse bigrading. If C is a chain complex, then $\partial\colon (\mathrm{r}C)_{(g,m)}\to(\mathrm{r}C)_{(g,m-1)}$. So ∂ is a degree $(0,-1)$ map on $\mathrm{r}C$ and $\mathrm{r}C$ is a $(\mathbb{G},\mathbb{Z}_2)$–bigraded chain complex. If C has a relative $(\mathbb{G},\mathbb{Z}_2)$–bigrading then $\mathrm{r}C$ has a well-defined relative $(\mathbb{G},\mathbb{Z}_2)$–bigrading. Note that we are only "reversing" the $\mathbb{G}$–grading.

We note that using the natural projection of $\mathbb{Z}$ to $\mathbb{Z}_2$, $\widehat{\mathrm{HFG}}(f)$, $\widetilde{\mathrm{HFG}}(f)$, $\widehat{\mathrm{HFG}}(g)$ and $\widetilde{\mathrm{HFG}}(g)$ become relatively $(H_1(E(f)),\mathbb{Z}_2)$–bigraded $\mathbb{F}$–vector spaces (similarly for the chain complexes defining them).

Lemma 6.4 *Let $f\colon G\to S^3$ be a sinkless and sourceless transverse spatial graph and g be a graph grid diagram representing f. Then*

$$\widetilde{\mathrm{HFG}}(g)\cong \mathrm{rSHF}(E(f),\gamma(g))$$

as relatively $(H_1(E(f)),\mathbb{Z}_2)$–bigraded $\mathbb{F}$–vector spaces.

Proof Let $(\Sigma(g),\boldsymbol{\alpha},\boldsymbol{\beta})$ be the specific Heegaard decomposition for $E(f)$ associated to the graph grid g as defined beforehand. Both $\mathrm{CFH}(\Sigma(g),\boldsymbol{\alpha},\boldsymbol{\beta})$ and $\widetilde{C}(g)$ are $\mathbb{F}$–vector spaces with the same generating set. One can check that the boundary maps are the same, giving an identification of the two chain complexes. Thus, we just need to compare their (relative) gradings.

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be generators (in either chain complex). If there is a rectangle connecting $\boldsymbol{x}$ and $\boldsymbol{y}$, then by (1), $M(\boldsymbol{x})-M(\boldsymbol{y})=1 \pmod 2$. Moreover, using the definition of $\boldsymbol{m}$ as described in the previous subsection, it is straightforward to check that $\boldsymbol{m}(\boldsymbol{x})\boldsymbol{m}(\boldsymbol{y})^{-1}=-1\in\{\pm1\}$. Now, for any $\boldsymbol{x}$ and $\boldsymbol{y}$, there is a sequence of rectangles $r_1,\dots,r_k$ connecting $\boldsymbol{x}$ and $\boldsymbol{y}$. Thus, $M(\boldsymbol{x})-M(\boldsymbol{y})=k \pmod 2$ and $\boldsymbol{m}(\boldsymbol{x})\boldsymbol{m}(\boldsymbol{y})^{-1}=(-1)^k\in\{\pm1\}$. Hence

$$(-1)^{M(\boldsymbol{x})-M(\boldsymbol{y})}=\boldsymbol{m}(\boldsymbol{x})\boldsymbol{m}(\boldsymbol{y})^{-1}.$$

By Lemma 4.13 and Equation (10), we see that

$$A(\boldsymbol{x})-A(\boldsymbol{y})=\varepsilon(\boldsymbol{y},\boldsymbol{x})=-(\mathfrak{s}(\boldsymbol{x})-\mathfrak{s}(\boldsymbol{y})). \qquad \square$$

Using [7, Proposition 5.4], we can relate $\mathrm{SFH}(E(f),\gamma(f))$ and $\mathrm{SFH}(E(f),\gamma(g))$. For $g\in\mathbb{G}$ and $m\in\mathbb{Z}_2$, let $W(-g,-1)$ (which equals $W(-g,1)$) be the 2–dimensional $(\mathbb{G},\mathbb{Z}_2)$–bigraded vector space over $\mathbb{F}$ spanned by one generator in degree $(0,0)$ and the other in degree $(-g,-1)$. (Note that we are slightly abusing notation since, in Section 4, we defined $W(-g,-1)$ to be the $(\mathbb{G},\mathbb{Z})$–bigraded vector space over $\mathbb{F}$ spanned by one generator in degree $(0,0)$ and the other in degree $(-g,-1)$.) If (C,∂) is a relatively bigraded $(\mathbb{G},\mathbb{Z}_2)$–chain complex over $\mathbb{F}$, then $C\otimes W(-g,-1)$ becomes a relatively bigraded $(\mathbb{G},\mathbb{Z}_2)$ chain complex with boundary $\partial\otimes\mathrm{id}$ in the usual way.

Proposition 6.5 *Let $f\colon G\to S^3$ be a sinkless and sourceless transverse spatial graph and let g be a graph grid diagram representing f. Then*

$$\mathrm{SFH}(E(f),\gamma(g))\cong\mathrm{SFH}(E(f),\gamma(f))\otimes\bigotimes_{e\in E(G)}W(w(e),1)^{\otimes n_e}$$

as relatively $(H_1(E(f)),\mathbb{Z}_2)$–bigraded $\mathbb{F}$–vector spaces, where n_e is the number of O*'s in g associated to the interior of e (not including the vertices).*

Proof Recall that $(E(f),\gamma(g))$ is a sutured manifold with $2n_e+1$ sutures associated to each edge e (n_e sutures are associated to O's on the interior of e and the other n_e+1 sutures are associated to X's on the interior of e). Pick an edge e. If $n_e=0$, leave the sutures on that edge alone. If $n_e\geq 1$, then there are at least three sutures associated to the edge that are parallel and have alternating orientations. Let S be the properly embedded surface in $E(f)$ as pictured in Figure 51 with either orientation. In this figure, the inner annulus is part of the boundary of the neighborhood of the edge of the graph. It contains the three parallel sutures with alternating orientations. Since G is sinkless and sourceless, no component of ∂S will bound a disk in $R(\gamma)$. Thus, one can verify that S is a decomposing surface and hence it defines a sutured manifold decomposition

$$(E(f),\gamma(g))\overset{S}{\rightsquigarrow}(E(f)',\gamma(g)').$$

See [7, Definition 2.5] for the definition of a decomposing surface and a sutured manifold decomposition. The resulting manifold $E(f)'$ is defined as $E(f)\setminus\mathrm{int}(N(S))$, where $N(S)$ is a neighborhood of S in $E(f)$ and hence is homeomorphic to the disjoint union of $E(f)$ and $S^1\times D^2$. To get the sutures on $E(f)\subset E(f)'$, we remove two of the three aforementioned sutures associated to e with opposite orientations. There are four sutures on $S^1\times D^2\subset E(f)'$; they are all parallel to $S^1\times\{p\}$ for $p\in\partial D^2$ and have alternating orientations. Let (M_1,γ_1) be the component of $(E(f)',\gamma(g)')$ with M_1 homeomorphic to $E(f)$ and (M_2,γ_2) be the component of $(E(f)',\gamma(g)')$ with M_2 homeomorphic to $S^1\times D^2$.

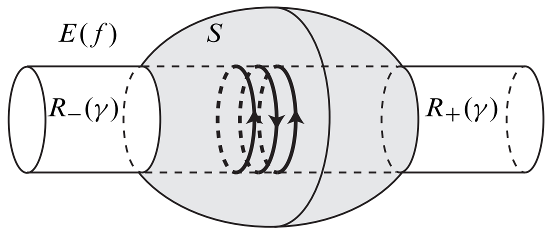

Figure 51: The decomposing surface S (shaded) properly embedded in $E(f)$

Since S is a nice decomposing surface (see [7, Definition 3.22] for the definition of nice), we can use [7, Proposition 5.4] to compute $\mathrm{SFH}(E(f)', \gamma(g)')$. First we need some notation. Let $i\colon E(f)' \to E(f)$ be the inclusion map and $i_*\colon H_1(E(f)') \to H_1(E(f))$ be the induced map on $H_1(-)$. For $\mathfrak{s} \in \mathrm{Spin}^c(E(f), \gamma(g))$, $\mathrm{CFH}(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathfrak{s})$ is defined as the chain complex generated by $\boldsymbol{x} \in \mathbb{T}_{\boldsymbol{\alpha}} \cap \mathbb{T}_{\boldsymbol{\beta}}$ with $\mathfrak{s}(\boldsymbol{x}) = \mathfrak{s}$, where $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is a balanced admissible diagram for $(E(f), \gamma(g))$. $\mathrm{SFH}(E(f), \gamma(g), \mathfrak{s})$ is the homology of $\mathrm{CFH}(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathfrak{s})$ (similarly for $(E(f)', \gamma(g)')$). By [7, Proposition 5.4], there is an affine map

$$f_S\colon \mathrm{Spin}^c(E(f)', \gamma(g)') \to \mathrm{Spin}^c(E(f), \gamma(g))$$

satisfying the following two conditions:

(1) There is an isomorphism

$$\widehat{p}\colon \mathrm{SFH}(E(f)', \gamma(g)') \xrightarrow{\cong} \bigoplus_{\mathfrak{s} \in \mathfrak{I}(f_S)} \mathrm{SFH}(E(f), \gamma(g), \mathfrak{s})$$

such that for every $\mathfrak{s}' \in \mathrm{Spin}^c(E(f)', \gamma(g)')$ we have

$$\widehat{p}(\mathrm{SFH}(E(f)', \gamma(g)', \mathfrak{s}')) \subset \mathrm{SFH}(E(f), \gamma(g), f_S(\mathfrak{s}')).$$

(2) If $\mathfrak{s}'_1, \mathfrak{s}'_2 \in \mathrm{Spin}^c(E(f)', \gamma(g)')$, then

$$i_*(\mathfrak{s}'_1 - \mathfrak{s}'_2) = f_S(\mathfrak{s}'_1) - f_S(\mathfrak{s}'_2) \in H_1(E(f)).$$

Note that in [6], Juhász identifies $\mathrm{Spin}^c(E(f), \gamma(g))$ with $H^2(E(f), \partial E(f))$, so the statement looks slightly different. Since $i_{|M_1}\colon M_1 \to E(f)$ is a homotopy equivalence, $i_*\colon H_1(E(f)') \to H_1(E_f)$ is surjective. Fix an $\mathfrak{s}'_0 \in \mathrm{Spin}^c(E(f)', \gamma(g)')$ and let $\mathfrak{s}_0 := f_S(\mathfrak{s}'_0)$. Use $\mathfrak{s}_0$ to identify $H_1(E(f))$ and $\mathrm{Spin}^c(E(f), \gamma(g))$. This identifies $v \in H_1(E(f))$ with $v + \mathfrak{s}_0 \in \mathrm{Spin}^c(E(f), \gamma(g))$. Let $\mathfrak{s} \in \mathrm{Spin}^c(E(f), \partial E(f))$. Then $\mathfrak{s} = v + \mathfrak{s}_0$ for some $v \in H_1(E(f))$. Since i_* is surjective, there is a $v' \in H_1(E(f)')$ with $i_*(v') = v$. The second statement above implies that

$$i_*(v') = i_*((v' + \mathfrak{s}'_0) - \mathfrak{s}'_0) = f_S(v' + \mathfrak{s}'_0) - f_S(\mathfrak{s}'_0) = f_S(v' + \mathfrak{s}'_0) - \mathfrak{s}_0.$$

Therefore, $\mathfrak{s} = v + \mathfrak{s}_0 = i_*(v') + \mathfrak{s}_0 = f_S(v' + \mathfrak{s}'_0)$, so f_S is surjective.

Choose an orientation on Σ, α_i and β_j for all i, j. Then we have an absolute $\mathbb{Z}_2$–grading on $\mathrm{SFH}(E(f), \gamma(g))$ given by $\mathrm{e}\boldsymbol{m} = \exp^{-1} \circ \boldsymbol{m}$. It can be shown that, as a relative grading, it agrees with $gr \pmod 2$. For a definition of gr see [6, Definition 8.1]. However, gr is only defined for two generators that have the same Spin^c class. Thus, we will need to consider the proof [7, Proposition 5.4], in order to show that $\mathrm{e}\boldsymbol{m}$ is preserved under $\hat{p}$. In this proof, he considers a balanced diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, P)$ adapted to the surface S in $(E(f), \gamma(g))$. Here $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is a Heegaard diagram for $(E(f), \gamma(g))$ and $P \subset \Sigma$ is a quasipolygon. Using P, he then constructs a Heegaard diagram $(\Sigma', \boldsymbol{\alpha}', \boldsymbol{\beta}')$ for $(E(f)', \gamma(g)')$ and a map

$$p\colon \Sigma' \to \Sigma$$

such that p sends α'_i to α_i and β'_j to β_j and

$$p_{|\Sigma' \setminus p^{-1}(P)}\colon \Sigma' \setminus p^{-1}(P) \to \Sigma \setminus P$$

is a diffeomorphism. Moreover, since f_S is onto in our case, it follows that all the intersections of α_i and β_j lie in $\Sigma \setminus P$ and all the intersections of α'_i and β'_j lie in $\Sigma' \setminus p^{-1}(P)$. The map p induces a bijection $\hat{p}\colon \mathbb{T}_{\boldsymbol{\alpha}'} \cap \mathbb{T}_{\boldsymbol{\beta}'} \to \mathbb{T}_{\boldsymbol{\alpha}} \cap \mathbb{T}_{\boldsymbol{\beta}}$. This gives the isomorphism $\hat{p}\colon \mathrm{SFH}(E(f)', \gamma(g)') \to \mathrm{SFH}(E(f), \gamma(g))$. Choose the orientations of Σ', α'_i and β'_j coming from Σ, α_i and β_j so that p preserves all the orientations. It then follows that if $\boldsymbol{x}' \in \mathbb{T}_{\boldsymbol{\alpha}'} \cap \mathbb{T}_{\boldsymbol{\beta}'}$ then

$$\boldsymbol{m}(\hat{p}(\boldsymbol{x}')) = \boldsymbol{m}(\boldsymbol{x}'). \tag{11}$$

Using f_S and using the above identification of $H_1(E(f))$ and $\mathrm{Spin}^c(E(f), \gamma(g))$, $\mathrm{SFH}(E(f)', \gamma(g)')$ inherits an $H_1(E(f))$–grading. This makes $\mathrm{SFH}(E(f), \gamma(g))$ and $\mathrm{SFH}(E(f)', \gamma(g)')$ into $(H_1(E(f)), \mathbb{Z}_2)$–bigraded $\mathbb{F}$–vector spaces. We can now show that $\hat{p}\colon \mathrm{SFH}(E(f)', \gamma(g)') \to \mathrm{SFH}(E(f), \gamma(g))$ is an $(H_1(E(f)), \mathbb{Z}_2)$–bigraded map. Let $x' \in \mathrm{SFH}(E(f)', \gamma(g)')_{(v,i)}$, where $(v, i) \in H_1(E(f)) \oplus \mathbb{Z}_2$. Since

$$f_S(v' + \mathfrak{s}'_0) = i_*(v') + \mathfrak{s}_0,$$

x' is in $\mathrm{SFH}(E(f)', \gamma(g)', v' + \mathfrak{s}'_0)$ for some $v' \in H_1(E(f)')$ with $i_*(v') = v$. So by [7, Proposition 5.4(1)],

$$\hat{p}(\mathrm{SFH}(E(f)', \gamma(g)', v' + \mathfrak{s}'_0)) \subset \mathrm{SFH}(E(f), \gamma(g), v + \mathfrak{s}_0).$$

Using this and (11), it follows that $\hat{p}(x') \in \mathrm{SFH}(E(f), \gamma(g))_{(v,i)}$.

We show that

$$\mathrm{SFH}(E(f)', \gamma(g)') \cong \mathrm{SFH}(M_1, \gamma_1) \otimes W(w(e), 1).$$

To see this, note that $(E(f)', \gamma(g'))$ is the disjoint union of (M_1, γ_1) and (M_2, γ_2). Moreover, it is easy to see that

$$\mathrm{SFH}(M_2, \gamma_2) \cong W(e(g), 1)$$

as a relatively $(H_1(E(f)), \mathbb{Z}_2)$–bigraded $\mathbb{F}$–vector space. To complete the proof, we keep removing pairs of sutures on each edge until we are left with $(E(f), \gamma(f))$. □

We can use this to complete the relationship between $\widehat{\mathrm{HFG}}(f)$ and $\mathrm{rSHF}(E(f), \gamma(f))$.

Theorem 6.6 *Let $f\colon G \to S^3$ be a sinkless and sourceless transverse spatial graph where G has no cut edges. Then*

$$\widehat{\mathrm{HFG}}(f) \cong \mathrm{rSHF}(E(f), \gamma(f))$$

as relatively $(H_1(E(f)), \mathbb{Z}_2)$–bigraded $\mathbb{F}$–vector spaces.

Proof Let g be a graph grid diagram representing f. By Proposition 4.32, Lemma 6.4, and Proposition 6.5, we have that

$$\begin{aligned}
\widehat{\mathrm{HFG}}(f) \otimes \bigotimes_{e \in E(G)} W(-w(e), -1)^{\otimes n_e} &\cong \widetilde{\mathrm{HFG}}(g) \\
&\cong \mathrm{rSHF}(E(f), \gamma(g)) \\
&\cong \mathrm{r}\Bigg(\mathrm{SFH}(E(f), \gamma(f)) \otimes \bigotimes_{e \in E(G)} W(w(e), 1)^{\otimes n_e}\Bigg) \\
&\cong \mathrm{rSHF}(E(f), \gamma(f)) \otimes \bigotimes_{e \in E(G)} W(-w(e), -1)^{\otimes n_e}
\end{aligned}$$

as relatively $(H_1(E(f)), \mathbb{Z}_2)$–bigraded $\mathbb{F}$–vector spaces.

The result follows from a slight generalization of [22, Lemma 3.18] (replace $\mathbb{Z}^2$ with $H_1(E(f)) \oplus \mathbb{Z}_2 \cong \mathbb{Z}^l \oplus \mathbb{Z}_2$ in the proof) that $\widehat{\mathrm{HFG}}(f) \cong \mathrm{rSHF}(E(f), \gamma(f))$ as relatively $(H_1(E(f)), \mathbb{Z}_2)$–bigraded $\mathbb{F}$–vector spaces. We will sketch of proof of this. Let $V = \bigotimes_{e \in E(G)} W(-w(e), -1)^{\otimes n_e}$. After shifting by an element of $H_1(E(f)) \oplus \mathbb{Z}_2$, we can assume that $\mathrm{rSHF}(E(f), \gamma(f))$ is (absolutely) bigraded and that $\widehat{\mathrm{HFG}}(f) \otimes V \cong \mathrm{rSHF}(E(f), \gamma(f)) \otimes V$ as $(H_1(E(f)), \mathbb{Z}_2)$ (absolutely) bigraded $\mathbb{F}$–vector spaces. Since G has no cut edges, $w(e) \neq 0$ for all $e \in E(G)$.

Suppose $V_1 \otimes W(a, m) \cong V_2 \otimes W(a, m)$ as $(H_1(E(f)), \mathbb{Z}_2)$–bigraded $\mathbb{F}$–vector spaces, where $(a, m) \in H_1(E(f)) \oplus \mathbb{Z}_2$, $a \neq 0$ and V_i is a finitely generated $\mathbb{F}$–vector space. V_i can be represented as a function $f_i\colon H_1(E(f)) \oplus \mathbb{Z}_2 \to \mathbb{Z}_{\geq 0}$, where $f_i(h, n) = \dim_{\mathbb{F}}(V_i)_{(h,n)}$. Since V_i is finitely generated, $f_i(h, n) = 0$ for all but finitely many pairs $(h,n) \in H_1(E(f)) \oplus \mathbb{Z}_2$. Then $V_i \otimes W(h,-1)$ can be represented by the function $g_i\colon H_1(E(f)) \oplus \mathbb{Z}_2 \to \mathbb{Z}_{\geq 0}$ where

$$g_i(h,n) = f_i(h,n) + f_i(h-a, n-m).$$

Since $V_1 \otimes W(a,m) \cong V_2 \otimes W(a,m)$, we have that $g_1 = g_2$. We now note that since $f_i(h,n) = 0$ for all but finitely many pairs (h,n) that $g_i(h,n) = 0$ for all but finitely many pairs. In addition, since $a \neq 0$ and $H_1(E(f)) \cong \mathbb{Z}^l$, $(h-ja, n-jm) \neq (h-j'a, n-j'm)$ whenever $j \neq j'$. Hence $\sum_{j=0}^{\infty} g_i(h-ja, n-jm)$ is a well-defined function and

$$f_i(h,n) = \sum_{j=0}^{\infty} g_i(h-ja, n-jm)$$

for all $(h,n) \in H_1(E(f)) \oplus \mathbb{Z}_2$. Since $g_1 = g_2$, it follows that $f_1 = f_2$ and hence $V_1 \cong V_2$ as $(H_1(E(f)), \mathbb{Z}_2)$–bigraded $\mathbb{F}$–vector spaces. □

As a result we see that the decategorification of $\widehat{\mathrm{HFG}}(f)$ is essentially the torsion invariant $T(E(f), \gamma(f))$ associated to sutured manifold $(E(f), \gamma(f))$.

Definition 6.7 Let $f\colon G \to S^3$ be a sinkless and sourceless transverse spatial graph. Define

$$\chi(\widehat{\mathrm{HFG}}(f)) = \sum_{(h,i)\in(H_1(E(f)),\mathbb{Z})} (-1)^i \operatorname{rank}_{\mathbb{F}} \widehat{\mathrm{HFG}}(f)_{(h,i)} h$$

considered as an element of $\mathbb{Z}[H_1(E(f)]$ modulo positive units; ie as an element of $H_1(E(f))$.

If $r \in \mathbb{Z}[H_1(E(f))]$ then $r = \sum_i a_i h_i$, where $a_i \in \mathbb{Z}$ and $h_i \in H_1(E(f))$. Define $\bar{r} := \sum_i a_i h_i^{-1}$, where here we are viewing $H_1(E(f))$ as a multiplicative group.

Corollary 6.8 *If $f\colon G \to S^3$ is a sinkless and sourceless transverse spatial graph where G has no cut edges then*

$$\chi(\widehat{\mathrm{HFG}}(f)) \doteq \bar{\Delta}_f.$$

That is, they are the same up to multiplication by units in $\mathbb{Z}[H_1(E(f))]$.

Proof Choose an $\mathfrak{s}_0 \in \mathrm{Spin}^c(E(f), \gamma(f))$. By [4, Theorem 1.1], we have that $\chi(\mathrm{SFH}(E(f), \gamma(g))) \doteq \Delta_f$, where

$$\chi(\mathrm{SFH}(E(f), \gamma(g))) = \sum_{\boldsymbol{x}\in\mathbb{T}_{\boldsymbol{\alpha}}\cap\mathbb{T}_{\boldsymbol{\beta}}} \boldsymbol{m}(\boldsymbol{x})\xi_{\mathfrak{s}_0}^{-1}(\mathfrak{s}(\boldsymbol{x}))$$

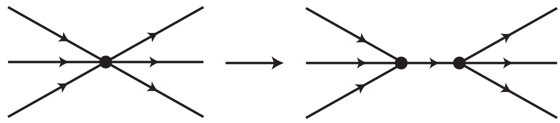

Figure 52: Transverse spatial graph to balanced bipartite graph with balanced orientation

for any admissible balanced Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ for $(E(f), \gamma(f))$. By a standard argument,

$$\sum_{\boldsymbol{x} \in \mathbb{T}_{\boldsymbol{\alpha}} \cap \mathbb{T}_{\boldsymbol{\beta}}} \boldsymbol{m}(\boldsymbol{x}) \xi_{\mathfrak{s}_0}^{-1}(\mathfrak{s}(\boldsymbol{x})) = \sum_{(h,i) \in (H_1(E(f)), \mathbb{Z})} (-1)^i \operatorname{rank}_{\mathbb{F}} \widehat{\mathrm{SFH}}(E(f), \gamma(g))_{(h,i)} h.$$

Hence, the result follows from Theorem 6.6. □

We now relate Bao's theory for balanced bipartite spatial graph with balanced orientations to ours. Given a transverse spatial graph, one can get a balanced bipartite spatial graph with balanced orientation by replacing each vertex with an edge; see Figure 52.

Conversely, suppose one has a balanced bipartite spatial graph $F\colon G_{V_1,V_2} \to S^3$ with balanced orientation. This means that $V_1 \cup V_2$ is the set of vertices, $|V_1| = |V_2|$ and each edge connects a vertex in V_1 with one in V_2. Moreover, since this has a balanced orientation, there are precisely $|V_1|$ edges that are oriented from V_1 to V_2, and these edges give a bijection between V_1 and V_2; see [1] for the precise definitions. Thus, we can isotope F so that all of these edges are very short straight arcs like on the right-hand side of Figure 52. We can collapse this edge to obtain a transverse spatial graph. Moreover, Bao constructs a sutured manifold which is the same as $(E(f), \gamma(f))$ for a transverse spatial graph f, and points out in [1, Proposition 4.11] that her hat version is the same as the sutured Floer homology of her sutured manifold. Thus, by Theorem 6.6 and Corollary 6.8, our hat theory (and decategorification) is the same as her hat theory (and decategorification), up to reversing the bigrading, as long as the underlying graph for the transverse graph has no sinks, sources or cut edges.

References

[1] **Y Bao**, *Heegaard Floer homology for embedded bipartite graphs*, preprint (2014) arXiv

[2] **P R Cromwell**, *Embedding knots and links in an open book, I: Basic properties*, Topology Appl. 64 (1995) 37–58 MR

[3] **I A Dynnikov**, *Arc-presentations of links: monotonic simplification*, Fund. Math. 190 (2006) 29–76 MR

[4] **S Friedl**, **A Juhász**, **J Rasmussen**, *The decategorification of sutured Floer homology*, J. Topol. 4 (2011) 431–478 MR

[5] **P Ghiggini**, *Knot Floer homology detects genus-one fibred knots*, Amer. J. Math. 130 (2008) 1151–1169 MR

[6] **A Juhász**, *Holomorphic discs and sutured manifolds*, Algebr. Geom. Topol. 6 (2006) 1429–1457 MR

[7] **A Juhász**, *The sutured Floer homology polytope*, Geom. Topol. 14 (2010) 1303–1354 MR

[8] **L H Kauffman**, *Invariants of graphs in three-space*, Trans. Amer. Math. Soc. 311 (1989) 697–710 MR

[9] **S Kinoshita**, *Alexander polynomials as isotopy invariants, I*, Osaka Math. J. 10 (1958) 263–271 MR

[10] **R Litherland**, *The Alexander module of a knotted theta-curve*, Math. Proc. Cambridge Philos. Soc. 106 (1989) 95–106 MR

[11] **C Manolescu**, **P Ozsváth**, **S Sarkar**, *A combinatorial description of knot Floer homology*, Ann. of Math. 169 (2009) 633–660 MR

[12] **C Manolescu**, **P Ozsváth**, **Z Szabó**, **D Thurston**, *On combinatorial link Floer homology*, Geom. Topol. 11 (2007) 2339–2412 MR

[13] **J McAtee**, **D S Silver**, **S G Williams**, *Coloring spatial graphs*, J. Knot Theory Ramifications 10 (2001) 109–120 MR

[14] **J McCleary**, *User's guide to spectral sequences*, Mathematics Lecture Series 12, Publish or Perish, Wilmington, DE (1985) MR

[15] **Y Ni**, *Link Floer homology detects the Thurston norm*, Geom. Topol. 13 (2009) 2991–3019 MR

[16] **P S Ozsváth**, **A I Stipsicz**, **Z Szabó**, *Grid homology for knots and links*, Mathematical Surveys and Monographs 208, Amer. Math. Soc., Providence, RI (2015) MR

[17] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and genus bounds*, Geom. Topol. 8 (2004) 311–334 MR

[18] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and knot invariants*, Adv. Math. 186 (2004) 58–116 MR

[19] **P Ozsváth**, **Z Szabó**, *Holomorphic disks, link invariants and the multivariable Alexander polynomial*, Algebr. Geom. Topol. 8 (2008) 615–692 MR

[20] **J A Rasmussen**, *Floer homology and knot complements*, PhD thesis, Harvard University (2003) MR Available at `http://search.proquest.com/docview/305332635`

[21] **S Sarkar**, *Grid diagrams and the Ozsváth–Szabó tau-invariant*, Math. Res. Lett. 18 (2011) 1239–1257 MR

[22] **E Tweedy**, *The antidiagonal filtration: reduced theory and applications*, Int. Math. Res. Not. 2015 (2015) 7979–8035 MR

[23] **C A Weibel**, *An introduction to homological algebra*, Cambridge Studies in Advanced Mathematics 38, Cambridge University Press (1994) MR

Department of Mathematics, Rice University
MS 136, PO Box 1892, Houston, TX 77251-1892, United States

Department of Mathematics, Indiana University
Rawles Hall, 831 E 3rd St, Bloomington, IN 47505, United States

shelly@rice.edu, odonnol@indiana.edu

Received: 23 September 2015 Revised: 11 October 2016

Algebraic & Geometric Topology 17 (2017) 1527–1555

Positive factorizations of mapping classes

R İNANÇ BAYKUR
NAOYUKI MONDEN
JEREMY VAN HORN-MORRIS

In this article, we study the maximal length of positive Dehn twist factorizations of surface mapping classes. In connection to fundamental questions regarding the uniform topology of symplectic 4–manifolds and Stein fillings of contact 3–manifolds coming from the topology of supporting Lefschetz pencils and open books, we completely determine which boundary multitwists admit arbitrarily long positive Dehn twist factorizations along nonseparating curves, and which mapping class groups contain elements admitting such factorizations. Moreover, for every pair of positive integers g and n, we tell whether or not there exist genus-g Lefschetz pencils with n base points, and if there are, what the maximal Euler characteristic is whenever it is bounded above. We observe that only symplectic 4–manifolds of general type can attain arbitrarily large topology regardless of the genus and the number of base points of Lefschetz pencils on them.

20F65, 53D35, 57R17

1 Introduction

Let Σ_g^n be a compact orientable genus-g surface with n boundary components, and Γ_g^n denote the mapping class group composed of orientation-preserving homeomorphisms of Σ_g^n which restrict to identity along $\partial\Sigma_g^n$, modulo isotopies fixing the same data. We denote by $t_c \in \Gamma_g^n$ the positive (right-handed) Dehn twist along the simple closed curve $c \subset \Sigma_g^n$. If $\Phi = t_{c_l} \cdots t_{c_1}$ in Γ_g^n, where c_i are nonseparating curves on Σ_g^n, we call the product of Dehn twists a *positive factorization* of Φ in Γ_g^n of *length* l.

Our motivation to study positive factorizations comes from their significance in the study of Stein fillings of contact 3–manifolds, as in Giroux [10] or Loi and Piergallini [17], and that of symplectic 4–manifolds via Lefschetz fibrations and pencils, as in Donaldson [9]. Provided all the twists are along homologically essential curves, a positive factorization of a mapping class in Γ_g^n, $n \geq 1$, prescribes an *allowable Lefschetz fibration* which supports a Stein structure filling the contact structure compatible with the induced genus-g open book on the boundary with n binding components; see Loi

Published: 17 July 2017 DOI: 10.2140/agt.2017.17.1527

and Piergallini [17] and Akbulut and Ozbagci [1]. Similarly, a positive factorization of a boundary multitwist Δ, ie the mapping class $t_{\delta_1} \cdots t_{\delta_n}$ for δ_i boundary parallel curves, describes a genus-g *Lefschetz pencil* with n base points. Conversely, given any allowable Lefschetz fibration or a Lefschetz pencil, one obtains such a factorization.

The main questions we will take on in this article are the following.

Question 1 *When does the page $F \cong \Sigma_g^n$ of an open book impose an a priori bound on the Euler characteristics of allowable Lefschetz fibrations with regular fiber F filling it?*

Question 2 *When does the fiber $F \cong \Sigma_g$ and a positive integer n imply an a priori bound on the Euler characteristics of allowable Lefschetz pencils with regular fiber F and n base points? When it does, what is the largest possible Euler characteristic? More specifically, for which g, n do there exist genus-g Lefschetz pencils with n base points?*

By Giroux, contact structures on 3–manifolds up to isotopies are in one-to-one correspondence with supporting open books up to positive stabilizations [10]. Moreover, a contact 3–manifold (Y, ξ) admits a Stein filling (X, J) if and only if (Y, ξ) admits a *positive open book*, ie an open book whose monodromy can be factorized into positive Dehn twists along homologically essential curves on the page [17]. Since $b_1(X) \leq b_1(Y)$ for any Stein filling X, Question 1 above, *up to stabilizations*, amounts to asking when the page of an open book (ie its genus g and number of binding components n) on a contact 3–manifold implies a uniform bound on the topology of its Stein fillings. On the other hand, Question 2 can be seen as a special version of Question 1 with page Σ_g^n and the particular monodromy Δ.

Related to our focus is the following natural function defined on mapping class groups. The *positive factorization length*, or *length* (as *the* length, in this paper) of a mapping class Φ, which we denote by $\mathrm{L}(\Phi)$, is defined to be the supremum taken over the lengths of all possible positive factorizations of Φ along nonseparating curves, and it is $-\infty$ if Φ does not admit any positive factorization. L is a superadditive function on Γ_g^n taking values in $\mathbb{N} \cup \{\pm\infty\}$. It is easy to see that $\mathrm{L} < +\infty$ translates to having the uniform bound in the above questions.

We will investigate the length of various mapping classes, leading to a surprisingly diverse picture, where our results will in particular provide complete solutions to the above problems. Below, the δ_i always denote boundary parallel curves along distinct boundary components of Σ_g^n for $i = 1, \ldots, n$.

Theorem A *Let* $\Delta = t_{\delta_1} \cdots t_{\delta_n}$ *on* Σ_g^n*, where* $n \geq 1$.[1] *Then*

$$\mathrm{L}(\Delta) = \begin{cases} -\infty & \text{if } g = 1 \text{ and } n > 9, \text{ or } g \geq 2 \text{ and } n > 4g+4, \\ +\infty & \text{if } n \leq 2g-4, \\ \text{finite} & \text{otherwise.} \end{cases}$$

When finite, *the exact value of* $\mathrm{L}(\Delta)$ *is*

$$\mathrm{L}(\Delta) = \begin{cases} 12 & \text{if } g = 1, \\ 40 & \text{if } g = 2, \\ 6g+18 & \text{if } 3 \leq g \leq 6, \\ 8g+4 & \text{if } g \geq 7. \end{cases}$$

In particular, when $\mathrm{L}(\Delta)$ *is finite, its value depends solely on* g*, and not* n*.*

Translating this to Lefschetz pencils, for every pair of fixed positive integers g, n, Theorem A allows us to tell (i) if there are any symplectic 4–manifolds admitting a genus-g Lefschetz pencil with n base points (and with only irreducible singular fibers), (ii) when the Euler characteristic of these 4–manifolds can get arbitrarily large, and (iii) if bounded, what the largest Euler characteristic exactly is. Parts (i) and (ii) completely answer [8, Question 5.1]. In the course of the proof, we will note that only symplectic 4–manifolds of general type, ie of Kodaira dimension 2, realize arbitrarily large Euler characteristic. In contrast, when there is a uniform bound, we will see that the largest Euler characteristic can be realized by a symplectic 4–manifold of Kodaira dimension $-\infty$, 0 or 2.

Our second theorem, inspired by partial observations in Dalyan, Korkmaz and Pamuk [8], shows that when Question 1 is formulated for higher powers of the boundary multitwist Δ (ie for Lefschetz fibrations with sections of self-intersection less than -1 instead of pencils) or for the product of Δ with a single Dehn twist along a nonseparating curve, the uniform bounds in Theorem A disappear for all $g \geq 2$.

Theorem B *Let* $\Delta = t_{\delta_1} \cdots t_{\delta_n}$ *on* Σ_g^n*, where* $n \geq 1$ *is any integer, and* a *be any nonseparating curve on* Σ*. Then*

(1) $\mathrm{L}(\Delta^k) = 12k$ *if* $g = 1$, $k \geq 1$, *and* $n \leq 9$,

(2) $\mathrm{L}(\Delta^k) = +\infty$ *if* $g \geq 2$, $k \geq 2$,

(3) $\mathrm{L}(\Delta\, t_a) = +\infty$ *if* $g \geq 2$.

[1] As we will review while proving the above theorem, it is well-known that for $\Delta = t_{\delta_1} \cdots t_{\delta_n}$ on $\Sigma_g^{n'}$ with $n < n'$, we have $\mathrm{L}(\Delta) = -\infty$. We have therefore expressed our results only for the nontrivial case $n = n'$.

A simple variation of L is obtained by allowing separating Dehn twists along *homologically essential* curves on Σ_g^n in the factorizations, which we denote by $\widetilde{\mathrm{L}}$. Clearly $\mathrm{L}(\Phi) \le \widetilde{\mathrm{L}}(\Phi)$ for all Φ, and $\mathrm{L}(\Gamma_g^n) \subset \widetilde{\mathrm{L}}(\Gamma_g^n)$. Our last theorem determines the *full image* of L and $\widetilde{\mathrm{L}}$ on mapping class groups Γ_g^n.

Theorem C *For $n \ge 1$, the image of Γ_g^n under $\widetilde{\mathrm{L}}$ and L is*

(1) $\widetilde{\mathrm{L}}(\Gamma_g^n) = \mathrm{L}(\Gamma_g^n) = \mathbb{N} \cup \{-\infty\}$ *if $g = 0$ and $n \ge 2$, or if $g = 1$,*

(2) $\widetilde{\mathrm{L}}(\Gamma_g^n) = \mathrm{L}(\Gamma_g^n) = \mathbb{N} \cup \{\pm\infty\}$ *if $g \ge 2$.*

Theorem C, along with parts of Theorems A and B, records the existence of mapping classes with arbitrarily long positive factorizations. In the course of its proof we will in fact spell out mapping classes (as multitwists along nonseparating curves) with positive factorizations of *unique* lengths.

Any Φ with $\mathrm{L}(\Phi) = +\infty$ provides an example of a contact 3–manifold with arbitrarily large Stein fillings. The mapping class Φ prescribes an open book, which in turn determines a contact 3–manifold by the work of Thurston and Winkelnkemper, whose Stein fillings are obtained from the allowable Lefschetz fibrations corresponding to respective positive factorizations. We therefore extend our earlier results in [6; 7], and obtain many more counter-examples to Stipsicz's conjecture [29], which predicted an a priori bound on the Euler characteristics of Stein fillings. Clearly, any contact 3–manifold Stein cobordant to one of these examples also bears the same property. Since having a supporting open book with infinite length monodromy is a contact invariant, our detailed analysis summarized in the results above can be used to distinguish contact structures on 3–manifolds.

The novelty in the proofs of the above theorems is the engagement of essentially four different methods:

(1) Underlying symplectic geometry and Seiberg–Witten theory The bounds and calculations of the maximal length in the finite cases in Theorem A will follow from our analysis of the underlying Kodaira dimension of the symplectic Lefschetz pencils corresponding to these factorizations. Here the symplectic Kodaira dimension will provide a useful way to organize the essential information. Indeed, we will observe that the only classes of 4–manifolds yielding pencils or fibrations with unbounded Euler characteristic are symplectic 4–manifolds of general type, ie of Kodaira dimension 2. Both our Kodaira dimension calculations following Sato [23] and Baykur and Hayano [3] — even the very fact that the Kodaira dimension is a well-defined invariant (see Li [13]) — and the sharp inequalities we obtain heavily depend on Taubes' seminal work on pseudoholomorphic curves and Seiberg–Witten equations on symplectic 4–manifolds.

(2) Dehn monoid and right-veering Realizing the finite lengths in Theorem B (the content of Proposition 4), as well as our recap of previous results which establishes the lack of any positive factorizations for certain mapping classes in Theorems A and C (Proposition 5) will be obtained using Thurston type right-veering arguments as in Short and Wiest [24] and Honda, Kazez and Matić [11], and will rely on the structure of the positive Dehn twist monoid of Γ_g^n.

(3) Homology of mapping class groups of small genus surfaces The precise calculation for the genus-1 case in Theorem B, and the bounds in Theorem C for low genus cases (Proposition 1), will follow from our complete understanding of the first homology group of the corresponding mapping class groups.

(4) Constructions of new monodromy factorizations We will show that an artful application of braid, chain and lantern substitutions applied to carefully tailored mapping class group factorizations allows one to obtain arbitrarily long positive factorizations in Theorems A and B (Theorems 9, 16 and 17). This greatly extends the earlier array of partial results of Baykur, Korkmaz and Monden [5], Baykur and Van Horn-Morris [6; 7] and Dalyan, Korkmaz and Pamuk [8] to the possible limits of these constructions as dictated by our results above.

The outline of our paper is as follows. In Section 2, we discuss mapping classes with bounded lengths, and in Section 3 we construct those with infinite lengths. These results will be assembled to complete the proofs of our main theorems in Section 4, where we will also present a couple more results on lengths of mapping classes and special subgroups of mapping class groups, and discuss some related questions.

Acknowledgements The first author was partially supported by the NSF Grant DMS-1510395 and the Simons Foundation Grant 317732. The second author was supported by the Grant-in-Aid for Young Scientists (B) (No. 13276356), Japan Society for the Promotion of Science. The third author was partially supported by the Simons Foundation Grant 279342.

2 Mapping classes with finite lengths

Here we investigate various examples of mapping classes with no positive factorizations or with only factorizations of bounded length. We first probe mapping classes on small genus surfaces, as well as those with small compact support in the interior, who have unique lengths. We then move on to showing that boundary multitwists involving too many boundary components have an a priori bound on the length of their positive factorizations.

2.1 Simple mapping classes with prescribed lengths

There are two tools that we will use to bound the number of Dehn twists in a given factorization. The first uses the fact that the mapping class group Γ_g^n for $g = 0, 1$ surjects to $\mathbb{Z}$, based on $H_1(\Gamma_g^n; \mathbb{Z})$ having a $\mathbb{Z}$ component. Secondly, we will use *right-veering* methods of Thurston [24] and Honda, Kazez and Matić [11].

We begin with analyzing the $g = 0, 1$ case, for any $n \geq 1$. When $g = 0$, equivalent arguments were given by Kaloti in [12] and Plamenevskaya in [19].

Proposition 1 *For $g \leq 1$ any positive factorization of $\Phi \in \Gamma_g^n$ along homologically essential curves has bounded length. So $\widetilde{\mathrm{L}}(\Phi)$ is bounded above for any Φ. Further, when $g = 1$, the length of any factorization into nonseparating Dehn twists is fixed.*

Proof While a careful look at $H_1(\Gamma_g^n)$ will yield more information, a bound on the number of nontrivial Dehn twists in a positive factorization of a mapping class Φ can be obtained by capping off to the base cases of $g = 0$, $n = 2$ or $g = 1$, $n = 1$.

(Genus 0) For genus 0, the base case is $n = 2$. The mapping class group of the annulus Γ_0^2 is isomorphic to $\mathbb{Z}$, where the right-handed Dehn twist about the core of the annulus is mapped to 1. In this case, $\widetilde{\mathrm{L}}(\Phi) = [\Phi]$, the image of Φ under this isomorphism, and moreover, a positive factorization, if it exists, is unique.

When $n > 2$, one can fix an outer boundary component, and identify Σ with a disk with holes. The homomorphism induced by capping off all but a single interior boundary component ∂_i counts the number of Dehn twists (in any factorization) that enclose ∂_i. Every essential curve must enclose at least one interior boundary component and so shows up in at least one of these counts. Adding up the images of Φ for all of these homomorphisms then gives a bound on the number of Dehn twists in any positive factorization of Φ.

(Genus 1) For genus 1, the base case is $n = 1$. The first homology of the mapping class group $H_1(\Gamma_1^1)$ isomorphic to $\mathbb{Z}$, where the right-handed Dehn twist about any nonseparating curve is mapped to 1. The boundary Dehn twist has $[t_\partial] = 12$. In this case, $\mathrm{L}(\Phi) = [\Phi]$, the image of Φ under this isomorphism, and $\widetilde{\mathrm{L}} \leq [\Phi]$.

When $n > 1$, then just as above, one can cap off all but one boundary component and calculate the value of Φ there. Any essential curve will remain essential for at least one of these maps, and so adding up the values of all of the images of Φ will give a bound on the length of any factorization of Φ into Dehn twists along homologically essential curves and an upper bound on $\widetilde{\mathrm{L}}(\Phi)$.

Even better, though, any nonseparating curve will remain nonseparating after every capping, so the length of any positive factorization into nonseparating Dehn twists can be found by capping off all but one boundary component of Σ and calculating $[\Phi]$ there. This determines $\mathrm{L}(\Phi)$ on the nose. □

Remark 2 Notice that combining the above with a theorem of Wendl [31] we recover the following theorem of [19; 12]:

> *If the open book prescribed by* $\Phi \in \Gamma_g^n$, $n \geq 1$, *is stably equivalent to a planar open book, then* $\widetilde{\mathrm{L}}(\Phi)$ *is finite.*

Equivalently, a mapping class $\Phi \in \Gamma_g^n$, $n \geq 1$ with $\widetilde{\mathrm{L}} = +\infty$ cannot be stably equivalent to a mapping class $\Psi \in \Gamma_0^m$, for any $m \geq 1$.

We now move on to producing particular mapping classes with prescribed finite lengths in Γ_g^n for any $g, n \geq 1$, for which we first review the notion of right-veering [11]. Let α and β two properly embedded oriented arcs in an oriented surface Σ with $\partial\Sigma \neq \varnothing$, having the same endpoints $\alpha(0) = \beta(0) = x_0$ and $\alpha(1) = \beta(1)$ on $\partial\Sigma$. Choose a lifted base point $\widetilde{x}_0$ of x_0 and lifts to the universal cover $\widetilde{\alpha}$ and $\widetilde{\beta}$ of the arcs α and β starting at $\widetilde{x}_0$. We say β is *to the right of* α *at* x_0 if the boundary component of $\widetilde{\Sigma}$ containing $\widetilde{\beta}(1)$ is to the right of that containing $\widetilde{\alpha}(1)$ when viewed from $\widetilde{x}_0$.

When Σ is an annulus, we can simply define a mapping class to be right-veering if it is a nonnegative power of the right-handed Dehn twist. For surfaces other than the annulus, think of $\widetilde{\Sigma}$ as sitting in the Poincaré disk model of the hyperbolic plane, lift the arcs to geodesics, and consider the endpoints $\widetilde{\alpha}(1)$ and $\widetilde{\beta}(1)$ radially from the boundary of $\widetilde{\Sigma}$ to the boundary of the disk. If the oriented path from $\widetilde{\beta}(1)$ to $\widetilde{\alpha}(1)$, avoiding $\widetilde{x}_0$, has the same orientation as the boundary orientation induced by the disk, then β is to the right of α. (More generally, one can take homotopic representatives of α and β which intersect minimally and are transverse at x_0, and ask whether in the neighborhood of x_0, β lies on the right or left of α.) A diffeomorphism Φ is called *right-veering* if for every base point x_0 and every properly embedded arc α starting at x_0, either $\Phi(\alpha)$ is *to the right* of α at x_0 or $\Phi(\alpha)$ is homotopic to α, fixing the end points. It is straightforward to see that this property is well-defined for an isotopy class of Φ rel boundary, and is independent of the choice of the base point. Hence we call a mapping class Φ right-veering if any representative of it is. It turns out that Veer_g^n, generated by right-veering mapping classes in Γ_g^n is a monoid of Γ_g^n, and contains Dehn_g^n, generated by all positive Dehn twists in Γ_g^n as a submonoid [11].

Proposition 3 *Let $\prod_{i=1}^{n} t_{c_i}$ be a positive Dehn twist factorization of a mapping class element Φ. If Φ preserves an arc α, then every curve c_i is disjoint from the arc α.*

Proof If Φ is a mapping class which preserves the homotopy class of an arc α, then for any factorization of Φ as a product of right-veering maps $\Phi = \Phi_r \circ \cdots \circ \Phi_1$, we can see that each of the Φ_i also preserve α. For if any Φ_i moves α, then it has to send it to the right, after which every Φ_j, for $i < j \leq r$, either fixes it, or sends it further to the right. Since Dehn twists are right-veering, the proposition follows. $\square$

A *multicurve* C on Σ, which is a collection of disjoint simple closed curves, is said to be *nonisolating* if every connected component of $\Sigma \setminus C$ contains a boundary component of Σ. Next, we observe that Dehn twists along such C can realize any finite length l.

Proposition 4 *Let $C = c_1 \cup \cdots \cup c_r$ be a nonisolating multicurve on Σ and $m_1, \ldots, m_r$ nonnegative integers. Then the multitwist $\prod_{i=1}^{r} t_{c_i}^{m_i}$ has a unique factorization in Γ_g^n into positive Dehn twists and hence $\widetilde{\mathrm{L}}\big(\prod_{i=1}^{r} t_{c_i}^{m_i}\big) = m_1 + \cdots + m_r$.*

Proof Let $\Phi = \prod_{i=1}^{r} t_{c_i}^{m_i}$ with $m_1, \ldots, m_r$ nonnegative integers as above. By Proposition 3, for any positive factorization of Φ, each factor fixes every arc which is disjoint from C. Since C is nonisolating, we can find arcs disjoint from C which cut Σ into disjoint annuli that respectively deformation retract onto $\bigcup c_i$. Thus all factors are supported on these annuli, must be Dehn twists along these annuli, and give the obvious factorization $\prod_{i=1}^{n} t_{c_i}^{m_i}$. $\square$

Lastly, we note a general source of mapping classes in Γ_g^n, $n \geq 1$, with no positive factorizations.

Proposition 5 *If Φ is a nontrivial element in Veer_g^n, where $n \geq 1$, then Φ^{-1} is not. Thus if Φ admits a nontrivial positive factorization, then Φ^{-1} is not right-veering. In particular, $\mathrm{L}(\Delta^k) = \widetilde{\mathrm{L}}(\Delta^k) = -\infty$ for $k < 0$ and $\mathrm{L}(1) = \widetilde{\mathrm{L}}(1) = 0$.*

Proof If Φ is nontrivial and right-veering, then it moves at least one arc α to the right. Then Φ^{-1} sends $\Phi(\alpha)$ to α; that is, to the left. $\square$

The particular case we noted above, that Δ^k for $k \leq 0$ does not admit any nontrivial positive factorizations, was first observed by Smith [25] (whose arguments are similar to ours) and also by Stipsicz [27] (who used Seiberg–Witten theory).

2.2 Boundary multitwists with finite lengths

We are going to prove:

Theorem 6 *Let $\Delta = t_{\delta_1} \cdot t_{\delta_2} \cdots t_{\delta_n}$ be the boundary multitwist on Σ_g^n, with $n \geq 2g-3 \geq 0$. If $n > 4g+4$, then $\mathrm{L}(\Delta) = -\infty$. If $n \leq 4g+4$, we have*

$$\mathrm{L}(\Delta) = \begin{cases} 40 & \text{if } g = 2, \\ 6g+18 & \text{if } 3 \leq g \leq 6, \\ 8g+4 & \text{if } g \geq 7. \end{cases}$$

In particular, when $\mathrm{L}(\Delta)$ is finite, its value depends solely on g, and not n.

Let us briefly review here the notion of *symplectic Kodaira dimension* we will repeatedly refer to in our proof of this theorem. The reader can turn to [13] for more details. First, we recall that a symplectic 4–manifold (X, ω) is called *minimal* if it does not contain any embedded symplectic sphere of square -1, and also that it can always be blown-down to a minimal symplectic 4–manifold $(X_{\mathrm{min}}, \omega')$. Let $\kappa_{X_{\mathrm{min}}}$ be the canonical class of a minimal model $(X_{\mathrm{min}}, \omega_{\mathrm{min}})$. We define the symplectic Kodaira dimension of (X, ω), denoted by $\kappa = \kappa(X, \omega)$, as follows:

$$\kappa(X,\omega) = \begin{cases} -\infty & \text{if } \kappa_{X_{\mathrm{min}}} \cdot [\omega_{\mathrm{min}}] < 0 \text{ or } \kappa^2_{X_{\mathrm{min}}} < 0, \\ 0 & \text{if } \kappa_{X_{\mathrm{min}}} \cdot [\omega_{\mathrm{min}}] = \kappa^2_{X_{\mathrm{min}}} = 0, \\ 1 & \text{if } \kappa_{X_{\mathrm{min}}} \cdot [\omega_{\mathrm{min}}] > 0 \text{ and } \kappa^2_{X_{\mathrm{min}}} = 0, \\ 2 & \text{if } \kappa_{X_{\mathrm{min}}} \cdot [\omega_{\mathrm{min}}] > 0 \text{ and } \kappa^2_{X_{\mathrm{min}}} > 0. \end{cases}$$

Here κ is independent of the minimal model $(X_{\mathrm{min}}, \omega_{\mathrm{min}})$ and is a smooth invariant of the 4–manifold X.

Proof of Theorem 6 Assume that Δ admits a positive Dehn twist factorization $W = t_{c_l} \cdots t_{c_1}$ along nonseparating curves c_i in Γ_g^n. Let (X, f) be the genus-g Lefschetz fibration with n disjoint (-1)–sphere sections, $S_1, \dots, S_n$, associated to this factorization. We can support (X, f) with a symplectic form ω, with respect to which all S_j are symplectic as well. Note that by the hypothesis, $g \geq 2$ and $n \geq 1$, the latter implying that X is *not* minimal.

For fixed g, the length l is maximized if and only if the Euler characteristic of X is, where $\mathrm{e}(X) = 4-4g+l$; whereas fixing n, along with g, will play a role in narrowing down the possible values of the symplectic Kodaira dimension $\kappa(X)$. We will read off $\kappa(X)$ based on the number of (-1)–sphere sections of f. In principle, we need to know that there are no other disjoint (-1)–sphere sections than $S_1, \dots, S_n$, that is, there are no lifts of the positive factorization to a boundary multitwist in $\Gamma_g^{n'}$ with

$n' > n$. We will overcome this issue by simply presenting our arguments starting with $\kappa = -\infty$ and going up to nonnegative $\kappa = 0$ cases. (Meanwhile, it will become evident that the $\kappa = 1$ and 2 cases cannot occur, so the proof will boil down to realizing and comparing the bounds we obtain in the $\kappa = -\infty$ and 0 cases.)

If $n > 2g-2$, we can blow-down the n (-1)–sphere sections $S_1, \dots, S_n$ to derive a symplectic surface F' from a regular fiber F of f, which has genus g and self-intersection n. Since the Seiberg–Witten adjunction inequality

$$2g-2 = -\mathrm{e}(F') \geq [F']^2 + |\beta \cdot F'| \geq [F']^2 = n$$

is violated by F', we conclude that X' (and thus X) should be a rational or ruled surface [16]. These are precisely the symplectic 4–manifolds with Kodaira dimension $\kappa = -\infty$.

If $n = 2g-2$, and X is not rational or ruled, it follows from Sato's work on the canonical class of genus $g \geq 2$ Lefschetz fibrations on nonminimal symplectic 4–manifolds that the canonical class K_X can be represented by the sum $\sum_{j=1}^n S_j$ of the exceptional sphere sections *in* $H_2(X;\mathbb{Q})$ (see [23] and also [3]).[2] Blowing down all S_j we get $K_{X'} = 0$ in $H_2(X;\mathbb{Q})$. In particular, the canonical class is torsion, and so X is a blow-up of a symplectic Calabi–Yau surface, $\kappa(X) = 0$. The minimal model of X should then have the rational homology type of a torus bundle over a torus, the Enrique surface, or the K3 surface by the work of Li and independently of Bauer [2; 14; 15].

Now if $n = 2g-3$, and X is not rational or ruled or a (blow-up of a) symplectic Calabi–Yau surface, then the collection $\sum_{j=1}^n S_j$ realizes the maximal disjoint collection of representatives of its exceptional classes intersecting the fiber. It therefore follows from Sato's work in [23] that, *provided* $g \geq 3$ for the genus-g Lefschetz fibration on X, the canonical class of X is represented by $2S_1 + \sum_{j=2}^n S_j + R$, where S_1 is a distinguished S_j we get by relabeling if necessary, and more importantly, R is a genus 1 irreducible component of a reducible fiber with $[R]^2 = -1$. The latter condition however is not realized by any Lefschetz fibration with only nonseparating vanishing cycles, which allows us to rule out this case. Finally, in the remaining $g = 2$ and $n = 1$ case, it follows from Smith's analysis of genus-2 pencils in [26, Theorem 5.5] that the maximal number of irreducible singular fibers is $l = 40$.

We have thus seen that for $n \geq 2g-3$, the $\kappa = 2$ and $\kappa = 1$ cases are already ruled out. It therefore suffices to discuss the $\kappa = -\infty$ and 0 cases, and compare the largest l we

[2] As the author's argument in [23] is based on positive intersections of holomorphic curves, it essentially captures the homology class of K_X only in $\mathbb{Q}$ coefficients, which otherwise would lead to a contradiction for pencils on the Enriques surface. Hence we quote here the result with this small correction [3].

get in these cases to determine the winner, all while remembering we have an additional candidate in the $g=2$ case as noted above.

Let $\kappa(X)=-\infty$. As X is not minimal, and because $\mathbb{CP}^2\#\overline{\mathbb{CP}}^2$ does not admit any genus $g>0$ Lefschetz fibration with a (-1)–section,[3] we have $X\cong S^2\times\Sigma_h\# m\overline{\mathbb{CP}}^2$ for some $h\geq 0$, $m\geq 1$. We have $\mathrm{e}(X)=2(2-2h)+m=4-4g+l$, so

$$l=4(g-h)+m\leq 4g+m. \tag{1}$$

On the other hand, it was shown in [28] that $4(b_1(X)-g)+b_2^-(X)\leq 5b_2^+(X)$. For X with $b_1(X)=2h$, $b_2^+(X)=1$, and $b_2^-(X)=m+1$, we get

$$m\leq 4-4(2h-g)\leq 4+4g. \tag{2}$$

Combining the inequalities (1) and (2), we conclude that $l\leq 8g+4$.

The first conclusion, namely that $\mathrm{L}(\Delta)=-\infty$ when $n>4g+4$, is rather immediate. Here $X\cong S^2\times\Sigma_h\# m\overline{\mathbb{CP}}^2$, and either $h>0$ and we have $m\geq n$ or $h=0$ and we only have $m+1\geq n$. The inequality (2) above, combined with our assumption $n>4g+4$, implies that the former is impossible, whereas the latter can hold only if $m=4g+4$.

We claim that there is no genus-g Lefschetz pencil on $\mathbb{CP}^2$ with $4g+5$ base points. Let H represent the generator of $H_2(\mathbb{CP}^2)$, and $F=aH$ represent the potential fiber class of the pencil. Since there exists a symplectic form for which the fiber is symplectic, and since $\mathbb{CP}^2$ has a unique symplectic structure up to deformations and symplectomorphisms, we can invoke the adjunction equality as

$$2g-2=F^2-3H\cdot F=4g+5-3a,$$

so $3a=2g+7$. As F is a fiber class, $a^2=F^2=4g+5$ should be satisfied as well. The only possible solution is when $a=3$, which is the case of a $g=1$ pencil (indeed, the well-known case of an elliptic pencil on $\mathbb{CP}^2$) we have excluded from our discussion (recall our hypothesis $n\geq 2g-3\geq 0$).

Moreover, note that the equality $l=8g+4$ holds only if $m=4g+4$ and $h=0$. There exist such genus-g Lefschetz fibrations with $l=8g+4$ irreducible fibers and $m=4g+4$ sections of square -1 on $\mathbb{CP}^2\#(4g+5)\overline{\mathbb{CP}}^2\cong S^2\times S^2\#(4g+4)\overline{\mathbb{CP}}^2$; see [30; 22].

Now, let $\kappa(X)=0$. Recall that l is maximal when $\mathrm{e}(X)$ is. Rational cohomology K3 has the largest Euler characteristics among all minimal candidates, and as discussed

[3] If it did, one would get a homology class $F=aH$ with $F^2=1$, where H is the generator of $H_2(\mathbb{CP}^2)$. This is only possible when $a=1$, which implies that the fiber genus is zero.

above, one can hope to have a genus-g Lefschetz fibration on at most $2g-2$ blow-ups of a symplectic Calabi–Yau surface. It follows that the maximal l is realized when X is a rational cohomology K3 surface blown up $2g-2$ times. So $2(2-2g)+l=24+2g-2$, implying $l=6g+18$. Such Lefschetz fibrations on symplectic Calabi–Yau K3 surfaces are constructed in [4]; also see [26, Proof of Theorem 3.10].

Hence all remaining conclusions of the theorem follow from a comparison of the maximal l we get in the $\kappa=-\infty$ and $\kappa=0$ cases, along with the additional ($\kappa=2$) case when $g=2$. □

Remark 7 As seen in our proof, there is an a priori upper bound, determined by the genus g and the number of base points n, on the number of critical points of Lefschetz pencils when $\kappa<1$, and for pencils with only irreducible fibers when $\kappa\leq 1$. So arbitrarily large topology is specific to pencils on *symplectic* 4*–manifolds of general type*, ie when $\kappa=2$. In contrast, when the uniform topology is bounded, the maximal Euler characteristic for a genus-g Lefschetz pencil with n base points can be realized by an (X,f) with $\kappa=2$ when $g=2$, $\kappa=0$ when $3\leq g\leq 7$, and $\kappa=-\infty$ when $g\geq 7$.

3 Mapping classes with infinite lengths

Here we will construct arbitrarily long positive factorizations of various mapping classes involving boundary multitwists in Γ_g^n, for $g\geq 2$, $n\geq 1$.

3.1 Preliminary results

We begin with a brief exposition of various recent results on arbitrarily long positive factorizations in [5; 6; 7; 8], which creates leverage for many of our results to follow. We hope that the proofs given below will help with making the current article self-contained in this aspect.

First examples of arbitrarily long positive factorizations were produced in [5] by Korkmaz and the first two authors of this article, for a *varying family* of single commutators in Γ_g^2, for any $g\geq 2$. The proof of this result is based on the following well-known relations. Let $c_1,c_2,\dots,c_{2h+1}$ be simple closed curves on Σ_g^n such that c_i and c_j are disjoint if $|i-j|\geq 2$ and that c_i and c_{i+1} intersect at one point. Then, a regular neighborhood of $c_1\cup c_2\cup\cdots\cup c_{2h+1}$ is a subsurface of genus h with two boundary components, b_1 and b_2. We then have the *chain relations*,

$$t_{b_1}t_{b_2}=(t_{c_1}t_{c_2}\cdots t_{c_{2h+1}})^{2h+2}=(t_{c_{2h+1}}\cdots t_{c_2}t_{c_1})^{2h+2}.$$

Now for a chain of length 3, we get $t_d t_e = (t_{c_1} t_{c_2} t_{c_3})^4$ and by applying the relation (6) below to $(t_{c_1} t_{c_2} t_{c_3})^2$, we obtain

$$t_d t_e = (t_{c_1} t_{c_2} t_{c_3})^2 t_{c_2} t_{c_3} t_{c_1} t_{c_2} t_{c_3} t_{c_3}.$$

Since d, e, c_1 and c_3 are disjoint, we have

$$(t_{c_1} t_{c_2} t_{c_3})^2 t_{c_2} t_{c_1} t_{c_3} t_{c_2} = t_d t_{c_3}^{-1} t_e t_{c_3}^{-1}.$$

Taking the m^{th} power of both sides, we obtain $T_{10m} = t_d^m t_{c_3}^{-m} t_e^m t_{c_3}^{-m}$ for any positive integer m, where $T_{10m} = \{(t_{c_1} t_{c_2} t_{c_3})^2 t_{c_2} t_{c_1} t_{c_3} t_{c_2}\}^m$. Now let

$$\phi_{12} = t_{c_4} t_{c_3} t_{c_2} t_{c_1} t_{c_1} t_{c_2} t_{c_3} t_{c_4} t_{c_4} t_d t_{c_3} t_{c_4}.$$

Since $\phi_{12}(c_3) = e$ and $\phi_{12}(d) = c_3$, we get

$$\begin{aligned}
\phi_{12} &= \phi_{12} t_d^{-m} t_{c_3}^m t_e^{-m} t_{c_3}^m T_{10m} \\
&= \phi_{12} t_d^{-m} t_{c_3}^m \phi_{12}^{-1} \phi_{12} t_e^{-m} t_{c_3}^m T_{10m} \\
&= t_{\phi_{12}(d)}^{-m} t_{\phi_{12}(c_3)}^m \phi_{12} t_e^{-m} t_{c_3}^m T_{10m} \\
&= t_{c_3}^{-m} t_e^m \phi_{12} t_e^{-m} t_{c_3}^m T_{10m}.
\end{aligned}$$

We thus obtain the commutator relation in [5],

$$C_m = [\phi_{12}, t_{c_3}^m t_e^m] = T_{10m},$$

the right-hand side of which contains arbitrarily long positive factorizations as m increases.

These commutator relations prescribe a family of genus-2 Lefschetz fibrations over T^2 with sections of self-intersection zero. Taking the complement of the regular fiber and the section, the first and the third authors of this article produced allowable Lefschetz fibrations filling a fixed spinal open book, leading to the first examples of contact 3–manifolds with arbitrarily large Stein fillings and arbitrarily negative signatures [6]. Guided by these examples, in a subsequent work [7], the same authors produced the first examples of mapping classes with arbitrarily long positive factorizations. They showed that *any* family of commutators $C_m = [A_m, B_m]$ with arbitrarily long positive factorizations can be crafted into arbitrarily long positive factorizations of the boundary multitwist $t_{\delta_1} t_{\delta_2}$ in Γ_g^2, for $g \geq 8$.

The arguments of [7] were taken further in an elegant article by Dalyan, Korkmaz and Pamuk in [8], who observed that for special commutators $C_m = [A, B_m]$, where one entry is a fixed mapping class, as in the commutator relation we reproduced above, one can manipulate the relations so as to produce arbitrarily long positive factorizations in Γ_2^2. Namely, by repeating the relation (6), we have

$$\begin{aligned}(t_{c_1}t_{c_2}t_{c_3}t_{c_4}t_{c_5})^6 &= (t_{c_1}t_{c_2}t_{c_3}t_{c_4})^5 t_{c_5}t_{c_4}t_{c_3}t_{c_2}t_{c_1}t_{c_1}t_{c_2}t_{c_3}t_{c_4}t_{c_5}\\ &= t_{c_1}t_{c_2}t_{c_3}t_{c_4}(t_{c_1}t_{c_2}t_{c_3}t_{c_4})^4 t_{c_5}t_{c_4}t_{c_3}t_{c_2}t_{c_1}t_{c_1}t_{c_2}t_{c_3}t_{c_4}t_{c_5}\\ &= t_{c_1}t_{c_2}t_{c_3}t_{c_4}(t_{c_1}t_{c_2}t_{c_3})^4 t_{c_4}t_{c_3}t_{c_2}t_{c_1}t_{c_5}t_{c_4}t_{c_3}t_{c_2}t_{c_1}t_{c_1}t_{c_2}t_{c_3}t_{c_4}t_{c_5}.\end{aligned}$$

As t_δ and $t_{\delta'}$ are center elements of Γ_2^2, by the chain relations $t_\delta t_{\delta'} = (t_{c_1}t_{c_2}t_{c_3}t_{c_4}t_{c_5})^6$ and $t_d t_e = (t_{c_1}t_{c_2}t_{c_3})^4$ we obtain

$$\begin{aligned}t_\delta t_{\delta'} &= t_{c_1}t_{c_2}t_{c_3}t_{c_4}t_d t_e t_{c_4}t_{c_3}t_{c_2}t_{c_1}t_{c_5}\cdot t_{c_4}t_{c_3}t_{c_2}t_{c_1}t_{c_1}t_{c_2}t_{c_3}t_{c_4}\cdot t_{c_5}\\ &= t_{c_4}t_{c_3}t_{c_2}t_{c_1}t_{c_1}t_{c_2}t_{c_3}t_{c_4}\cdot t_{c_5}\cdot t_{c_1}t_{c_2}t_{c_3}t_{c_4}t_d t_e t_{c_4}t_{c_3}t_{c_2}t_{c_1}t_{c_5}\\ &= t_{c_4}t_{c_3}t_{c_2}t_{c_1}t_{c_1}t_{c_2}t_{c_3}t_{c_4}\cdot t_{c_4}t_d t_{c_3}\cdot D_9\\ &= D_9\cdot t_{c_4}t_{c_3}t_{c_2}t_{c_1}t_{c_1}t_{c_2}t_{c_3}t_{c_4}\cdot t_{c_4}t_d t_{c_3},\end{aligned}$$

where $D_9 = t_{(t_{c_4}t_d t_{c_3})^{-1}(c_5)}t_{c_1}t_{t_{c_3}^{-1}(c_2)}t_{(t_{c_4}t_d t_{c_3})^{-1}(c_3)}t_e t_{t_{c_3}^{-1}(c_4)}t_{c_2}t_{c_1}t_{c_5}$. By multiplying both sides of this relation by t_{c_4}, we obtain

$$t_\delta t_{\delta'} t_{c_4} = D_9\cdot\phi_{12,m}\cdot T_{10m}.$$

We sum these up in the following theorem.

Theorem 8 [5; 8] *Let d, e and c_i, $i=1,2,3,4,5$, be the simple closed curves on Σ_2^2 as in Figure 1, and let*

$$\begin{aligned}\phi_{12} &= t_{c_4}t_{c_3}t_{c_2}t_{c_1}t_{c_1}t_{c_2}t_{c_3}t_{c_4}t_{c_4}t_d t_{c_3}t_{c_4},\\ \phi_{12,m} &= t_{c_3}^{-m}t_e^m\phi_{12}t_e^{-m}t_{c_3}^m,\\ T_{10m} &= \{(t_{c_1}t_{c_2}t_{c_3})^2 t_{c_2}t_{c_1}t_{c_3}t_{c_2}\}^m,\\ D_9 &= t_{(t_{c_4}t_d t_{c_3})^{-1}(c_5)}t_{c_1}t_{t_{c_3}^{-1}(c_2)}t_{(t_{c_4}t_d t_{c_3})^{-1}(c_3)}t_e t_{t_{c_3}^{-1}(c_4)}t_{c_2}t_{c_1}t_{c_5}.\end{aligned}$$

Then, for all positive integers m, the following relations hold in Γ_2^2 :

$$\phi_{12} = \phi_{12,m}\cdot T_{10m}, \qquad \text{(Baykur, Korkmaz and Monden)} \tag{3}$$

$$t_\delta t_{\delta'} t_{c_4} = D_9\cdot\phi_{12,m}\cdot T_{10m}. \qquad \text{(Dalyan, Korkmaz and Pamuk)} \tag{4}$$

Finally, let us recall the following generalization of the *lantern relation*, now often called the *daisy relation* [20; 5] (also see [18; 21]). This relation will be key for inflating number of boundary components, and extending Theorem 8 to Γ_g^n for any $1\le n\le 2g-4$,

$$t_{\delta_0}^{p-1}t_{\delta_1}t_{\delta_2}\cdots t_{\delta_{p+1}} = t_{x_1}t_{x_2}\cdots t_{x_{p+1}},$$

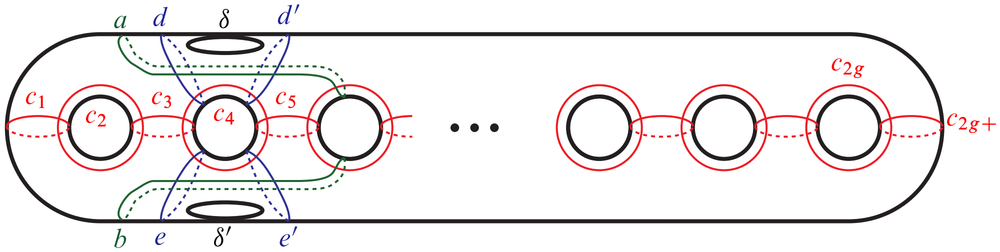

Figure 1: The curves $c_1, c_2, \dots, c_{2g+1}$ and a, b, d, d', e, e' and the boundary curves δ, δ' on Σ_g^2

in Γ_0^{p+2}, the mapping class group of a 2–sphere with $p+2 \geq 4$ boundary components. Here $\delta_0, \delta_1, \delta_2, \dots, \delta_{p+1}$ denote the $p+2$ boundary curves of Σ_0^{p+2}, and $x_1, x_2, \dots, x_{p+1}$ are the interior curves as shown in Figure 2. The $p=2$ case is the usual lantern relation.

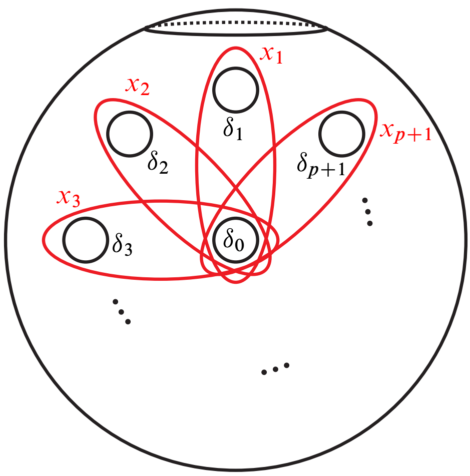

Figure 2: The curves $\delta_0, \delta_1, \dots, \delta_{p+1}$ and $x_1, x_2, \dots, x_{p+1}$

3.2 Boundary multitwist of infinite length

Theorem 9 *Let $g \geq 3$. Then, in Γ_g^{2g-4}, the multitwist*

$$t_{\delta_1} t_{\delta_2} \cdots t_{\delta_{2g-4}}$$

can be written as a product of arbitrarily large number of right-handed Dehn twists about nonseparating curves.

Let a, b, d, d', e, e' and c_i $(i = 1, 2, \dots, 2g+1)$ be the simple closed curves on Σ_g^2, and let δ and δ' be the two boundary curves of Σ_g^2 as in Figure 1.

We will now introduce the key lemma for the proofs of Theorems 9 and 16, namely Lemma 10.

Let l be a positive integer such that $l \leq n$. Let β and α, α' be the separating curve and the nonseparating curves on Σ_g^n in Figure 3, respectively. Note that β separates Σ_g^n into a surface of genus g with one boundary β and a sphere with $l+1$ boundaries $d, \delta_1, \delta_2, \dots, \delta_l$ and that α and α' separate Σ_g^n into a surface of genus $g-1$ with 2 boundaries α and α' and a sphere with $l+2$ boundaries $\alpha, \alpha', \delta_1, \delta_2, \dots, \delta_l$. Let $x_1, x_2, \dots, x_l$ be the nonseparating curves on Σ_g^n in Figure 3.

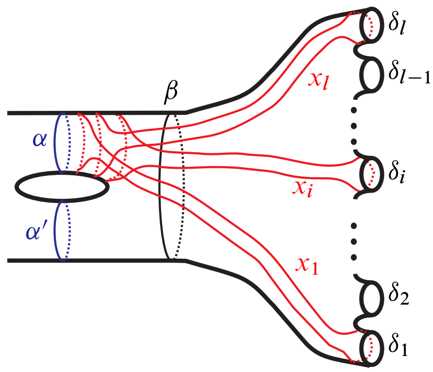

Figure 3: The curves α, α', β and $x_1, x_2, \dots, x_l$

Lemma 10 *Suppose that the following relation holds in Γ_g^n:*

$$U \cdot t_\beta = T \cdot t_\alpha^{l-1} t_{\alpha'},$$

where U and T are elements in Γ_g^n. Then, the following relation holds in Γ_g^n:

$$U \cdot t_{\delta_1} t_{\delta_2} \cdots t_{\delta_l} = T \cdot t_{x_1} \cdots t_{x_l}.$$

This is a generalization of a technical lemma from [4], which we provide a different proof of below.

Proof Multiplying both sides of the relation $U \cdot t_\beta = T \cdot t_\alpha^{l-1} t_{\alpha'}$ by $\delta_1 \delta_2 \cdots \delta_l$, we obtain the following relation:

$$U \cdot t_\beta t_{\delta_1} t_{\delta_2} \cdots t_{\delta_l} = T \cdot t_\alpha^{l-1} t_{\alpha'} t_{\delta_1} t_{\delta_2} \cdots t_{\delta_l}.$$

Since $t_{\delta_1}, t_{\delta_2}, \dots, t_{\delta_l}$ are elements in the center of Γ_g^n, we can rewrite this relation as follows:

$$U \cdot t_{\delta_1} t_{\delta_2} \cdots t_{\delta_l} t_\beta = T \cdot t_\alpha^{l-1} t_{\delta_1} t_{\delta_2} \cdots t_{\delta_l} t_{\alpha'}.$$

Here, by the daisy relation $t_\alpha^{l-1} t_{\delta_1} t_{\delta_2} \cdots t_{\delta_l} t_{\alpha'} = t_{x_1} t_{x_2} \cdots t_{x_l} t_\beta$, we have

$$U \cdot t_{\delta_1} t_{\delta_2} \cdots t_{\delta_l} t_\beta = T \cdot t_{x_1} t_{x_2} \cdots t_{x_l} t_\beta.$$

Removing t_β from both sides of this relation we get the desired relation. □

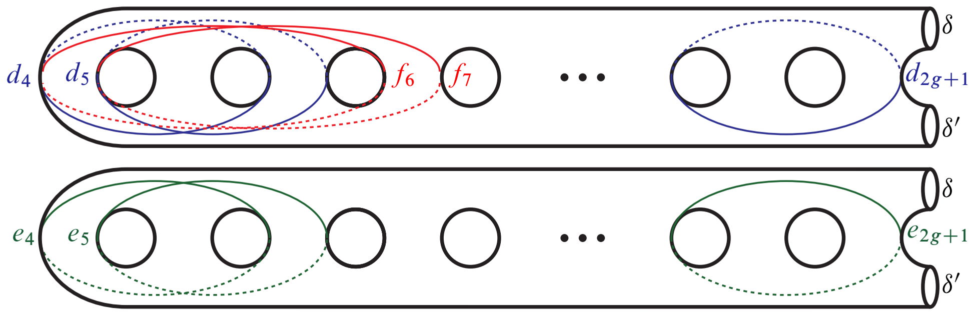

Figure 4: The curves d_j, e_j $(j = 4, 5, \ldots, 2g+1)$ and f_h $(h = 6, 7, 8, 9)$ on Σ_g^2

Let d_j, e_j $(j = 4, 5, \ldots, 2g+1)$, f_h $(h = 6, 7, 8, 9)$ be the simple closed curves on Σ_g^2 as in Figure 4 which are defined by

$$d_j = t_{c_{j-3}}^{-1} t_{c_{j-2}}^{-1} t_{c_{j-1}}^{-1}(c_j), \qquad e_j = t_{c_{j-3}} t_{c_{j-2}} t_{c_{j-1}}(c_j),$$
$$f_h = t_{c_{h-5}} t_{c_{h-4}} t_{c_{h-3}} t_{c_{h-2}} t_{c_{h-1}}(c_h).$$

Letting i, l, m be positive integers such that $l+1 \le i \le m-1$, the following relations hold from the braid relations:

$$t_{c_{i-1}} \cdot t_{c_m} t_{c_{m-1}} \cdots t_{c_l} = t_{c_m} t_{c_{m-1}} \cdots t_{c_l} \cdot t_{c_i}, \tag{5}$$
$$t_{c_l} t_{c_{l+1}} \cdots t_{c_m} \cdot t_{c_{i-1}} = t_{c_i} \cdot t_{c_l} t_{c_{l+1}} \cdots t_{c_m}. \tag{6}$$

Next is a lemma from [4], whose proof we include here for completeness.

Lemma 11 [4] *For $k = 1, 2, \ldots, 2g-2$, the following relations hold in Σ_g^2:*

$$\prod_{i=k}^{2g-2} t_{c_{i+3}} t_{c_{i+2}} t_{c_{i+1}} t_{c_i} = (t_{c_{k+2}} t_{c_{k+1}} t_{c_k})^{2g-1-k} t_{d_{k+3}} t_{d_{k+4}} \cdots t_{d_{2g+1}} \tag{7}$$

$$\prod_{i=2g-2}^{k} t_{c_i} t_{c_{i+1}} t_{c_{i+2}} t_{c_{i+3}} = t_{e_{2g+1}} \cdots t_{e_{k+4}} t_{e_{k+3}} (t_{c_k} t_{c_{k+1}} t_{c_{k+2}})^{2g-1-k} \tag{8}$$

$$\prod_{i=4}^{1} t_{c_i} t_{c_{i+1}} t_{c_{i+2}} t_{c_{i+3}} t_{c_{i+4}} t_{c_{i+5}} = t_{f_9} t_{f_8} t_{f_7} t_{f_6} (t_{c_1} t_{c_2} t_{c_3} t_{c_4} t_{c_5})^4 \tag{9}$$

Proof First, we prove the relation (7) by induction on $2g-1-k$. Suppose that $k=2g-2$. Then, we have

$$t_{c_{2g+1}} \cdot t_{c_{2g}} t_{c_{2g-1}} t_{c_{2g-2}} = t_{c_{2g}} t_{c_{2g-1}} t_{c_{2g-2}} \cdot t_{d_{2g+1}}.$$

Therefore, the conclusion of the relation holds for $k=1$. Let us assume, inductively, that the relation holds for $k+1<2g-2$. By (5), we have

$$\begin{aligned}
&\prod_{i=k}^{2g-2} t_{c_{i+3}} t_{c_{i+2}} t_{c_{i+1}} t_{c_i} \\
&\qquad = t_{c_{k+3}} t_{c_{k+2}} t_{c_{k+1}} t_{c_k} \cdot \prod_{i=k+1}^{2g-2} t_{c_{i+3}} t_{c_{i+2}} t_{c_{i+1}} t_{c_i} \\
&\qquad = t_{c_{k+3}} t_{c_{k+2}} t_{c_{k+1}} t_{c_k} \cdot (t_{c_{k+3}} t_{c_{k+2}} t_{c_{k+1}})^{2g-1-(k+1)} t_{d_{k+4}} t_{d_{k+5}} \cdots t_{d_{2g+1}} \\
&\qquad = (t_{c_{k+2}} t_{c_{k+1}} t_{c_k})^{2g-1-(k+1)} \cdot t_{c_{k+3}} t_{c_{k+2}} t_{c_{k+1}} t_{c_k} \cdot t_{d_{k+4}} t_{d_{k+5}} \cdots t_{d_{2g+1}} \\
&\qquad = (t_{c_{k+2}} t_{c_{k+1}} t_{c_k})^{2g-1-(k+1)} \cdot t_{c_{k+2}} t_{c_{k+1}} t_{c_k} \cdot t_{d_{k+3}} \cdot t_{d_{k+4}} t_{d_{k+5}} \cdots t_{d_{2g+1}}.
\end{aligned}$$

Hence, the relation (7) is proved.

Next, we prove the relation (8) by induction on $2g-1-k$. Suppose that $k=2g-2$. Then, we have

$$t_{c_{2g-2}} t_{c_{2g-1}} t_{c_{2g}} \cdot t_{c_{2g+1}} = t_{e_{2g+1}} \cdot t_{c_{2g-2}} t_{c_{2g-1}} t_{c_{2g}}.$$

Therefore, the conclusion of the relation holds for $k=2g-2$. Let us assume, inductively, that the relation holds for $k+1<2g-2$. By (6), we have

$$\begin{aligned}
&\prod_{i=2g-2}^{k} t_{c_i} t_{c_{i+1}} t_{c_{i+2}} t_{c_{i+3}} \\
&\qquad = \left(\prod_{i=2g-2}^{k+1} t_{c_i} t_{c_{i+1}} t_{c_{i+2}} t_{c_{i+3}} \right) \cdot t_{c_k} t_{c_{k+1}} t_{c_{k+2}} t_{c_{k+3}} \\
&\qquad = t_{e_{2g+1}} \cdots t_{e_{k+5}} t_{e_{k+4}} (t_{c_{k+1}} t_{c_{k+2}} t_{c_{k+3}})^{2g-2-k} \cdot t_{c_k} t_{c_{k+1}} t_{c_{k+2}} t_{c_{k+3}} \\
&\qquad = t_{e_{2g+1}} \cdots t_{e_{k+5}} t_{e_{k+4}} \cdot t_{c_k} t_{c_{k+1}} t_{c_{k+2}} t_{c_{k+3}} \cdot (t_{c_k} t_{c_{k+1}} t_{c_{k+2}})^{2g-2-k} \\
&\qquad = t_{e_{2g+1}} \cdots t_{e_{k+5}} t_{e_{k+4}} \cdot t_{e_{k+3}} \cdot t_{c_k} t_{c_{k+1}} t_{c_{k+2}} \cdot (t_{c_k} t_{c_{k+1}} t_{c_{k+2}})^{2g-2-k}.
\end{aligned}$$

Hence we obtain the relation (8). The proof for the relation (9) is very similar, and we leave it to the reader. □

Let

$$H_i := t_{c_1} t_{c_2} \cdots t_{c_i}, \qquad \overline{H}_i := t_{c_i} \cdots t_{c_2} t_{c_1}, \quad I_{2g-8} := t_{d_{10}} t_{d_{11}} \cdots t_{d_{2g+1}},$$

$$J_{2g-6} := t_{e_{2g+1}} \cdots t_{e_9} t_{e_8}, \quad K_4 := t_{f_9} t_{f_8} t_{f_7} t_{f_6}, \qquad L_{16} := \prod_{i=1}^{4} t_{c_{i+3}} t_{c_{i+2}} t_{c_{i+1}} t_{c_i}.$$

Lemma 12 *For $g \geq 4$, the following relation holds in Σ_g^2:*

$$(t_{c_1} t_{c_2} \cdots t_{c_{2g+1}})^4 = (H_3)^4 (\overline{H}_3)^{2g-8} K_4 (H_5)^4 I_{2g-8}$$

Proof It is easy to check that from the braid relations we have

$$(t_{c_1} t_{c_2} \cdots t_{c_{2g+1}})^4 = (t_{c_1} t_{c_2} t_{c_3})^4 \prod_{i=4}^{1} t_{c_i} \cdots t_{c_{i+2g-3}} = (H_3)^4 \prod_{i=1}^{2g-2} t_{c_{i+3}} t_{c_{i+2}} t_{c_{i+1}} t_{c_i}.$$

By the relation (7) for $k = 7$ repeating (5), we obtain

$$\begin{aligned}(t_{c_1} t_{c_2} \cdots t_{c_{2g+1}})^4 &= (H_3)^4 \left(\prod_{i=1}^{6} t_{c_{i+3}} t_{c_{i+2}} t_{c_{i+1}} t_{c_i} \right) (t_{c_9} t_{c_8} t_{c_7})^{2g-8} I_{2g-8} \\ &= (H_3)^4 (t_{c_3} t_{c_2} t_{c_1})^{2g-8} \left(\prod_{i=1}^{6} t_{c_{i+3}} t_{c_{i+2}} t_{c_{i+1}} t_{c_i} \right) I_{2g-8}.\end{aligned}$$

It is easy to check that from the braid relations we have

$$\prod_{i=1}^{6} t_{c_{i+3}} t_{c_{i+2}} t_{c_{i+1}} t_{c_i} = \prod_{i=4}^{1} t_{c_i} t_{c_{i+1}} t_{c_{i+2}} t_{c_{i+3}} t_{c_{i+4}} t_{c_{i+5}}.$$

Hence, by (9), we obtain the desired relation. □

Lemma 13 *For $g \geq 4$, the following relation holds in Σ_g^2:*

$$(t_{c_1} t_{c_2} \cdots t_{c_{2g+1}})^{2g-2} = J_{2g-6} L_{16} (H_3)^{2g-6} t_{d'} t_{e'}$$

Proof From the braid relations we have

$$(t_{c_1} t_{c_2} \cdots t_{c_{2g+1}})^{2g-2} = \left(\prod_{i=2g-2}^{1} t_{c_i} t_{c_{i+1}} t_{c_{i+2}} t_{c_{i+3}} \right) (t_{c_5} t_{c_6} \cdots t_{c_{2g+1}})^{2g-2}.$$

By the chain relation $t_{d'}t_{e'} = (t_{c_5}t_{c_6}\cdots t_{c_{2g+1}})^{2g-2}$, the relation (8) for $k=5$ and repeating (6),

$$\begin{aligned}(t_{c_1}t_{c_2}\cdots t_{c_{2g+1}})^{2g-2} &= J_{2g-6}(t_{c_5}t_{c_6}t_{c_7})^{2g-6}\Biggl(\prod_{i=4}^{1} t_{c_i}t_{c_{i+1}}t_{c_{i+2}}t_{c_{i+3}}\Biggr)t_{d'}t_{e'}\\ &= J_{2g-6}\Biggl(\prod_{i=4}^{1} t_{c_i}t_{c_{i+1}}t_{c_{i+2}}t_{c_{i+3}}\Biggr)(t_{c_1}t_{c_2}t_{c_3})^{2g-6}t_{d'}t_{e'}.\end{aligned}$$

Since it is easy to check that from the braid relations we have

$$\prod_{i=4}^{1} t_{c_i}t_{c_{i+1}}t_{c_{i+2}}t_{c_{i+3}} = \prod_{i=1}^{4} t_{c_{i+3}}t_{c_{i+2}}t_{c_{i+1}}t_{c_i},$$

we obtain the relation in the statement. □

Proposition 14 *Let $g \geq 4$. Then the following relation holds in Σ_g^2. If g is even, then we have*

$$t_\delta t_{\delta'} = K_4(H_5)^4 I_{2g-8}J_{2g-6}L_{16}(H_3)^2 t_d^{g-3}t_{d'}t_e^{g-3}t_{e'}.$$

If g is odd, then we have

$$t_\delta t_{\delta'} = K_4(H_5)^4 I_{2g-8}J_{2g-6}L_{16}(\overline{H}_3)^2 t_d^{g-3}t_{d'}t_e^{g-3}t_{e'}.$$

Proof By Lemmas 12, 13 and the chain relation

$$t_\delta t_{\delta'} = (t_{c_1}t_{c_2}\cdots t_{c_{2g+1}})^{2g+2} = (t_{c_1}t_{c_2}\cdots t_{c_{2g+1}})^4\cdot(t_{c_1}t_{c_2}\cdots t_{c_{2g+1}})^{2g-2},$$

we have

$$t_\delta t_{\delta'} = (H_3)^4(\overline{H}_3)^{2g-8}K_4(H_5)^4 I_{2g-8}J_{2g-6}L_{16}(H_3)^{2g-6}t_{d'}t_{e'}.$$

Since c_1, c_2, c_3 are disjoint from d' and e', by conjugation by $(H_3)^4(\overline{H}_3)^{2g-8}$ we obtain

$$t_\delta t_{\delta'} = K_4(H_5)^4 I_{2g-8}J_{2g-6}L_{16}(H_3)^{2g-2}(\overline{H}_3)^{2g-8}t_{d'}t_{e'}.$$

The claim follows from this relation and the chain relations $t_dt_e = (H_3)^4 = (\overline{H}_3)^4$. □

Lemma 15 *Let $g \geq 4$. For any positive integer m, we have*

$$L_{16}(H_5)^4 = \phi_{12,m}T_{10m}M_9\cdot t_{c_5}t_{c_3}t_{c_4}t_{c_2}t_{c_3},$$

where $M_9 = t_{(t_{c_4}t_dt_{c_3}t_{c_4})^{-1}(c_5)}t_{c_2}t_{(t_{c_5}t_{c_4})^{-1}(c_6)}t_dt_{t_{c_4}^{-1}(c_3)}t_{c_7}t_{(t_{c_5}t_{c_4})^{-1}(c_6)}t_dt_{t_{c_4}^{-1}(e)}$.

Proof By (6) and the braid relations, we have

$$\begin{aligned}
L_{16}(H_5)^4 &= t_{c_4}t_{c_3}t_{c_2}t_{c_1}H_5t_{c_4}t_{c_3}t_{c_2}t_{c_1}t_{t_{c_5}^{-1}(c_6)}t_{c_4}t_{c_3}t_{c_2}t_{c_7}t_{t_{c_5}^{-1}(c_6)}t_{c_4}t_{c_3}(H_5)^3 \\
&= t_{c_4}t_{c_3}t_{c_2}t_{c_1}H_5t_{c_4}t_{c_3}t_{c_2}t_{t_{c_5}^{-1}(c_6)}t_{c_4}t_{c_3}t_{c_7}t_{t_{c_5}^{-1}(c_6)}t_{c_4}\cdot t_{c_1}t_{c_2}t_{c_3}(H_5)^3 \\
&= t_{c_4}t_{c_3}t_{c_2}t_{c_1}H_5t_{c_4}t_{c_3}t_{c_2}t_{t_{c_5}^{-1}(c_6)}t_{c_4}t_{c_3}t_{c_7}t_{t_{c_5}^{-1}(c_6)}t_{c_4}(H_3)^4t_{c_4}t_{c_5}t_{c_3}t_{c_4}t_{c_2}t_{c_3}.
\end{aligned}$$

Here, by the chain relation $t_dt_e=(H_3)^4$,

$$\begin{aligned}
t_{c_2}t_{t_{c_5}^{-1}(c_6)}t_{c_4}t_{c_3}t_{c_7}t_{t_{c_5}^{-1}(c_6)}t_{c_4}(H_3)^4t_{c_4} &= t_{c_2}t_{t_{c_5}^{-1}(c_6)}t_{c_4}t_{c_3}t_{c_7}t_{t_{c_5}^{-1}(c_6)}t_{c_4}t_dt_et_{c_4} \\
&= t_{c_2}t_{t_{c_5}^{-1}(c_6)}t_{c_4}t_{c_3}t_{c_7}t_{t_{c_5}^{-1}(c_6)}t_{c_4}t_dt_{c_4}t_{t_{c_4}^{-1}(e)} \\
&= t_dt_{c_4}\cdot N_7\cdot t_{c_4}t_{t_{c_4}^{-1}(e)},
\end{aligned}$$

where $N_7=(t_dt_{c_4})^{-1}(t_{c_2}t_{t_{c_5}^{-1}(c_6)}t_{c_4}t_{c_3}t_{c_7}t_{t_{c_5}^{-1}(c_6)}t_{c_4})(t_dt_{c_4})$. Note that it is easy to check that $N_7=t_{c_2}t_{(t_{c_5}t_{c_4})^{-1}(c_6)}t_dt_{t_{c_4}^{-1}(c_3)}t_{c_7}t_{(t_{c_5}t_{c_4})^{-1}(c_6)}t_d$. Therefore, we have

$$\begin{aligned}
L_{16}(H_5)^4 &= t_{c_4}t_{c_3}t_{c_2}t_{c_1}H_5t_{c_4}t_{c_3}\cdot t_dt_{c_4}\cdot N_7\cdot t_{c_4}t_{t_{c_4}^{-1}(e)}\cdot t_{c_5}t_{c_3}t_{c_4}t_{c_2}t_{c_3} \\
&= t_{c_4}t_{c_3}t_{c_2}t_{c_1}\cdot t_{c_1}t_{c_2}t_{c_3}t_{c_4}t_{c_5}\cdot t_{c_4}t_dt_{c_3}t_{c_4}\cdot N_7\cdot t_{c_4}t_{t_{c_4}^{-1}(e)}\cdot t_{c_5}t_{c_3}t_{c_4}t_{c_2}t_{c_3} \\
&= t_{c_4}t_{c_3}t_{c_2}t_{c_1}\cdot t_{c_1}t_{c_2}t_{c_3}t_{c_4}\cdot t_{c_4}t_dt_{c_3}t_{c_4}\cdot M_9\cdot t_{c_5}t_{c_3}t_{c_4}t_{c_2}t_{c_3} \\
&= \phi_{12}\cdot M_9\cdot t_{c_5}t_{c_3}t_{c_4}t_{c_2}t_{c_3}.
\end{aligned}$$

The lemma follows from the relation (3). □

Proof of Theorem 9 Suppose that $g\geq 4$. Let S and S' be two spheres with $g-1$ boundary components, and let $\delta,\delta_1,\delta_2,\dots,\delta_{g-2}$ and $\delta',\delta_{g-1},\delta_g,\dots,\delta_{2g-4}$ denote the boundary curves of S and S', respectively. We attach S and S' to Σ_g^2 along δ and δ'. Then we obtain a compact oriented surface of genus g with $2g-4$ boundary components $\delta_1,\delta_2,\dots,\delta_{2g-4}$, denoted by Σ_g^{2g-4}. By Proposition 14 and Lemma 10, there are simple close curves $x_1,x_2,\dots,x_{2g-4}$ such that the following relations hold in Γ_g^{2g-4}. Let

$$Z_{g-2}:=t_{x_1}t_{x_2}\cdots t_{x_{g-2}},\quad W_{g-2}:=t_{x_{g-1}}t_{x_g}\cdots t_{x_{2g-4}}.$$

If g is even, then we have

$$t_{\delta_1}t_{\delta_2}\cdots t_{\delta_{2g-4}}=K_4(H_5)^4I_{2g-8}J_{2g-6}L_{16}(H_3)^2Z_{g-2}W_{g-2}.$$

If g is odd, then we have

$$t_{\delta_1}t_{\delta_2}\cdots t_{\delta_{2g-4}}=K_4(H_5)^4I_{2g-8}J_{2g-6}L_{16}(\overline{H}_3)^2Z_{g-2}W_{g-2}.$$

By conjugation by L_{16} and Lemma 15, we have the following relation. If g is even, then we have

$$t_{\delta_1}t_{\delta_2}\cdots t_{\delta_{2g-4}} = K'_4\phi_{12,m}T_{10m}M_9\cdot t_{c_5}t_{c_3}t_{c_4}t_{c_2}t_{c_3}\cdot I_{2g-8}J_{2g-6}(H'_3)^2 Z'_{g-2}W'_{g-2},$$

where $K'_4 = L_{16}K_4L_{16}^{-1}$, $H'_3 = L_{16}H_3L_{16}^{-1}$ $Z'_{g-2} = L_{16}Z_{g-2}L_{16}^{-1}$ and $W'_{g-2} = L_{16}W_{g-2}L_{16}^{-1}$. If g is odd, then we have

$$t_{\delta_1}t_{\delta_2}\cdots t_{\delta_{2g-4}} = K'_4\phi_{12,m}T_{10m}M_9\cdot t_{c_5}t_{c_3}t_{c_4}t_{c_2}t_{c_3}\cdot I_{2g-8}J_{2g-6}(\overline{H'}_3)^2 Z'_{g-2}W'_{g-2},$$

where $\overline{H'}_3 = L_{16}\overline{H}_3L_{16}^{-1}$.

Note that K'_4, H'_3, $\overline{H'}_3$, Z'_{g-2} and W'_{g-2} are also products of 4, 3, 3, $g-2$ and $g-2$ right-handed Dehn twists about nonseparating curves, respectively. Therefore, for any positive integer m, the mapping class $t_{\delta_1}t_{\delta_2}\cdots t_{\delta_{2g-4}}$ may can be written as a product of $6g+2+10m$ right-handed Dehn twists about nonseparating curves. This completes the proof. □

3.3 Factorizations of boundary multitwist and a single Dehn twist

Theorem 16 *Let $g \geq 2$. Let a be a nonseparating curve on Σ_g^n. Then, for any positive integer n, in the mapping class group Γ_g^n, the multitwist*

$$t_{\delta_1}t_{\delta_2}\cdots t_{\delta_n}t_a$$

can be written as a product of an arbitrarily large number of right-handed Dehn twists about nonseparating curves.

The proof of Theorem 16 is a direct application of Lemma 10 (and Theorem 8).

Proof of Theorem 16 By the relation (4) and Theorem 9 we may assume that $n \geq 3$. Let k be a positive integer, and recall that $T_{10m} = \{(t_{c_1}t_{c_2}t_{c_3})^2 t_{c_2}t_{c_1}t_{c_3}t_{c_2}\}^m$. By the chain relation $t_dt_e = (t_{c_1}t_{c_2}t_{c_3})^4$, we may write

$$T_{10\cdot 2(n-2)} = T\cdot (t_{c_1}t_{c_2}t_{c_3})^{4(n-2)} = T\cdot t_d^{n-2}t_e^{n-2} = T\cdot t_e^{n-2}t_d^{n-2},$$

where T is a product of $8(n-2)$ right-handed Dehn twists about nonseparating curves. Therefore, if $m > 2(n-2)$, then we can rewrite T_{10m} in Γ_2^2 as

$$T_{10m} = T_{10\{m-2(n-2)\}}\cdot T_{10\cdot\{2(n-2)\}} = O_{10m-11(n-2)}\cdot t_d^{n-2}, \tag{10}$$

where $O_{10m-11(n-2)} = T_{10\{m-2(n-2)\}}\cdot T\cdot t_e^{n-2}$, so it is a product of $10m-11(n-2)$ right-handed Dehn twists about nonseparating curves.

Let S be a sphere with n boundary components $\delta, \delta_1, \delta_2, \dots, \delta_{n-1}$. We attach S to Σ_g^2 along δ. Set $\delta' = \delta_n$. Then we obtain a compact oriented surface of genus g with n boundary components $\delta_1, \delta_2, \dots, \delta_n$, denoted by Σ_g^n. Note that we obtain a separating curve on Σ_g^n from δ. We continue to write δ for the resulting separating curve on Σ_g^n.

Suppose that $g = 2$. Let $D_8 = D_9 t_{c_5}^{-1}$. By the relations (4) and (10), we have

$$\begin{aligned} t_\delta t_{\delta_n} t_{c_4} &= D_8 \cdot t_{c_5} \cdot \phi_{12,m} \cdot O_{10m-11(n-2)} \cdot t_d^{n-2} \\ &= D_8 \cdot \phi'_{12,m} \cdot O_{10m-11(n-2)} \cdot t_d^{n-2} t_{c_5}, \end{aligned}$$

where $\phi'_{12,m} = t_{c_5} \phi_{12,m} t_{c_5}^{-1}$. Note that $O_{10m-11(n-2)} = t_{c_5} O_{10m-11(n-2)} t_{c_5}^{-1}$ since c_5 is disjoint from c_1, c_2, c_3. By Lemma 10, there are nonseparating curves $x_1, \dots, x_{n-1}$ on Σ_2^n such that the following relation holds in Γ_2^n:

$$t_{\delta_1} \cdots t_{\delta_{n-1}} t_{\delta_n} t_{c_4} = D_8 \cdot \phi'_{12,m} \cdot O_{10m-11(n-2)} \cdot t_{x_1} t_{x_2} \cdots t_{x_{n-1}} \tag{11}$$

Therefore, if $g = 2$, then the element $t_{\delta_1} \cdots t_{\delta_{n-1}} t_{\delta_n} t_{c_4}$ can be written as a product of $10m - 10n + 41$ right-handed Dehn twists about nonseparating simple closed curves for any $m > 2(n-2)$.

Suppose that $g \geq 3$. Let a, b and d' be the simple closed curves on Σ_g^2 as in Figure 1. We obtain three nonseparating simple closed curves on Σ_g^n from a, b, d' by attaching S to Σ_g^2 along δ. We use the same letter a, b, d' for the three resulting curves on Σ_g^n. By the chain relation, the relation

$$t_\delta t_{\delta_n} t_{d'} = (t_{c_1} t_{c_2} \cdots t_{c_{2g+1}})^{2g+2} \cdot t_{d'}$$

holds in Γ_g^n. By the chain relation $t_a t_b = (t_{c_1} t_{c_2} \cdots t_{c_5})^6$ we may write

$$t_\delta t_{\delta_n} t_{d'} = (t_{c_1} t_{c_2} \cdots t_{c_5})^6 t_{c_4} \cdot P_{4g^2+6g-29} t_{d'} = t_a t_b t_{c_4} \cdot P_{4g^2+6g-29} \cdot t_{d'},$$

where $P_{4g^2+6g-29}$ is a product of $8(n-2)$ right-handed Dehn twists about nonseparating curves. Therefore, by relations (4) and (10), we have

$$\begin{aligned} t_\delta t_{\delta_n} t_{d'} &= D_9 \cdot \phi_{12,m} \cdot O_{10m-11(n-2)} \cdot t_d^{n-2} \cdot P_{4g^2+6g-29} \cdot t_{d'} \\ &= D_9 \cdot \phi_{12,m} \cdot O_{10m-11(n-2)} \cdot P'_{4g^2+6g-29} \cdot t_d^{n-2} t_{d'}, \end{aligned} \tag{12}$$

where $P'_{4g^2+6g-29} = t_d^{n-2} P_{4g^2+6g-29} t_d^{-n+2}$. By Lemma 10, there are nonseparating curves $x_1, \dots, x_{n-1}$ on Σ_g^n such that the following relation holds in Γ_g^n:

$$t_{\delta_1} \cdots t_{\delta_n} t_{d'} = D_9 \cdot \phi_{12,m} \cdot O_{10m-11(n-2)} \cdot P'_{4g^2+6g-29} \cdot t_{x_1} \cdots t_{x_{n-1}} \tag{13}$$

Therefore, if $g \geq 3$, then the element $t_{\delta_1} \cdots t_{\delta_{n-1}} t_{\delta_n} t_{d'}$ can be written as a product of $4g^2 + 6g + 13 + 10m - 10n$ right-handed Dehn twists about nonseparating simple closed curves for any $m > 2(n-2)$. □

3.4 Powers of boundary multitwists have infinite length

Theorem 17 *Let $g \geq 2$, and let $k \geq 2$ be a positive integer. Then, for any k and n, in the mapping class group Γ_g^n the element*

$$(t_{\delta_1} t_{\delta_2} \cdots t_{\delta_n})^k$$

can be written as a product of an arbitrarily large number of right-handed Dehn twists about nonseparating curves.

Proof of Theorem 17 Suppose that $k \in \{2, 3\}$ and $n \geq 3$.

Suppose that $g = 2$. Since $O_{10m-11(n-2)}$ contains at least two t_{c_1}, we may write the relation (11) as

$$t_{\delta_1} \cdots t_{\delta_{n-1}} t_{\delta_n} t_{c_4} = Q_{10m-10n+39} \cdot t_{c_1}^2,$$

where $Q_{10m-10n+39}$ is a product of $10m - 10n + 39$ right-handed Dehn twists about nonseparating curves. Since c_1 and c_4 are nonseparating curves and disjoint from each other, there is an element Ψ_1 in Γ_2^n such that $\Psi_1(c_4) = c_1$ and $\Psi_1(c_1) = c_4$. Therefore, by the relation $t_{\Psi_1(c)} = \Psi_1 t_c \Psi_1^{-1}$, we obtain the relation

$$t_{\delta_1} \cdots t_{\delta_{n-1}} t_{\delta_n} t_{c_1} = Q'_{10m-10n+39} \cdot t_{c_4}^2,$$

where $Q'_{10m-10n+39} = \Psi_1 Q_{10m-10n+39} \Psi_1^{-1}$. From the above relations, we have

$$\begin{aligned}(t_{\delta_1} \cdots t_{\delta_{n-1}} t_{\delta_n})^k t_{c_4}^{k-1} t_{c_1} &= (t_{\delta_1} \cdots t_{\delta_{n-1}} t_{\delta_n} t_{c_4})^{k-1} (t_{\delta_1} \cdots t_{\delta_{n-1}} t_{\delta_n} t_{c_1}) \\ &= (Q_{10m-10n+39} \cdot t_{c_1}^2)^{k-1} (Q'_{10m-10n+39} \cdot t_{c_4}^2).\end{aligned}$$

Since $k = 2, 3$, we can remove $t_{c_4}^{k-1} t_{c_1}$ from both sides of this relation. Hence, $(t_{\delta_1} \cdots t_{\delta_{n-1}} t_{\delta_n})^k$ can be written as a product of $k(10m - 10n + 39) + k$ right-handed Dehn twists about nonseparating curves. The proof for $g \geq 3$ is similar. In this case, we use the relation (13) and an element Ψ_2 in Γ_g^n such that $\Psi_2(d') = c_1$ and $\Psi_2(c_1) = d'$. The existence of Φ_2 follows from the fact that c_1 and d' are disjoint simple closed curves.

Suppose that $k \geq 4$. Since $k = 2q + 3\epsilon$ for $q \geq 1$ and $\epsilon = 0, 1$, the element $(t_{\delta_1} \cdots t_{\delta_n})^k = \{(t_{\delta_1} \cdots t_{\delta_n})^2\}^q (t_{\delta_1} \cdots t_{\delta_n})^{3\epsilon}$ in Γ_g^n $(g \geq 2)$ can be written as a product of an arbitrarily large number of right-handed Dehn twists about nonseparating curves. The case of $n = 1, 2$ follows from gluing disks along boundary components of Σ_g^n. □

Remark 18 A close look at the proof of Theorem 6 makes it evident that whenever we have arbitrarily long positive factorizations of *any* multitwist along boundary curves $t_{\delta_1}^{k_1} t_{\delta_2}^{k_2} \cdots t_{\delta_n}^{k_n}$, all but finitely many of the corresponding Lefschetz fibrations will be on symplectic 4–manifolds of general type. In particular, the total spaces of the positive factorizations of the multitwists $(t_{\delta_1} t_{\delta_2} \cdots t_{\delta_n})^k$, $k \geq 2$, in the above proof should have symplectic Kodaira dimension $\kappa = 2$, no matter what n is, which is very different than the case of $k = 1$ corresponding to Lefschetz pencils.

4 Completing the proofs of main theorems and further remarks

We now bring together various results we have obtained to complete the proofs of Theorems A, B and C. We will also discuss the length function for further mapping classes, as well as for its restrictions to subgroups of mapping classes, and list a few interesting questions.

4.1 Proofs of Theorems A, B and C

To prove our main theorems, we will simply provide navigational guides to the relevant results one needs to assemble, many of which we have obtained in the previous sections.

Proof of Theorem A It is well-known that there is a unique genus-1 Lefschetz *fibration* with (-1)–sphere sections, whose total space is $X = E(1) = \mathbb{CP}^2 \# 9\overline{\mathbb{CP}}^2$. Since $b^-(X) = 9$, there are no more than 9 disjoint (-1)–sphere sections in this fibration. It follows that $\mathrm{L}(\Delta) = -\infty$ if $n > 9$ and 12 if $1 \leq n \leq 9$.

All the remaining values of $\mathrm{L}(\Delta)$ are given by Theorem 6 and by Theorem 9. □

Proof of Theorem B The mapping class group Γ_1^1 is isomorphic to the braid group on three strands, and it is generated by t_a, t_b for any two nonseparating simple closed curves intersecting at one point. Here $H_1(\Gamma_1^1; \mathbb{Z}) \cong \mathbb{Z}$, generated by any Dehn twist along a nonseparating curve. By the 1–boundary chain relation, we have $t_\delta = (t_a t_b)^6$ in Γ_g^1. So for $[t_a] = 1$ in $H_1(\Gamma_1^1; \mathbb{Z})$, we have $[t_\delta] = 12$ in $H_1(\Gamma_1^1; \mathbb{Z})$.

If $n > 1$, we can cap off all boundary components of Σ_1^n but one, which induces a homomorphism from Γ_g^n onto Γ_g^1. Thus, any positive product of Δ^k in Γ_1^n, *if it exists*, yields a positive product of t_δ^k in Γ_1^1, which by the above calculation is equal to $12k$. It follows that Δ^k has a positive factorization of length $12k + l$ in Γ_g^n, where l is the number of Dehn twists along curves that separate some of the $n-1$ boundary components we capped off. Since L is calculated only for nonseparating curves, the latter contribution does not occur, completing the proof of our claim that Δ^k is precisely $12k$, part (1) of the theorem. It is a standard fact that any elliptic surface $E(k)$ admits $n \leq 9$ sections of self-intersection $-k$, so Δ^k admits a positive factorization provided $n \leq 9$.

Part (2) is covered by Theorem 17 and part (3) by Theorem 16. $\square$

Proof of Theorem C In both parts, the value $-\infty$ of $\widetilde{\mathrm{L}}$ or L is realized by $1 \in \Gamma_g^n$ by Proposition 5, whereas any positive k is realized by t_c^k along any homologically essential curve c by Proposition 4. Note that for $g = 0$ and $n = 1$, there are no homologically essential curves, and thus no positive factorizations to consider.

The fact that $\widetilde{\mathrm{L}}(\Gamma_g^n)$ does not contain $+\infty$ under the assumptions in part (1) follows from Proposition 1. However for $g \geq 2$, either by Theorem 16 or Theorem 17, we have mapping classes in Γ_g^n with infinite length, completing the proof of part (2) of the theorem. $\square$

4.2 Further observations and questions

As our results demonstrate, knowing that a mapping class admits a positive factorization in the mapping class group of a surface (say the page of an open book) does not in general mean that there is an upper bound on the length of all its positive factorizations. The exceptions occur in low genus cases which is essentially due to positive factorizations being lifts of quasipositive braid factorizations, where for the latter, it is known that the degree of a factorization determines the length of all possible factorizations. We can thus ask for which subgroups $N < \Gamma_g^n$, the restriction of L to N, which we denote by L_N, has bounded image.

Consider the subgroup $\mathcal{H}_g^1$ of Γ_g^1, which consists of mapping classes that commute with a fixed hyperelliptic involution on Σ_g^1. This group has a nontrivial abelianization, namely $\mathbb{Z}$, which in a similar fashion to our arguments above provides a bound on the length of any factorization into hyperelliptic Dehn twists in $\mathcal{H}_g^1$: the length of any factorization into hyperelliptic Dehn twists along nonseparating curves is fixed. The quotient of Σ_g^1 under the hyperelliptic involution gives the disk with $2g+1$ marked points. Since any $\Phi \in \mathcal{H}_g^1$ commutes with the hyperelliptic involution, it gives a mapping class of the disk with $2g+1$ marked points. Projecting the branch locus to the quotient then gives a braid in S^3, and the class $[f]$ in the abelianization is exactly the writhe of this braid under the obvious identification of $\mathrm{Ab}(\widetilde{\Gamma}_g^1)$ with $\mathbb{Z}$.

Now for $g \geq 3$, let $\Phi = t_\delta$ in $\mathcal{H}_g^1$. By the above observation, Φ has finite length in this subgroup. On the other hand, we have the 1–boundary component chain relation

$$t_\delta = (t_{c_1} t_{c_2} \cdots t_{c_{2g}})^{4g+2}.$$

It is easy to see that by applying braid relators successively, we get

$$\Phi = (t_{c_1} t_{c_2} \cdots t_{c_{2g-1}})^{4g+2} \cdot W,$$

where W is a positive word that consists of products of conjugates of $t_{c_{2g}}$. By the 2–boundary chain relation, we get $(t_{c_1} t_{c_2} \cdots t_{c_{2g}})^{2g+2} = t_{b_1} t_{b_2}$, which is a mapping class with infinite length. It follows that Φ has infinite length in Γ_g^1, even though it has finite length in the subgroup $\mathcal{H}_g^1$. We have thus seen:

Proposition 19 *The image of the positive factorization length function on the subgroup* $N = \mathrm{L}_{\mathcal{H}_g^1}$ *is strictly smaller than its image on the mapping class group* Γ_g^1*. Namely,* $\mathrm{L}_{\mathcal{H}_g^1}(\mathcal{H}_g^1) = \mathbb{N} \cup \{-\infty\}$*, whereas* $\mathrm{L}(\Gamma_g^1) = \mathbb{N} \cup \{\pm\infty\}$*.*

We therefore see that if the related geometric problem is restrained by positive factorizations in a subgroup of the mapping class group, one can achieve uniform bounds on the topology of the fillings, which are in addition asked to come from branched coverings of the 4–ball in the above case. This raises a question that is interesting in its own right:

Question 20 *For which subgroups* $N < \Gamma_g^n$ *does* L_N *have finite, positive image? What is the geometric significance of such* N*?*

References

[1] **S Akbulut**, **B Ozbagci**, *Lefschetz fibrations on compact Stein surfaces*, Geom. Topol. 5 (2001) 319–334 MR

[2] **S Bauer**, *Almost complex* 4*–manifolds with vanishing first Chern class*, J. Differential Geom. 79 (2008) 25–32 MR

[3] **R İ Baykur**, **K Hayano**, *Multisections of Lefschetz fibrations and topology of symplectic* 4*–manifolds*, Geom. Topol. 20 (2016) 2335–2395 MR

[4] **R İ Baykur**, **K Hayano**, **N Monden**, *Unchaining surgery and symplectic* 4*–manifolds* in preparation

[5] **R İ Baykur**, **M Korkmaz**, **N Monden**, *Sections of surface bundles and Lefschetz fibrations*, Trans. Amer. Math. Soc. 365 (2013) 5999–6016 MR

[6] **R İ Baykur**, **J Van Horn-Morris**, *Families of contact* 3*–manifolds with arbitrarily large Stein fillings*, J. Differential Geom. 101 (2015) 423–465 MR

[7] **R İ Baykur**, **J Van Horn-Morris**, *Topological complexity of symplectic 4–manifolds and Stein fillings*, J. Symplectic Geom. 14 (2016) 171–202 MR

[8] **E Dalyan**, **M Korkmaz**, **M Pamuk**, *Arbitrarily long factorizations in mapping class groups*, Int. Math. Res. Not. 2015 (2015) 9400–9414 MR

[9] **S K Donaldson**, *Lefschetz pencils on symplectic manifolds*, J. Differential Geom. 53 (1999) 205–236 MR

[10] **E Giroux**, *Géométrie de contact: de la dimension trois vers les dimensions supérieures*, from "Proceedings of the International Congress of Mathematicians, II" (T Li, editor), Higher Ed. Press, Beijing (2002) 405–414 MR

[11] **K Honda**, **W H Kazez**, **G Matić**, *Right-veering diffeomorphisms of compact surfaces with boundary*, Invent. Math. 169 (2007) 427–449 MR

[12] **A Kaloti**, *Stein fillings of planar open books*, preprint (2013) arXiv

[13] **T-J Li**, *The Kodaira dimension of symplectic 4–manifolds*, from "Floer homology, gauge theory, and low-dimensional topology" (D A Ellwood, P S Ozsváth, A I Stipsicz, Z Szabó, editors), Clay Math. Proc. 5, Amer. Math. Soc., Providence, RI (2006) 249–261 MR

[14] **T-J Li**, *Quaternionic bundles and Betti numbers of symplectic 4–manifolds with Kodaira dimension zero*, Int. Math. Res. Not. 2006 (2006) art. id. 37385, 28 pp. MR

[15] **T-J Li**, *Symplectic 4–manifolds with Kodaira dimension zero*, J. Differential Geom. 74 (2006) 321–352 MR

[16] **T J Li**, **A Liu**, *Symplectic structure on ruled surfaces and a generalized adjunction formula*, Math. Res. Lett. 2 (1995) 453–471 MR

[17] **A Loi**, **R Piergallini**, *Compact Stein surfaces with boundary as branched covers of B^4*, Invent. Math. 143 (2001) 325–348 MR

[18] **D Margalit**, **J McCammond**, *Geometric presentations for the pure braid group*, J. Knot Theory Ramifications 18 (2009) 1–20 MR

[19] **O Plamenevskaya**, *On Legendrian surgeries between lens spaces*, J. Symplectic Geom. 10 (2012) 165–181 MR

[20] **O Plamenevskaya**, **J Van Horn-Morris**, *Planar open books, monodromy factorizations and symplectic fillings*, Geom. Topol. 14 (2010) 2077–2101 MR

[21] **A Putman**, *An infinite presentation of the Torelli group*, Geom. Funct. Anal. 19 (2009) 591–643 MR

[22] **M-H Saitō**, **K-I Sakakibara**, *On Mordell–Weil lattices of higher genus fibrations on rational surfaces*, J. Math. Kyoto Univ. 34 (1994) 859–871 MR

[23] **Y Sato**, *Canonical classes and the geography of nonminimal Lefschetz fibrations over S^2*, Pacific J. Math. 262 (2013) 191–226 MR

[24] **H Short**, **B Wiest**, *Orderings of mapping class groups after Thurston*, Enseign. Math. 46 (2000) 279–312 MR

[25] **I Smith**, *Geometric monodromy and the hyperbolic disc*, Q. J. Math. 52 (2001) 217–228 MR

[26] **I Smith**, *Lefschetz pencils and divisors in moduli space*, Geom. Topol. 5 (2001) 579–608 MR

[27] **A I Stipsicz**, *Sections of Lefschetz fibrations and Stein fillings*, Turkish J. Math. 25 (2001) 97–101 MR

[28] **A I Stipsicz**, *Singular fibers in Lefschetz fibrations on manifolds with* $b_2^+ = 1$, Topology Appl. 117 (2002) 9–21 MR

[29] **A I Stipsicz**, *On the geography of Stein fillings of certain* 3*–manifolds*, Michigan Math. J. 51 (2003) 327–337 MR

[30] **S Tanaka**, *On sections of hyperelliptic Lefschetz fibrations*, Algebr. Geom. Topol. 12 (2012) 2259–2286 MR

[31] **C Wendl**, *Strongly fillable contact manifolds and* J*–holomorphic foliations*, Duke Math. J. 151 (2010) 337–384 MR

Department of Mathematics and Statistics, University of Massachusetts
Lederle Graduate Research Tower, 710 North Pleasant Street, Amherst, MA 01003-9305, United States

Department of Engineering Science, Osaka Electro-Communication University
Hatsu-cho 18-8, Neyagawa 572-8530, Japan

Department of Mathematical Sciences, The University of Arkansas
Fayetteville, AR 72701, United States

`baykur@math.umass.edu, monden@isc.osakac.ac.jp, jvhm@uark.edu`

Received: 24 September 2015 Revised: 13 May 2016

Algebraic & Geometric Topology 17 (2017) 1557–1674

On bordered theories for Khovanov homology

ANDREW MANION

We describe how to formulate Khovanov's functor-valued invariant of tangles in the language of bordered Heegaard Floer homology. We then give an alternate construction of Lawrence Roberts' type D and type A structures in Khovanov homology, and his algebra $\mathcal{B}\Gamma_n$, in terms of Khovanov's theory of modules over the ring H^n. We reprove invariance and pairing properties of Roberts' bordered modules in this language. Along the way, we obtain an explicit generators-and-relations description of H^n which may be of independent interest.

57M27

1 Introduction

We consider two tangle theories for Khovanov homology which are inspired by the bordered Heegaard Floer homology of Lipshitz, Ozsváth and Thurston [5]. The first theory is a reformulation of Khovanov's functor-valued invariant [4] in the bordered language. The second theory was introduced by Lawrence Roberts in [11; 12].

These bordered Khovanov theories share the same basic structure. Each assigns a differential bigraded algebra $\mathcal{B}$ to a collection of $2n$ points on the line $\{0\} \times \mathbb{R}$ in the plane $\mathbb{R} \times \mathbb{R}$. To a tangle diagram T_1 in $\mathbb{R}_{\geq 0} \times \mathbb{R}$ with $2n$ endpoints on $\{0\} \times \mathbb{R}$, these theories assign a (left) type D structure $\widehat{D}_{T_1}$ over $\mathcal{B}$. The definitions of type D structures, and other elements of the algebra of bordered Floer homology, will be given in Section 2.

To a tangle diagram T_2 in $\mathbb{R}_{\leq 0} \times \mathbb{R}$ with $2n$ endpoints on $\{0\} \times \mathbb{R}$, bordered theories assign a (right) type A structure (ie an $\mathcal{A}_\infty$–module) $\widehat{A}_{T_2}$ over $\mathcal{B}$. There is a natural pairing operation between type D and type A structures over $\mathcal{B}$ called the box tensor product, denoted $\boxtimes$ (or $\boxtimes_{\mathcal{B}}$ when $\mathcal{B}$ is unclear). If $T_2 T_1$ denotes the link diagram obtained by concatenating T_2 and T_1 horizontally, bordered theories compute the Khovanov complex $\mathrm{CKh}(T_2 T_1)$ using the following pairing formula:

$$\mathrm{CKh}(T_2 T_1) \cong \widehat{A}_{T_2} \boxtimes_{\mathcal{B}} \widehat{D}_{T_1}.$$

In Section 3, we will obtain a bordered theory with the above structure by taking $\mathcal{B}$ to be Khovanov's arc algebra H^n from [4], viewed as a differential bigraded algebra with the differential and one of the two gradings identically equal to zero. The type D and

Published: 17 July 2017 DOI: 10.2140/agt.2017.17.1557

type A structures $\widehat{D}_{T_1}$ and $\widehat{A}_{T_2}$ will be referred to as $\widehat{D}(T_1)$ and $\widehat{A}(T_2)$ in this setting. Both come from Khovanov's tangle invariants $[T_i]^{\mathrm{Kh}}$, which are chain complexes of projective graded H^n modules up to homotopy equivalence.

1.0.1 Theorem (Theorem 3.2.1) *After multiplying the intrinsic gradings on $\widehat{A}(T_2) \boxtimes_{H^n} \widehat{D}(T_1)$ by -1,*

$$\mathrm{CKh}(T_2 T_1) \cong \widehat{A}(T_2) \boxtimes_{H^n} \widehat{D}(T_1).$$

Roberts [11; 12] has a different construction of a bordered theory for Khovanov homology, including a differential bigraded algebra $\mathcal{B}\Gamma_n$ as well as type D and type A structures for tangles. The goal of Section 5 and Section 6 is to construct Roberts' theory using Khovanov's theory. The basic idea is to refine Khovanov's proofs of the existence and invariance of his tangle invariants by splitting the equations involved into subequations, each of which holds individually.

The construction of Roberts' theory from Khovanov's is not straightforward or trivial; the combinatorics is quite involved. Moreover, at various points we take our inspiration directly from [11; 12] rather than from abstract algebraic definitions. In particular, see Remark 5.3.5 below. While it would be interesting to search for the most general or natural possible explanation for the connection between these two theories, we do not pursue this goal here.

We take the first step toward relating Roberts' and Khovanov's theories in Section 4. In Section 4.1, we discuss quadratic and linear-quadratic algebras following Polishchuk and Positselski [10]. In Section 4.2, we show that H^n may be viewed as a linear-quadratic algebra.

1.0.2 Theorem (Theorem 4.2.1) *With the set of generators specified at the beginning of Section 4.2, H^n is a linear-quadratic algebra.*

This theorem allows us to write H^n as the quotient of the tensor algebra on the specified generators by an ideal generated by certain explicitly given relations, which are listed in items (1)–(4) of the proof of Theorem 4.2.1. See Corollary 4.2.7 for a more precise statement.

A combinatorial lemma about noncrossing partitions, Lemma 4.2.4, is needed to prove Theorem 1.0.2. While Theorem 1.0.2 is not necessary for the remainder of the paper, Lemma 4.2.4 is important for Section 5. Proofs of Lemma 4.2.4 were found by Dömötör Pálvölgyi [9] and independently by Aaron Potechin in a private email communication. This lemma, and Theorem 1.0.2, may be of interest to readers independently of the other constructions in this paper.

In Section 4.3, we consider a notion of Polishchuk and Positselski [10] of quadratic duality for linear-quadratic algebras. In Section 4.5, we discuss a bordered-algebra version of this duality using type DD bimodules. Generalized Koszul duality between two algebras $\mathcal{B}$ and $\mathcal{B}'$ in bordered Floer homology is defined (see Lipshitz, Ozsváth and Thurston [6]) by the existence of a quasi-invertible rank-one type DD bimodule over $\mathcal{B}$ and $\mathcal{B}'$. The algebras used in Lipshitz, Ozsváth and Thurston's construction have interesting Koszul self-duality properties. However, it seems that no such properties hold for H^n. Viewing H^n as a linear-quadratic algebra, we will see in Section 4.4 that its quadratic dual is infinite-dimensional, whereas H^n is always finitely generated over $\mathbb{Z}$.

One could ask whether the duality between H^n and this infinite-dimensional algebra is a (generalized) Koszul duality; one could also explore related theories in which everything stays finite-dimensional. We will take the second option here.

In Section 5, we will outline an alternate construction, based on H^n, of Roberts' algebra $\mathcal{B}\Gamma_n$. We define an algebra $\mathcal{B} = \mathcal{B}_R(H^n)$, and we show in Proposition 5.1.7 that the algebra $\mathcal{B}$ is linear-quadratic. The proof is very similar to the proof of Theorem 4.2.1 asserting that H^n is linear-quadratic, and it also uses Lemma 4.2.4 in an essential way. We deduce that $\mathcal{B}$ is isomorphic to the subalgebra $\mathcal{B}_R\Gamma_n$ of $\mathcal{B}\Gamma_n$ generated by right-pointing generators $\overrightarrow{e}$.

The quadratic dual $\mathcal{B}^!$ of $\mathcal{B}$ is closely related to the subalgebra $\mathcal{B}_L\Gamma_n$ of $\mathcal{B}\Gamma_n$ generated by left-pointing generators $\overleftarrow{e}$. In more detail, a mirroring operation m is defined on certain algebras in Definition 5.2.6. We will see in Proposition 5.2.8 that $\mathcal{B}_L\Gamma_n$ is a quotient of the mirror $m(\mathcal{B}^!)$ of $\mathcal{B}^!$ by certain additional relations, listed in that proposition. As Remark 5.2.3 points out, $\mathcal{B}^!$ is finitely generated for idempotent reasons.

In Section 5.3, we define a product algebra $m(\mathcal{B}^!) \odot \mathcal{B}$ of $m(\mathcal{B}^!)$ and $\mathcal{B}$. We may describe Roberts' full algebra $\mathcal{B}\Gamma_n$ as a quotient of $m(\mathcal{B}^!) \odot \mathcal{B}$.

1.0.3 Theorem (Corollary 5.3.4) *$\mathcal{B}\Gamma_n$ is isomorphic to the quotient of $m(\mathcal{B}^!) \odot \mathcal{B}$ by the extra relations on $m(\mathcal{B}^!)$ listed in Proposition 5.2.8.*

The duality properties of $\mathcal{B}\Gamma_n$ and $m(\mathcal{B}^!) \odot \mathcal{B}$ seem more promising than those of H^n. In Proposition 5.3.6 we define a rank-one type DD bimodule over $m(\mathcal{B}^!) \odot \mathcal{B}$ and its mirror version $m(m(\mathcal{B}^!) \odot \mathcal{B})$. Conjecture 5.3.9 predicts that this DD bimodule is quasi-invertible and thus yields a Koszul duality. By taking quotients of the type DD algebra outputs, we can obtain a related rank-one DD bimodule over $\mathcal{B}\Gamma_n$ and its mirror version $m(\mathcal{B}\Gamma_n)$. Thus, we could also ask if Conjecture 5.3.9 is true with $m(\mathcal{B}^!) \odot \mathcal{B}$ replaced by $\mathcal{B}\Gamma_n$. A proof of either conjecture would establish that, with regard to

Koszul duality, Roberts' bordered theory (or the version over $m(\mathcal{B}^!) \odot \mathcal{B}$) has closer formal parallels with bordered Floer homology than Khovanov's H^n theory does.

In Section 6, we show how to obtain type A and type D structures over $m(\mathcal{B}^!) \odot \mathcal{B}$ from chain complexes of graded projective H^n–modules satisfying certain algebraic conditions.

1.0.4 Theorem *The following constructions are well-defined:*

- *Let M be a chain complex of projective graded right H^n–modules satisfying the algebraic condition C_{module} of Definition 6.1.1. To M we may associate a type A structure $\widehat{A}(M)$ over $m(\mathcal{B}^!) \odot \mathcal{B}$.*
- *Let N be a chain complex of projective graded left H^n–modules satisfying the condition C_{module} of Definition 6.3.3. To N we may associate a type D structure $\widehat{D}(N)$ over $m(\mathcal{B}^!) \odot \mathcal{B}$.*

Theorem 1.0.4 is a summary of Definition 6.2.4, Proposition 6.2.5 and Definition 6.3.7 (as well as the definitions and propositions preceding them).

The chain complexes $[T_i]^{\mathrm{Kh}}$ associated to tangles by Khovanov satisfy C_{module}, so Theorem 1.0.4 gives us type A and type D structures $\widehat{A}([T_2]^{\mathrm{Kh}})$ and $\widehat{D}([T_1]^{\mathrm{Kh}})$ over $m(\mathcal{B}^!) \odot \mathcal{B}$. By Proposition 6.2.6, the extra relations of Theorem 1.0.3 act as zero on the type A structure $\widehat{A}([T_2]^{\mathrm{Kh}})$, so we get a type A structure over the quotient algebra $\mathcal{B}\Gamma_n$. We may also take quotients of the algebra outputs of the type D structure $\widehat{D}([T_1]^{\mathrm{Kh}})$ to get a type D structure over $\mathcal{B}\Gamma_n$.

1.0.5 Theorem (Proposition 6.2.7 and Proposition 6.3.10) *The type A structure $\widehat{A}([T_2]^{\mathrm{Kh}})$ over $\mathcal{B}\Gamma_n$, and the type D structure $\widehat{D}([T_1]^{\mathrm{Kh}})$ over $\mathcal{B}\Gamma_n$, are isomorphic to the type A and D structures Roberts associates to T_2 and T_1 in [11; 12].*

We show that the pairing of the bordered modules over $m(\mathcal{B}^!) \odot \mathcal{B}$ agrees with the tensor product of the original chain complexes over H^n.

1.0.6 Theorem (Proposition 6.4.1) *Given M and N as in Theorem 1.0.4, we have*

$$\widehat{A}(M) \boxtimes_{m(\mathcal{B}^!)\odot\mathcal{B}} \widehat{D}(N) \cong M \otimes_{H^n} N,$$

after multiplying the intrinsic gradings on $M \otimes_{H^n} N$ by -1.

By Proposition 6.4.3, the pairing $\boxtimes$ is the same over $m(\mathcal{B}^!) \odot \mathcal{B}$ and its quotient $\mathcal{B}\Gamma_n$. Thus, we get an alternate proof that the pairing of Roberts' type D and type A structures computes Khovanov homology.

Finally, in Section 6.5 and Section 6.6 we show that the homotopy types of $\widehat{A}([T_2]^{\mathrm{Kh}})$ and $\widehat{D}([T_1]^{\mathrm{Kh}})$, as type A and type D structures over $m(\mathcal{B}^!) \odot \mathcal{B}$, are invariants of the tangles underlying the diagrams T_1 and T_2.

1.0.7 Theorem (Corollary 6.5.21 and Corollary 6.6.9) *Performing a Reidemeister move on T_2 or T_1 yields a homotopy equivalence between the corresponding type A structures $\widehat{A}([T_2]^{\mathrm{Kh}})$ or type D structures $\widehat{D}([T_1]^{\mathrm{Kh}})$ over $m(\mathcal{B}^!) \odot \mathcal{B}$.*

With the help of Proposition 6.5.22, we also obtain an alternate proof that Roberts' type A and type D structures over $\mathcal{B}\Gamma_n$ are homotopy-invariant under Reidemeister moves.

Acknowledgments I would like to thank Zoltán Szabó, Lawrence Roberts, Victor Reiner and Robert Lipshitz for interesting comments and discussions during the writing of this paper. I would especially like to thank Dömötör Pálvölgyi and Aaron Potechin for independently finding proofs of Lemma 4.2.4. Finally, I would like to thank the referee for useful suggestions.

This work was supported by the National Science Foundation GRFP under Award No DGE 1148900 and the National Science Foundation MSPRF under Award No 1502686.

2 Some bordered algebra

The standard reference for the algebra of bordered Floer homology is Lipshitz, Ozsváth and Thurston [7]. We will use only a subset of the full algebraic machinery; however, we will work with coefficients in $\mathbb{Z}$ rather than $\mathbb{Z}/2\mathbb{Z}$. For this sign lift, we will follow the conventions of Roberts in [11; 12].

2.1 Differential graded algebras and modules

2.1.1 Convention Unless otherwise specified, all algebras and modules discussed in this paper will be assumed to be finitely generated over $\mathbb{Z}$.

The following is the notion of differential graded algebra which will be most useful for us; we will not need to use more general $\mathcal{A}_\infty$–algebras. In this paper, the coefficient ring R is always a direct product of finitely many copies of $\mathbb{Z}$.

2.1.2 Definition A *differential bigraded algebra*, or *dg algebra*, is a bigraded unital associative algebra $\mathcal{B}$ over a coefficient ring R, equipped with an R–bilinear differential

μ_1 which is homogeneous of degree $(0,+1)$ with respect to the bigrading. The two gradings on a dg algebra will be called the intrinsic and homological gradings (in that order). Thus, the differential should preserve the intrinsic grading and increase the homological grading by 1.

The differential must satisfy the following Leibniz rule:

$$\mu_1(xy) = (-1)^{\deg_h y}(\mu_1(x))y + x(\mu_1(y)),$$

where $\deg_h$ denotes the homological degree, for elements x and y of $\mathcal{B}$ which are homogeneous with respect to the homological grading. The coefficient ring R is required to coincide with the summand $\mathcal{B}_{0,0}$ of $\mathcal{B}$ in bigrading $(0,0)$.

2.1.3 Definition Suppose $R = \mathbb{Z}^{\times k}$. The elements $e_1 = (1,0,\dots,0),\dots,e_k = (0,\dots,0,1)$ will be called the *minimal*, or *elementary*, idempotents of $\mathcal{B}$. The coefficient ring R will also be referred to as the *idempotent ring* of $\mathcal{B}$. For each elementary idempotent e_i, there is a left R–module $Re_i \simeq \mathbb{Z}$.

2.1.4 Remark The usual convention in bordered Floer homology is to have the differential decrease the homological grading by 1; we have chosen to reverse this convention since the differentials in Khovanov homology increase homological grading by 1.

2.1.5 Remark Bordered Floer homology requires more general gradings by a (possibly nonabelian) group G and a distinguished element λ in the center of G. We use here only the special case where G is the abelian group $\mathbb{Z}^2$ and λ is $(0,1)$.

When dealing with bigraded algebras or modules, we will use the following degree shift convention: if $X = \bigoplus_{i,j} X_{i,j}$ is any type of bigraded object, then $X[m,n]$ is the same type of bigraded object, and the summand of $X[m,n]$ in bigrading (i,j) is $X_{i-m,j-n}$.

Since we are working over $\mathbb{Z}$, the following notation will also be useful, following Roberts [11; 12]. If X is any type of bigraded object, then $|\mathrm{id}|\colon X \to X$ is defined by multiplication by $(-1)^{\deg_h}$, where $\deg_h$ denotes the homological degree. Similarly, $|\mathrm{id}|^j\colon X \to X$ is defined by multiplication by $(-1)^{j \deg_h}$, and $|\mathrm{id}|^{j\otimes k}$ is the k–fold tensor product of $|\mathrm{id}|^j$. In this notation, if μ_2 denotes the multiplication on a dg algebra $\mathcal{B}$, then the Leibniz rule for the differential μ_1 on $\mathcal{B}$ can be written as

$$\mu_1 \circ \mu_2 = \mu_2 \circ (\mu_1 \otimes |\mathrm{id}|) + \mu_2 \circ (\mathrm{id} \otimes \mu_1).$$

2.1.6 Definition A *left differential bigraded module*, or *left dg module*, over a dg algebra $\mathcal{B}$, is a bigraded left $\mathcal{B}$–module M equipped with a differential d of bidegree

$(0,+1)$, such that the Leibniz rule

$$d \circ m = m \circ (\mu_1 \otimes |\mathrm{id}|) + m \circ (\mathrm{id} \otimes d)$$

is satisfied, where $m\colon \mathcal{B} \otimes_R M \to M$ is the action of $\mathcal{B}$ on M and μ_1 is the differential on $\mathcal{B}$.

2.1.7 Definition A *right differential bigraded module*, or *right dg module*, over a dg algebra $\mathcal{B}$, is a bigraded right $\mathcal{B}$–module M equipped with a differential d of bidegree $(0,+1)$, such that the Leibniz rule

$$d \circ m = m \circ (d \otimes |\mathrm{id}|) + m \circ (\mathrm{id} \otimes \mu_1)$$

is satisfied, where $m\colon M \otimes_R \mathcal{B} \to M$ is the action of $\mathcal{B}$ on M and μ_1 is the differential on $\mathcal{B}$.

If M is a right dg module and M' is a left dg module over $\mathcal{B}$, then we can take the tensor product of M and M' over $\mathcal{B}$ to produce a chain complex of graded abelian groups, or equivalently a differential bigraded $\mathbb{Z}$–module.

2.1.8 Definition Let M be a right dg module and M' be a left dg module over $\mathcal{B}$. The differential on the tensor product $M \otimes_{\mathcal{B}} M'$ is defined to be

$$d_{M \otimes_{\mathcal{B}} M'} := d_M \otimes |\mathrm{id}_{M'}| + \mathrm{id}_M \otimes d_{M'}.$$

2.2 Type D structures

2.2.1 Definition Let $\mathcal{B}$ be a differential bigraded algebra over R as in Definition 2.1.2. Let μ_1 and μ_2 denote the differential and multiplication on $\mathcal{B}$, respectively.

A *type D structure* over $\mathcal{B}$ is, firstly, a bigraded left R–module $\widehat{D}$ which is isomorphic to a finite direct sum of R–modules $Re_{i_\alpha}[j_\alpha,k_\alpha]$, where the e_{i_α} are elementary idempotents of $\mathcal{B}$ (all in bigrading $(0,0)$) and $[j_\alpha,k_\alpha]$ is a grading shift. The module $\widehat{D}$ should be equipped with a bigrading-preserving R–linear map

$$\delta\colon \widehat{D} \to (\mathcal{B} \otimes_R \widehat{D})[0,-1],$$

such that

$$(\mu_1 \otimes |\mathrm{id}|) \circ \delta + (\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta) \circ \delta = 0.$$

2.2.2 Remark The condition that $\widehat{D} = \bigoplus_\alpha Re_{i_\alpha}[j_\alpha,k_\alpha]$ would be unnecessary if R were a direct product of copies of $\mathbb{Z}/2\mathbb{Z}$, rather than $\mathbb{Z}$. But over $\mathbb{Z}$, we want to exclude cases like $\mathcal{B} = R = \mathbb{Z}$, $\widehat{D} = \mathbb{Z}/2\mathbb{Z}$, $\delta = 0$ from being valid type D structures. The reason for this restriction is that we want Proposition 2.2.3 below, which is true over $\mathbb{Z}/2\mathbb{Z}$, to hold over $\mathbb{Z}$ as well.

2.2.3 Proposition *If $(\widehat{D},\delta)$ is a type D structure over $\mathcal{B}$, then $\mathcal{B}\otimes_R\widehat{D}$ is a projective left dg $\mathcal{B}$–module when equipped with the differential*

$$d := \mu_1\otimes|\mathrm{id}| + (\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta),$$

where μ_1 and μ_2 denote the differential and multiplication on $\mathcal{B}$, respectively.

Proof First, since $\widehat{D}$ (as an R–module) is a direct sum of R–modules $Re_{i_\alpha}[j_\alpha,k_\alpha]$, $\mathcal{B}\otimes_R\widehat{D}$ is a direct sum of $\mathcal{B}$–modules $\mathcal{B}e_{i_\alpha}[j_\alpha,k_\alpha]$. These are each projective because they are summands of grading shifts of $\mathcal{B}$: if $R=\mathbb{Z}^{\times k}$, we have $\mathcal{B}=\bigoplus_{i=1}^k\mathcal{B}e_i$ as left $\mathcal{B}$–modules. Thus, $\mathcal{B}\otimes_R\widehat{D}$ is a projective $\mathcal{B}$–module.

Before showing that $d^2=0$, we check that d satisfies the Leibniz rule. The action of the algebra $\mathcal{B}$ on $\mathcal{B}\otimes_R\widehat{D}$ is given by the following map:

$$m := \mu_2\otimes\mathrm{id}\colon \mathcal{B}\otimes_R(\mathcal{B}\otimes_R\widehat{D}) = (\mathcal{B}\otimes_R\mathcal{B})\otimes_R\widehat{D}\to\mathcal{B}\otimes_R\widehat{D}.$$

We want to show that $d\circ m = m\circ(\mu_1\otimes|\mathrm{id}|)+m\circ(\mathrm{id}\otimes d)$, as maps from $\mathcal{B}\otimes_R(\mathcal{B}\otimes_R\widehat{D})$ to $(\mathcal{B}\otimes_R\widehat{D})$. We can write out the left side:

$$\begin{aligned}
d\circ m &= (\mu_1\otimes|\mathrm{id}| + (\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta))\circ(\mu_2\otimes\mathrm{id})\\
&= (\mu_1\circ\mu_2)\otimes|\mathrm{id}| + (\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta)\circ(\mu_2\otimes\mathrm{id})\\
&= (\mu_2\circ(\mu_1\otimes|\mathrm{id}|))\otimes|\mathrm{id}| + (\mu_2\circ(\mathrm{id}\otimes\mu_1))\otimes|\mathrm{id}|\\
&\qquad\qquad + (\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta)\circ(\mu_2\otimes\mathrm{id})\\
&= (\mu_2\otimes\mathrm{id})\circ(\mu_1\otimes|\mathrm{id}|\otimes|\mathrm{id}|) + (\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\mu_1\otimes|\mathrm{id}|)\\
&\qquad\qquad + (\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta)\circ(\mu_2\otimes\mathrm{id}).
\end{aligned}$$

Meanwhile, the right side is this:

$$\begin{aligned}
&m\circ(\mu_1\otimes|\mathrm{id}|)+m\circ(\mathrm{id}\otimes d)\\
&\quad= (\mu_2\otimes\mathrm{id})\circ(\mu_1\otimes|\mathrm{id}|\otimes|\mathrm{id}|)+(\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes d)\\
&\quad= (\mu_2\otimes\mathrm{id})\circ(\mu_1\otimes|\mathrm{id}|\otimes|\mathrm{id}|)+(\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes(\mu_1\otimes|\mathrm{id}|+(\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta)))\\
&\quad= (\mu_2\otimes\mathrm{id})\circ(\mu_1\otimes|\mathrm{id}|\otimes|\mathrm{id}|)+(\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\mu_1\otimes|\mathrm{id}|)\\
&\qquad\qquad +(\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes((\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta))).
\end{aligned}$$

The first two terms on the left side cancel with those on the right side, and we only need show that

$$(\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta)\circ(\mu_2\otimes\mathrm{id}) = (\mu_2\otimes\mathrm{id})\circ\big(\mathrm{id}\otimes((\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta))\big).$$

This identity follows since

$$\begin{aligned}(\mu_2\otimes\mathrm{id})\circ\big(\mathrm{id}\otimes((\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta))\big) &= (\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\mathrm{id}\otimes\delta)\\ &= (\mu_2\otimes\mathrm{id})\circ(\mu_2\otimes\mathrm{id}\otimes\mathrm{id})\circ(\mathrm{id}\otimes\mathrm{id}\otimes\delta)\\ &= (\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta)\circ(\mu_2\otimes\mathrm{id}).\end{aligned}$$

Now suppose $a\otimes x$ is a generator of $\mathcal{B}\otimes_R\widehat{D}$; we want to show $d^2(a\otimes x)=0$. We may write $a\otimes x$ as $m(a,1\otimes x)$ and apply the Leibniz rule

$$d(a\otimes x)=(-1)^{\deg_h x}m(\mu_1(a),1\otimes x)+m(a,\delta(x)),$$

so

$$\begin{aligned}d^2(a\otimes x) &= (-1)^{\deg_h x}d(m(\mu_1(a),1\otimes x))+d(m(a,\delta(x)))\\ &= (-1)^{\deg_h x}m(\mu_1(a),\delta(x))+(-1)^{\deg_h x+1}m(\mu_1(a),\delta(x))+m(a,d(\delta(x))).\end{aligned}$$

The first two terms cancel each other, so it suffices to show that $d(\delta(x))=0$. Writing out d, this equation amounts to

$$(\mu_1\otimes|\mathrm{id}|)\circ\delta+(\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta)\circ\delta=0.$$

This is exactly the type D structure relation. □

The following propositions will be useful in the description of Khovanov's functor-valued invariant as a bordered theory.

2.2.4 Proposition *Let $\mathcal{B}$ be a dg algebra over R. Suppose that $\mathcal{B}$ is concentrated in homological degree 0 (it may have nontrivial intrinsic gradings). Then a dg module over $\mathcal{B}$ is the same as a chain complex of singly graded $\mathcal{B}$–modules with $\mathcal{B}$–linear grading-preserving differential maps.*

Proof Since $\mathcal{B}$ is concentrated in homological degree 0, the differential on $\mathcal{B}$ must be zero. Let M be a dg module over $\mathcal{B}$, with summand $M_{j,k}$ in bigrading (j,k). Then, for each homological grading k, the summand $\bigoplus_j M_{j,k}$ of M is preserved when multiplying by $\mathcal{B}$; it is a singly graded $\mathcal{B}$–module. Define a chain complex with chain module $C_k=\bigoplus_j M_{j,k}$. The differential $C_k\to C_{k+1}$ is the differential on M; it is $\mathcal{B}$–linear by the Leibniz rule, since $\mathcal{B}$ has no differential.

In the other direction, taking direct sums over chain modules yields a map from chain complexes to dg modules. These operations are inverse to each other. □

The next proposition involves isomorphisms of type D structures; see Definition 6.6.1 for the basic definitions.

2.2.5 Proposition *Let $\mathcal{B}$ be a dg algebra over R. Suppose that $\mathcal{B}$ is concentrated in homological degree 0, and that all intrinsic gradings of $\mathcal{B}$ are nonnegative. Then a type D structure over $\mathcal{B}$ is the same, up to isomorphism, as a chain complex of singly graded projective left $\mathcal{B}$–modules with $\mathcal{B}$–linear grading-preserving differential maps.*

Proof Given a type D structure $\widehat{D}$ over $\mathcal{B}$, Proposition 2.2.3 shows that $\mathcal{B} \otimes_R \widehat{D}$ is a dg module over $\mathcal{B}$, or equivalently a chain complex of graded left $\mathcal{B}$–modules by Proposition 2.2.4. In fact, each term of the chain complex is projective, since it is a direct sum of modules $\mathcal{B}e_{i_\alpha}[j_\alpha, k_\alpha]$.

Conversely, suppose $\cdots \to C_k \to C_{k+1} \to \cdots$ is a chain complex of graded projective left $\mathcal{B}$–modules. Since each C_k is assumed to be finitely generated, it may be written as a direct sum of indecomposable graded projective left R–modules $C_{k,\alpha}$. By Khovanov [4, Lemma 1 of Section 2.5], which assumes that the intrinsic gradings of $\mathcal{B}$ are nonnegative, we see that each $C_{k,\alpha}$ is isomorphic to $\mathcal{B}e_{i_{k,\alpha}}[j_{k,\alpha}]$ for some uniquely determined elementary idempotent $e_{i_{k,\alpha}}$ and grading shift $j_{k,\alpha}$. Define $\widehat{D}$ as a bigraded R–module to be the direct sum, over all k and α, of $Re_{i_{k,\alpha}}[j_{k,\alpha}, k]$.

We may identify $\bigoplus_k C_k$ with $\mathcal{B} \otimes_R \widehat{D}$, since $\bigoplus_k C_k = \bigoplus_{k,\alpha} \mathcal{B}e_{i_{k,\alpha}}[j_{k,\alpha}, k]$ and $\widehat{D} = \bigoplus_{k,\alpha} Re_{i_{k,\alpha}}[j_{k,\alpha}, k]$. Let d denote the differential on the dg module $\bigoplus_k C_k$. Then the type D operation $\delta\colon \widehat{D} \to \mathcal{B} \otimes_R \widehat{D}$ is obtained by restricting d to $\widehat{D} \cong 1 \otimes_R \widehat{D} \subset \mathcal{B} \otimes_R \widehat{D}$. It has the correct grading properties because d does.

Since d satisfies the Leibniz rule, we may write $d = \mu_1 \otimes |\mathrm{id}| + (\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta)$. Thus, the type D relations for δ are equivalent to $d \circ \delta = 0$, which holds because δ is a restriction of d.

Finally, we show the two constructions given above are inverses up to isomorphism. Suppose we start with a chain complex $\cdots \to C_k \to C_{k+1} \to \cdots$, decompose each C_k as $\bigoplus_\alpha C_{k,\alpha}$ and take the corresponding type D structure $\widehat{D}$. Then $\mathcal{B} \otimes_R \widehat{D}$ is clearly isomorphic to the dg module associated to $\cdots \to C_k \to C_{k+1} \to \cdots$. On the other hand, suppose we start with a type D structure $\widehat{D}$ and then obtain a type D structure by decomposing $\mathcal{B} \otimes_R \widehat{D}$ into indecomposable projectives. The resulting type D structure has the same number of generators as $\widehat{D}$, with the same idempotents and bigradings. However, the type D operation may be different: in decomposing $\mathcal{B} \otimes_R \widehat{D}$, we may have incorporated a change of basis.

From Definition 6.6.1, we see that if $\widehat{D}$ and $\widehat{D}'$ are two type D structures such that $\mathcal{B} \otimes_R \widehat{D}$ and $\mathcal{B} \otimes_R \widehat{D}'$ are isomorphic as dg modules, then $\widehat{D}$ and $\widehat{D}'$ are isomorphic as type D structures. The required isomorphisms of type D structures may be obtained by restricting the isomorphisms of dg modules. Thus, any type D structure obtained by decomposing $\mathcal{B} \otimes_R \widehat{D}$ as above is isomorphic to $\widehat{D}$. $\square$

2.3 Type A structures and pairing

2.3.1 Definition Let $\mathcal{B}$ be a dg algebra over R as in Definition 2.1.2. Let μ_1 and μ_2 denote the differential and multiplication on $\mathcal{B}$, respectively.

A *type A structure* $\widehat{A}$ over $\mathcal{B}$, synonymous with $\mathcal{A}_\infty$*–module* over $\mathcal{B}$, is a bigraded right R–module $\widehat{A}$, finitely generated over R as usual by Convention 2.1.1, together with R–linear bigrading-preserving maps $m_i\colon \widehat{A}\otimes_R \mathcal{B}^{\otimes(i-1)} \to \widehat{A}[0,i-2]$, $i\in\mathbb{Z}_{\geq 1}$, satisfying

$$\sum_{i+j=n+1}(-1)^{j(i+1)}m_i\circ(m_j\otimes|\mathrm{id}|^{j\otimes(i-1)}) + (-1)^{n+1}\sum_{k=1}^{n-1} m_n\circ(\mathrm{id}^{\otimes k}\otimes\mu_1\otimes|\mathrm{id}|^{\otimes(n-k-1)}) + \sum_{k=1}^{n-2}(-1)^k m_{n-1}\circ(\mathrm{id}^{\otimes k}\otimes\mu_2\otimes\mathrm{id}^{\otimes(n-k-2)}) = 0$$

for every $n\geq 1$. The type A structure $\widehat{A}$ is called *strictly unital* if $m_2(-,1)=\mathrm{id}_{\widehat{A}}$ and $m_n=0$ for $n>2$ when any of the algebra inputs to m_n is 1.

2.3.2 Example If M is a (right) dg $\mathcal{B}$–module, then M is a strictly unital type A structure over $\mathcal{B}$ with $m_i=0$ for $i\neq 1,2$. If M is an ordinary bigraded module over $\mathcal{B}$, with no differential, then M is a strictly unital type A structure with $m_i=0$ for $i\neq 2$.

2.3.3 Remark We will only need to work with type A structures which come from dg modules as in Example 2.3.2. Thus, all our type A structures will be strictly unital, so we will omit mention of this condition in what follows. However, although our type A structures will have no nontrivial higher action terms, we will eventually need to work with $\mathcal{A}_\infty$*–morphisms* between these type A structures. We will need to consider morphisms which do have nontrivial higher $\mathcal{A}_\infty$–terms; see Section 6.5.

Given a type D structure $(\widehat{D},\delta)$ and a type A structure $(\widehat{A},\{m_i\mid i\geq 1\})$ over $\mathcal{B}$, with either $\widehat{D}$ or $\widehat{A}$ operationally bounded in an appropriate sense, the natural way to pair them is known as the box tensor product. It yields a differential bigraded abelian group $\widehat{A}\boxtimes\widehat{D}$. We will not worry about boundedness in this paper since all type D and type A structures under consideration are bounded. See Lipshitz, Ozsváth and Thurston [7] for more details and algebraic properties of $\boxtimes$ over $\mathbb{Z}/2\mathbb{Z}$. The material below follows Roberts [11]; we include proofs for completeness.

To define $\boxtimes$, the following notation will be useful.

2.3.4 Definition Let $(\widehat{D}, \delta)$ be a type D structure over $\mathcal{B}$. The map $\delta^k \colon \widehat{D} \to \mathcal{B}^k \otimes_R \widehat{D}$ is

$$\delta^k := (\overbrace{\mathrm{id} \otimes \cdots \otimes \mathrm{id}}^{k-1} \otimes \delta) \circ \cdots \circ (\mathrm{id} \otimes \delta) \circ \delta,$$

where δ is applied k times. In particular, $\delta = \delta^1$.

2.3.5 Definition [11, Definition 79] $\widehat{A} \boxtimes \widehat{D}$, as a bigraded abelian group, is the tensor product $\widehat{A} \otimes_R \widehat{D}$. The differential on $\widehat{A} \boxtimes \widehat{D}$ is

$$\partial^{\boxtimes} := \sum_{n=1}^{\infty} (m_n \otimes |\mathrm{id}|^n) \circ (\mathrm{id} \otimes \delta^{n-1}).$$

Since we are implicitly assuming boundedness, only finitely many terms of the sum are nonzero.

2.3.6 Proposition [11, Theorem 80] *The operator $\partial^{\boxtimes}$, as defined in Definition 2.3.5, satisfies*

$$(\partial^{\boxtimes})^2 = 0.$$

Proof In this proof, when referring to identity operators, we will use subscripts to explicitly indicate which identity operators we mean.

First, note that as maps from $\widehat{A} \otimes_R \mathcal{B}^{j-1} \otimes_R \widehat{D}$ to $\widehat{A} \otimes_R \mathcal{B}^{i-1} \otimes_R \widehat{D}$, we have

$$(\mathrm{id}_{\widehat{A}} \otimes \delta^{i-1}) \circ (m_j \otimes |\mathrm{id}_{\widehat{D}}|^j) = (-1)^{j(i+1)} (m_j \otimes |\mathrm{id}_{\mathcal{B}^{i-1} \otimes \widehat{D}}|^j) \circ (\mathrm{id}_{\widehat{A} \otimes \mathcal{B}^{j-1}} \otimes \delta^{i-1}).$$

This identity is immediate over $\mathbb{Z}/2\mathbb{Z}$, and we need only verify that the signs are right. On the right side of the equality, we have $|\mathrm{id}_{\mathcal{B}^{i-1} \otimes \widehat{D}}|^j$ which is computed from the homological degree of an output of δ^{i-1}. Since δ increases homological degree by 1, δ^{i-1} increases homological degree by $i-1$. Thus, compared with the left side, the right side has an extra factor of $(-1)^{j(i-1)} = (-1)^{j(i+1)}$.

Thus,

$$\begin{aligned}
(\partial^{\boxtimes})^2 &= \sum_{n \geq 1} \sum_{i+j=n+1} (m_i \otimes |\mathrm{id}_{\widehat{D}}|^i) \circ (\mathrm{id}_{\widehat{A}} \otimes \delta^{i-1}) \circ (m_j \otimes |\mathrm{id}_{\widehat{D}}|^j) \circ (\mathrm{id}_{\widehat{A}} \otimes \delta^{j-1}) \\
&= \sum_{n \geq 1} \sum_{i+j=n+1} (-1)^{j(i+1)} (m_i \otimes |\mathrm{id}_{\widehat{D}}|^i) \circ (m_j \otimes |\mathrm{id}_{\mathcal{B}^{i-1}}|^j \otimes |\mathrm{id}_{\widehat{D}}|^j) \\
&\qquad \circ (\mathrm{id}_{\widehat{A}} \otimes \mathrm{id}_{\mathcal{B}^{j-1}} \otimes \delta^{i-1}) \circ (\mathrm{id}_{\widehat{A}} \otimes \delta^{j-1}) \\
&= \sum_{n \geq 1} \sum_{i+j=n+1} (-1)^{j(i+1)} ((m_i \circ (m_j \otimes |\mathrm{id}_{\mathcal{B}^{i-1}}|^j)) \otimes |\mathrm{id}_{\widehat{D}}|^{n+1}) \circ (\mathrm{id}_{\widehat{A}} \otimes \delta^{n-1})
\end{aligned}$$

$$= \sum_{n\geq 1}\Bigg(\Bigg((-1)^n \sum_{k=1}^{n-1} m_n \circ (\mathrm{id}_{\hat{A}} \otimes \mathrm{id}_{\mathcal{B}^{k-1}} \otimes \mu_1 \otimes |\mathrm{id}_{\mathcal{B}^{n-k-1}}|) \\ - \sum_{k=1}^{n-2} (-1)^k m_{n-1} \circ (\mathrm{id}_{\hat{A}} \otimes \mathrm{id}_{\mathcal{B}^{k-1}} \otimes \mu_2 \otimes \mathrm{id}_{\mathcal{B}^{n-k-2}})\Bigg) \otimes |\mathrm{id}_{\hat{D}}|^{n+1}\Bigg) \\ \circ (\mathrm{id}_{\hat{A}} \otimes \delta^{n-1}),$$

where the type A relations for $\hat{A}$ were used in the final equality. It remains to show that the derivative terms

$$\sum_{n\geq 1}\Bigg(\Bigg((-1)^n \sum_{k=1}^{n-1} m_n \circ (\mathrm{id}_{\hat{A}} \otimes \mathrm{id}_{\mathcal{B}^{k-1}} \otimes \mu_1 \otimes |\mathrm{id}_{\mathcal{B}^{n-k-1}}|)\Bigg) \otimes |\mathrm{id}_{\hat{D}}|^{n+1}\Bigg) \circ (\mathrm{id}_{\hat{A}} \otimes \delta^{n-1})$$

are equal to the multiplication terms

$$\sum_{n\geq 1}\Bigg(\Bigg(\sum_{k=1}^{n-2} (-1)^k m_{n-1} \circ (\mathrm{id}_{\hat{A}} \otimes \mathrm{id}_{\mathcal{B}^{k-1}} \otimes \mu_2 \otimes \mathrm{id}_{\mathcal{B}^{n-k-2}})\Bigg) \otimes |\mathrm{id}_{\hat{D}}|^{n+1}\Bigg) \circ (\mathrm{id}_{\hat{A}} \otimes \delta^{n-1}).$$

For a fixed $n \geq 1$ and $1 \leq k \leq n-1$, we claim that the derivative term is equal to

$$(-1)^{k+1} (m_n \otimes |\mathrm{id}_{\hat{D}}|^n) \circ \big(\mathrm{id}_{\hat{A}} \otimes \big((\mathrm{id}_{\mathcal{B}^{n-2}} \otimes \delta) \circ \cdots \\ \circ (\mathrm{id}_{\mathcal{B}^{k-1}} \otimes \mu_1 \otimes |\mathrm{id}_{\hat{D}}|) \circ (\mathrm{id}_{\mathcal{B}^{k-1}} \otimes \delta) \circ \cdots \circ (\mathrm{id}_{\mathcal{B}} \otimes \delta) \circ \delta\big)\big).$$

Over $\mathbb{Z}/2\mathbb{Z}$, this equality follows from expanding out δ^{n-1}. To see that the formula holds over $\mathbb{Z}$, note that when $k = n-1$, the sign in front of the above expression is $(-1)^n$, in agreement with the original expression for the derivative term. Each time k is decreased by 1, the sign should flip because $|\mathrm{id}_{\hat{D}}|$ occurs after one fewer instance of δ in the second expression, compared to the original.

Now,

$$(\mathrm{id}_{\mathcal{B}^{k-1}} \otimes \mu_1 \otimes |\mathrm{id}_{\hat{D}}|) \circ (\mathrm{id}_{\mathcal{B}^{k-1}} \otimes \delta) = \mathrm{id}_{\mathcal{B}^{k-1}} \otimes ((\mu_1 \otimes |\mathrm{id}_{\hat{D}}|) \circ \delta) \\ = -\mathrm{id}_{\mathcal{B}^{k-1}} \otimes ((\mu_2 \otimes \mathrm{id}_{\hat{D}}) \circ (\mathrm{id}_{\mathcal{B}} \otimes \delta) \circ \delta)$$

by the type D relations for $\hat{D}$. Thus, the sum of the derivative terms is

$$\sum_{n\geq 1}\sum_{k=1}^{n-1} (-1)^k (m_n \otimes |\mathrm{id}_{\hat{D}}|^n) \circ \big(\mathrm{id}_{\hat{A}} \otimes \big((\mathrm{id}_{\mathcal{B}^{n-2}} \otimes \delta) \circ \cdots \\ \circ (\mathrm{id}_{\mathcal{B}^k} \otimes \delta) \circ (\mathrm{id}_{\mathcal{B}^{k-1}} \otimes \mu_2 \otimes \mathrm{id}_{\hat{D}}) \circ (\mathrm{id}_{\mathcal{B}^k} \otimes \delta) \circ (\mathrm{id}_{\mathcal{B}^{k-1}} \otimes \delta) \circ \cdots \circ \delta\big)\big) \\ = \sum_{n\geq 1}\sum_{k=1}^{n-1} (-1)^k (m_n \otimes |\mathrm{id}_{\hat{D}}|^n) \circ (\mathrm{id}_{\hat{A} \otimes \mathcal{B}^{k-1}} \otimes \mu_2 \otimes \mathrm{id}_{\mathcal{B}^{n-k-1} \otimes \hat{D}}) \circ (\mathrm{id}_{\hat{A}} \otimes \delta^n).$$

Since the $n=1$ multiplication term is zero, and $|\mathrm{id}_{\widehat{D}}|^{n+1} = |\mathrm{id}_{\widehat{D}}|^{n-1}$, the sum of the multiplication terms is

$$\sum_{n\geq 2}\sum_{k=1}^{n-2}(-1)^k(m_{n-1}\otimes|\mathrm{id}_{\widehat{D}}|^{n-1})\circ(\mathrm{id}_{\widehat{A}\otimes\mathcal{B}^{k-1}}\otimes\mu_2\otimes\mathrm{id}_{\mathcal{B}^{n-k-2}\otimes\widehat{D}})\circ(\mathrm{id}_{\widehat{A}}\otimes\delta^{n-1})$$
$$=\sum_{n\geq 1}\sum_{k=1}^{n-1}(-1)^k(m_{n}\otimes|\mathrm{id}_{\widehat{D}}|^{n})\circ(\mathrm{id}_{\widehat{A}\otimes\mathcal{B}^{k-1}}\otimes\mu_2\otimes\mathrm{id}_{\mathcal{B}^{n-k-1}\otimes\widehat{D}})\circ(\mathrm{id}_{\widehat{A}}\otimes\delta^{n}).$$

The sums of the derivative and multiplication terms agree, proving that $(\partial^{\boxtimes})^2 = 0$. $\square$

2.3.7 Proposition [5, Example 2.2.7] *Let $\mathcal{B}$ be a dg algebra over R as in Definition 2.1.2. Let $\widehat{D}$ be a type D structure over $\mathcal{B}$, and let $\widehat{A}$ be a right dg module over $\mathcal{B}$. Then $\widehat{A}\boxtimes\widehat{D}$ and $\widehat{A}\otimes_{\mathcal{B}}(\mathcal{B}\otimes_R\widehat{D})$ are isomorphic as differential bigraded abelian groups.*

For completeness, and since we are working over $\mathbb{Z}$, we will give a proof of this proposition.

Proof of Proposition 2.3.7 As bigraded abelian groups,

$$\widehat{A}\otimes_{\mathcal{B}}(\mathcal{B}\otimes_R\widehat{D})\cong(\widehat{A}\otimes_{\mathcal{B}}\mathcal{B})\otimes_R\widehat{D}\cong\widehat{A}\otimes_R\widehat{D}.$$

Thus, $\widehat{A}\otimes_{\mathcal{B}}(\mathcal{B}\otimes_R\widehat{D})$ and $\widehat{A}\boxtimes\widehat{D}$ have the same underlying group; we must verify that the differentials agree.

The differential on $\widehat{A}\otimes_{\mathcal{B}}(\mathcal{B}\otimes_R\widehat{D})$ may be written as

$$d_{\widehat{A}}\otimes(|\mathrm{id}_{\mathcal{B}}|\otimes|\mathrm{id}_{\widehat{D}}|)+\mathrm{id}_{\widehat{A}}\otimes(\mu_1\otimes|\mathrm{id}_{\widehat{D}}|+(\mu_2\otimes\mathrm{id}_{\widehat{D}})\circ(\mathrm{id}_{\mathcal{B}}\otimes\delta))$$
$$=d_{\widehat{A}}\otimes(|\mathrm{id}_{\mathcal{B}}|\otimes|\mathrm{id}_{\widehat{D}}|)+\mathrm{id}_{\widehat{A}}\otimes(\mu_1\otimes|\mathrm{id}_{\widehat{D}}|)$$
$$+(\mathrm{id}_{\widehat{A}}\otimes(\mu_2\otimes\mathrm{id}_{\widehat{D}}))\circ(\mathrm{id}_{\widehat{A}}\otimes(\mathrm{id}_{\mathcal{B}}\otimes\delta)).$$

Regrouping the parentheses, we get

$$(d_{\widehat{A}}\otimes|\mathrm{id}_{\mathcal{B}}|)\otimes|\mathrm{id}_{\widehat{D}}|+(\mathrm{id}_{\widehat{A}}\otimes\mu_1)\otimes|\mathrm{id}_{\widehat{D}}|+((\mathrm{id}_{\widehat{A}}\otimes\mu_2)\otimes\mathrm{id}_{\widehat{D}})\circ((\mathrm{id}_{\widehat{A}}\otimes\mathrm{id}_{\mathcal{B}})\otimes\delta).$$

Identifying $\widehat{A}$ with $\widehat{A}\otimes_{\mathcal{B}}\mathcal{B}$, the differential on $\widehat{A}$ becomes $d_{\widehat{A}}\otimes|\mathrm{id}_{\mathcal{B}}|+\mathrm{id}_{\widehat{A}}\otimes\mu_1$. Similarly, the algebra multiplication $m\colon \widehat{A}\otimes\mathcal{B}\to\widehat{A}$ becomes

$$\mathrm{id}_{\widehat{A}}\otimes\mu_2\colon(\widehat{A}\otimes_{\mathcal{B}}\mathcal{B})\otimes\mathcal{B}\to(\widehat{A}\otimes_{\mathcal{B}}\mathcal{B}).$$

Thus, we can identify the above formula for the differential on $\widehat{A}\otimes_{\mathcal{B}}(\mathcal{B}\otimes_R\widehat{D})\cong\widehat{A}\otimes_R\widehat{D}$ with

$$d_{\widehat{A}}\otimes|\mathrm{id}_{\widehat{D}}|+(m\otimes\mathrm{id}_{\widehat{D}})\circ(\mathrm{id}_{\widehat{A}}\otimes\delta).$$

This is also the differential on $\widehat{A}\boxtimes\widehat{D}$. $\square$

3 Khovanov's functor-valued invariant as a bordered theory

We will assume some familiarity with Khovanov's paper [4]. Here we briefly introduce some useful conventions and notation.

Khovanov's arc algebra H^n has one $\mathbb{Z}$–grading. We will view H^n as a differential bigraded algebra, concentrated in homological degree 0 and with no differential. The usual $\mathbb{Z}$–grading on H^n becomes the intrinsic component of the bigrading. The intrinsic gradings of H^n are nonnegative. Thus, both Proposition 2.2.4 and Proposition 2.2.5 apply to H^n as a dg algebra.

The component of H^n in degree 0 (or, with our conventions, in bidegree $(0,0)$) will be denoted $\mathcal{I}_n$ and referred to as the idempotent ring of H^n. It is isomorphic to $\mathbb{Z}^{\times(C_n)}$, where C_n is the n^{th} Catalan number. The elementary idempotents of H^n are the idempotents 1_a described by [4, Section 2.4]. The index a runs over elements of the set B^n of crossingless matchings of $2n$ points; since this set will be important later, we recall its definition here.

3.0.1 Definition Let P be a set of $2n$ distinct points $p_1, \dots, p_{2n}$ on the line $\{0\}\times\mathbb{R} \subset \mathbb{R}\times\mathbb{R}$, ordered from top to bottom. A *crossingless matching* a of P is a partition of P into n pairs of points, such that there exists an embedding of n arcs $[0,1]^{\sqcup n}$ disjointly into $\mathbb{R}_{\geq 0}\times\mathbb{R}$ with each arc connecting a pair of points matched in a. The set of crossingless matchings of $2n$ points will be denoted B^n (different choices of P yield canonical bijections between the relevant sets B^n).

3.0.2 Remark The set B^n is also in bijection with the set NC_n of noncrossing partitions of n points. A noncrossing partition a of a set Q of n points $q_1, \dots, q_n$ on $\{0\}\times\mathbb{R}$ (ordered from top to bottom again) is defined to be any partition of Q into k disjoint subsets, such that there exists an embedding of k acyclic graphs disjointly into $\mathbb{R}_{\geq 0}\times\mathbb{R}$, with each graph bounding one of the k subsets of $\{q_1, \dots, q_n\}$ comprising a.

To go from a crossingless matching a of $2n$ points $p_1, \dots, p_{2n}$ to a noncrossing partition a' of n points $q_1, \dots, q_n$, checkerboard color the half-plane $\mathbb{R}_{\geq 0}$ with respect to some embedding of arcs representing a, such that the unbounded region of the half-plane is colored white. Without loss of generality, we may put the point q_i on the line $\{0\}\times\mathbb{R}$ between the points p_{2i-1} and p_{2i}. In the noncrossing partition a', two points q_i and q_j are placed in the same subset if they can be connected in $\mathbb{R}_{\geq 0}\times\mathbb{R}$ by a path through the black region of the checkerboard coloring. The skeleton of the black region provides the planar graphs which verify that a' is a noncrossing partition.

On the other hand, given a noncrossing partition a' of n points, one can pick an embedding of graphs representing a', and fatten each graph to obtain a planar surface.

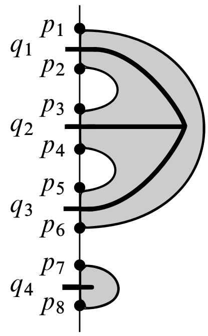

Figure 1: Bijection between crossingless matchings on $2n$ points and noncrossing partitions on n points

The boundary of this surface is a crossingless matching of $2n$ points. These two constructions are inverse to each other; see Figure 1 for an illustration.

Let T be an oriented tangle diagram in the half-plane, with $2n$ endpoints, and assume we have chosen an ordering of the crossings of T. Khovanov's construction assigns a bounded chain complex of finitely generated projective graded H^n–modules, with H^n–linear differential maps, to T. We will call this complex $[T]^{\mathrm{Kh}}$; we will often view $[T]^{\mathrm{Kh}}$ as a dg H^n–module using Proposition 2.2.4. If T lies in $\mathbb{R}_{\geq 0} \times \mathbb{R}$, then $[T]^{\mathrm{Kh}}$ is a left dg module; if T lies in $\mathbb{R}_{\leq 0} \times \mathbb{R}$, then $[T]^{\mathrm{Kh}}$ is a right dg module.

If T_1 is an oriented tangle diagram in $\mathbb{R}_{\geq 0} \times \mathbb{R}$, Proposition 2.2.5 gives us an isomorphism class of type D structures $\widehat{D}(T_1)$ over H^n, such that

$$[T_1]^{\mathrm{Kh}} \cong H^n \otimes_{\mathcal{I}_n} \widehat{D}(T_1).$$

As we will see in Section 3.1, Khovanov's construction of $[T_1]^{\mathrm{Kh}}$ naturally gives us an explicit type D structure $\widehat{D}(T_1)$ with this property.

If T_2 is an oriented tangle diagram in $\mathbb{R}_{\leq 0} \times \mathbb{R}$, we will simply take the type A structure $\widehat{A}(T_2)$ of T_2 to be the right dg module $[T_2]^{\mathrm{Kh}}$. Suppose T_2 and T_1 have consistent orientations; put them together to obtain an oriented link diagram L. Order the crossings of L so that those of T_1 come before those of T_2 and let $\mathrm{CKh}(L)$ be the Khovanov complex of L. Khovanov shows in [4] that

$$\mathrm{CKh}(L) \cong [T_2]^{\mathrm{Kh}} \otimes_{H^n} [T_1]^{\mathrm{Kh}},$$

after multiplying the intrinsic gradings on $[T_2]^{\mathrm{Kh}} \otimes_{H^n} [T_1]^{\mathrm{Kh}}$ by -1. By Proposition 2.3.7, we have

$$\mathrm{CKh}(L) \cong \widehat{A}(T_2) \boxtimes \widehat{D}(T_1),$$

as in bordered Floer homology, after applying the same intrinsic-grading reversal to $\widehat{A}(T_2) \boxtimes \widehat{D}(T_1)$. We will summarize this discussion more formally below in Theorem 3.2.1.

3.0.3 Remark The reversal of the gradings here comes from Khovanov's choice [4, pages 672–673] to make H^n positively rather than negatively graded. It is only a convention; one could define the basic generators of H^n to live in degrees -1 and -2, rather than 1 and 2, and then no grading reversal would be necessary.

3.0.4 Remark Up to isomorphism, the bigraded chain complex $\mathrm{CKh}(L)$ does not depend on the ordering of the crossings. Indeed, suppose we reverse the ordering of two adjacent crossings i and $i+1$. Then an isomorphism

$$F\colon (\mathrm{CKh}(L),\ \text{first ordering}) \to (\mathrm{CKh}(L),\ \text{second ordering})$$

can be defined, on the summand of $\mathrm{CKh}(L)$ corresponding to a vertex ρ of the cube of resolutions, to be $F := (-1)^{f(\rho)} \cdot \mathrm{id}$, where $f(\rho) := 1$ if ρ resolves crossings i and $i+1$ both as 1, rather than 0, and $f(\rho) := 0$ otherwise.

The same argument applies unchanged to the tangle complexes $[T]^{\mathrm{Kh}}$: the isomorphism type of $[T]^{\mathrm{Kh}}$ does not depend on the ordering of the crossings.

3.0.5 Remark Khovanov avoids having to choose an ordering of the crossings by using the skew-commutative cubes formalism. We will not do this here, but we will usually suppress mention of the choice of ordering of the crossings.

3.1 Type D structures

In this section we give a concrete definition of $\widehat{D}(T_1)$. First, we recall some properties of H^n and Khovanov's dg modules $[T]^{\mathrm{Kh}}$. This section has some overlap with the author's PhD thesis [8, Sections 4.2–4.3].

The algebra H^n has an additive basis β, over $\mathbb{Z}$, consisting of elements which we will denote $((W(a)b), \sigma)$. Here, a and b are elements of B^n, the set of crossingless matchings of $2n$ points, and the operation W mirrors the matching from the right half-plane to the left half-plane. The horizontal concatenation $W(a)b$ is a collection of disjoint circles in $\mathbb{R}^2$. The remaining datum σ consists of a choice of sign, $+$ or $-$, on each of these circles. For $a \in B^n$, the idempotent 1_a is $(W(a)a, \text{all plus})$. Multiplication in H^n is defined using minimal cobordisms and a two-dimensional topological quantum field theory; we refer the reader to Khovanov [4] for a precise definition.

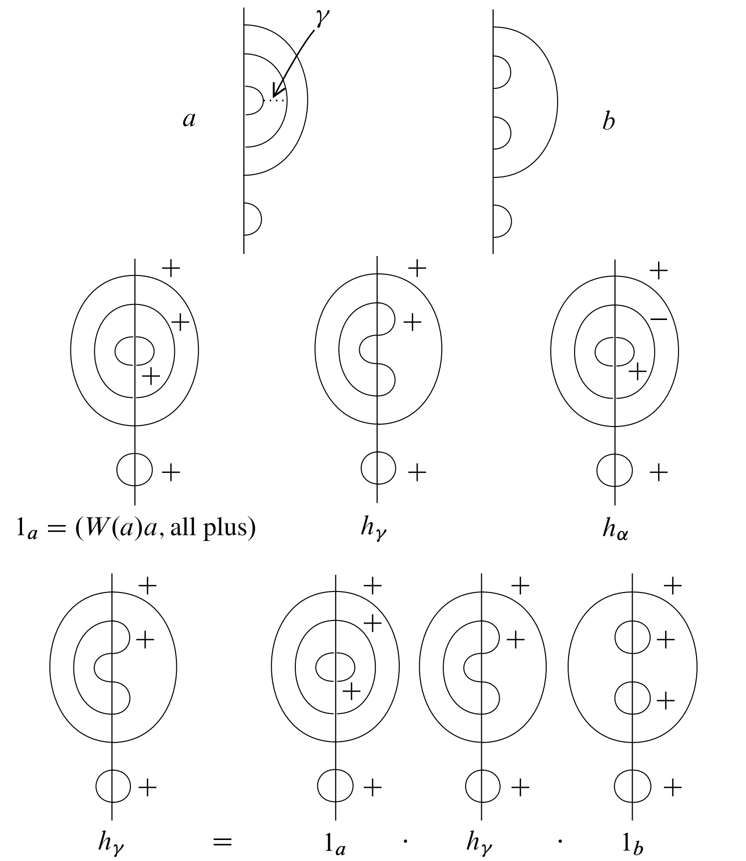

Figure 2: Top line: crossingless matchings a and b, with b obtained by surgering a bridge γ in a. Second line: the idempotent element $(W(a)a, \text{all plus})$ and the multiplicative generators $h_\gamma = (W(a)b, \text{all plus})$ and $h_\alpha = (W(a)a,\ \text{minus on } W(\alpha)\alpha)$. Third line: left and right idempotents of h_γ.

Below we use the notion of a bridge of a crossingless matching; see Roberts [12, Definition 8]. The dotted arc γ in the first line of Figure 2 is a bridge of the crossingless matching a.

Certain of the basis elements $((W(a)b), \sigma)$ form a natural set of multiplicative generators for H^n. These generators come in two forms: the first are elements $h_\gamma = (W(a)b, \text{all plus})$, where $a \in B^n$, the element $b \in B^n$ is obtained from a by surgering one pair of arcs along a bridge γ, and all circles of $W(a)b$ are labeled $+$. The other generators are elements $h_\alpha = (W(a)a,\ \text{minus on } W(\alpha)\alpha)$, where $a \in B^n$ and all circles of $W(a)a$ are labeled $+$ except one circle, $W(\alpha)\alpha$ for some arc α of a, which is labeled $-$. Each generator h_γ and h_α has a unique left idempotent and right

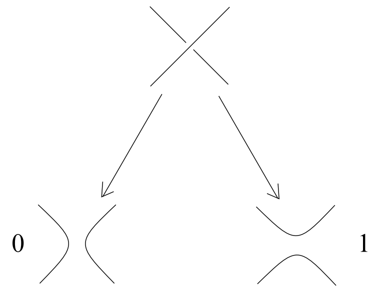

Figure 3: Zero- and one-resolutions of a crossing

idempotent in $\mathcal{I}_n$. We will denote the set of multiplicative generators $\{h_\gamma, h_\alpha\}$ as β_{mult}; it is a subset of β. Some examples are shown in Figure 2.

3.1.1 Proposition *The elements h_γ and h_α of β_{mult} generate H^n multiplicatively.*

Proof We have not defined the multiplication on H^n here, so we will only sketch the proof. It suffices to show that any element of the form $(W(a)b),$ all plus$)$ may be written as a product of h_γ generators; the rest of the elements of β may then be obtained using these elements and h_α generators.

The element $(W(a)b),$ all plus$)$ may be identified with a disjoint union of disks embedded in $D^2 \times I$ with boundary restricting to a on $D^2 \times \{0\}$, b on $D^2 \times \{1\}$, and $2n$ straight lines on $(\partial D^2) \times I$ (in other words, a cobordism from a to b). Here we identify crossingless matchings in the right half-plane with crossingless matchings in D^2; see Figure 5 below.

We may assume (after an isotopy if necessary) that the I–coordinate of $D^2 \times I$ gives a Morse function on the disjoint union of disks which has only index-1 critical points, each of which occurs at a distinct value of the I–coordinate. In such a configuration, the disjoint union of disks can be viewed as a composition of elementary saddle cobordisms beginning at a and ending at b. Each saddle cobordism corresponds to a generator h_γ. Furthermore, the composition of the saddle cobordisms in this sense agrees with the result of multiplying the elements h_γ in H^n using minimal cobordisms; see [8, Figure 4.3] for an illustration of this fact. Thus, we may write $(W(a)b),$ all plus $)$ as a product of generators h_γ. $\square$

Let T be an oriented tangle diagram in $\mathbb{R}_{\geq 0} \times \mathbb{R}$. To specify a generator x_i of $[T]^{\mathrm{Kh}}$, we first specify a resolution ρ_i of all crossings of T; we can view ρ_i as a function from the set of crossings to the two-element set $\{0, 1\}$ (see Figure 3). If T_{ρ_i} denotes the diagram T with the crossings resolved according to ρ_i, then T_{ρ_i} consists of a

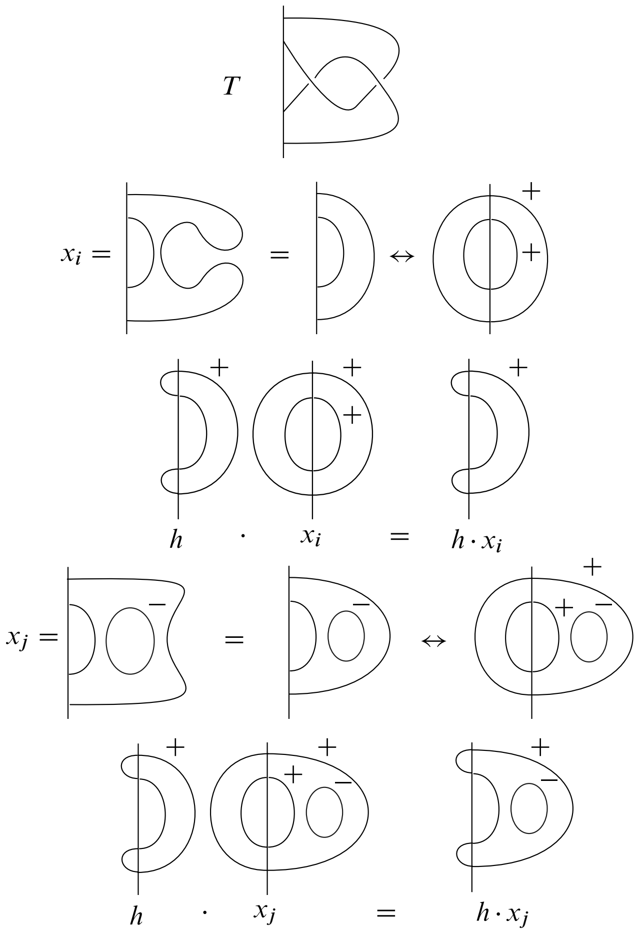

Figure 4: Some generators x_i and $h \cdot x_i$ of $[T]^{\mathrm{Kh}}$

crossingless matching of $2n$ points together with some circles contained in $\mathbb{R}_{>0} \times \mathbb{R}$. Following Roberts [12], these circles will be called free circles. The remaining data needed to specify x_i are a choice of $+$ or $-$ on each free circle.

Identify x_i with the diagram obtained by gluing the mirror of the crossingless-matching part of T_{ρ_i} to the left side of T_{ρ_i} and labeling all resulting circles with $+$. Then $[T]^{\mathrm{Kh}}$ has a $\mathbb{Z}$–basis consisting of elements $h \cdot x_i$, where the right idempotent of h agrees with the crossingless-matching part of T_{ρ_i}. (By multiplying h with x_i in Khovanov's usual minimal-cobordism sense, one obtains the basis for $[T]^{\mathrm{Kh}}$ described in [4].) See Figure 4 for an illustration of the generators x_i and basis elements $h \cdot x_i$.

The remainder of this section may also be found in the author's thesis [8, Section 4.3.1], with minor modifications.

3.1.2 Definition [8, Definition 4.3.1] Let T_1 be an oriented tangle diagram in $\mathbb{R}_{\geq 0}\times\mathbb{R}$, with n_+ positive crossings and n_- negative crossings, and let $0\leq r\leq n_+ + n_-$. Define $\widehat{D}(T_1)$ to be generated as an (intrinsically) graded abelian group, in homological degree $r-n_-$, by the generators $1\cdot x_i$ of $([T_1]^{\mathrm{Kh}})_{r-n_-}$, where $([T_1]^{\mathrm{Kh}})_{r-n_-}$ is the chain space of $[T_1]^{\mathrm{Kh}}$ in degree $r-n_-$. These generators have the same crossingless matching on the left and right sides of $\{0\}\times\mathbb{R}$, and all circles touching the boundary line have a $+$ sign. With this definition, $\widehat{D}(T_1)$ is an $\mathcal{I}_n$–submodule of $[T_1]^{\mathrm{Kh}}$.

3.1.3 Proposition [8, Proposition 4.3.2] *As H^n–modules,*

$$[T_1]^{\mathrm{Kh}} \cong H^n \otimes_{\mathcal{I}_n} \widehat{D}(T_1),$$

with $\widehat{D}(T_1)$ as defined in Definition 3.1.2.

Proof This follows from Definition 3.1.2. □

Let $\iota := \iota_{\widehat{D}(T_1)}$ (the inclusion of $\widehat{D}(T_1)$ into $[T_1]^{\mathrm{Kh}}$) and let d be the differential on $[T_1]^{\mathrm{Kh}}$. Let μ denote the multiplication on H^n.

3.1.4 Definition [8, Definition 4.3.3] The type D differential δ on $\widehat{D}(T_1)$ is defined by restricting the differential d to the $\mathcal{I}_n$–submodule $\widehat{D}(T_1)$ of $[T_1]^{\mathrm{Kh}}$:

$$\delta := \widehat{D}(T_1) \xrightarrow{\iota} [T_1]^{\mathrm{Kh}} \xrightarrow{d} [T_1]^{\mathrm{Kh}} \cong H^n \otimes_{\mathcal{I}_n} \widehat{D}(T_1).$$

It is an $\mathcal{I}_n$–linear map because ι and d are.

Lemma 3.1.5 and Proposition 3.1.6 below follow from the proof of Proposition 2.2.5, but we give short justifications to keep this section self-contained.

3.1.5 Lemma [8, Lemma 4.3.4] *Under the identification $[T_1]^{\mathrm{Kh}} \cong H^n \otimes_{\mathcal{I}_n} \widehat{D}(T_1)$ from Proposition 3.1.3, we have*

$$d = (\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta),$$

where $\mu_2\colon H^n \otimes H^n \to H^n$ is the algebra multiplication.

Proof Let $h\cdot\iota(x)$ denote a generator of $[T_1]^{\mathrm{Kh}}$. Then, by the Leibniz property for $[T_1]^{\mathrm{Kh}}$, we have $d(h\cdot\iota(x)) = h\cdot d\iota(x)$, since H^n has no differential. But since $d\iota(x)=\delta(x)$, we can conclude that

$$d(h\cdot\iota(x)) = (\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta)(h\cdot\iota(x)).$$ □

3.1.6 Proposition [8, Proposition 4.3.5] *$(\widehat{D}(T_1), \delta)$ satisfies the type D relations:*

$$(\mu_1 \otimes |\mathrm{id}|) \circ \delta + (\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta) \circ \delta = 0.$$

Proof There is no differential on H^n, so the μ_1 term is zero. For the other term, if x is a generator of $\widehat{D}(T_1)$, then

$$((\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta) \circ \delta)(x) = (d) \circ (d \circ \iota)(x) = 0,$$

since $d^2 = 0$ on $[T_1]^{\mathrm{Kh}}$. □

In Definition 3.1.2 and Definition 3.1.4, we constructed $\widehat{D}(T_1)$ as an $\mathcal{I}_n$–submodule of $[T]^{\mathrm{Kh}}$ following the proof of Proposition 2.2.5. Thus, the isomorphism class of $\widehat{D}(T_1)$ agrees with the isomorphism class of type D structures obtained from $[T]^{\mathrm{Kh}}$ by using Proposition 2.2.5.

3.2 Type A structures and pairing

Let T_2 be an oriented tangle diagram in $\mathbb{R}_{\leq 0} \times \mathbb{R}$. Since dg modules are special cases of type A structures, Proposition 2.2.4 tells us that the right dg module $[T_2]^{\mathrm{Kh}}$ is a valid example of a type A structure over H^n. We define $\widehat{A}(T_2)$ to be $[T_2]^{\mathrm{Kh}}$.

3.2.1 Theorem (Theorem 1.0.1) *Let T_1 and T_2 be oriented tangle diagrams in $\mathbb{R}_{\geq 0} \times \mathbb{R}$ and $\mathbb{R}_{\leq 0} \times \mathbb{R}$, respectively, with orderings chosen of the crossings of T_1 and T_2. Assume that T_1 and T_2 have consistent orientations, so that their horizontal concatenation is an oriented link diagram L in $\mathbb{R}^2$. Order the crossings of L such that those of T_1 come before those of T_2. Then*

$$\mathrm{CKh}(L) \cong \widehat{A}(T_2) \boxtimes \widehat{D}(T_1),$$

after multiplying the intrinsic gradings on $\widehat{A}(T_2) \boxtimes \widehat{D}(T_1)$ by -1.

Proof Since, up to a grading reversal, $\mathrm{CKh}(L) \cong [T_2]^{\mathrm{Kh}} \otimes_{H^n} [T_1]^{\mathrm{Kh}}$, which is the same as $\widehat{A}(T_2) \otimes_{H^n} (H^n \otimes_{\mathcal{I}_n} \widehat{D}(T_1))$, this proposition follows from Proposition 2.3.7, Remark 3.0.3, Lemma 3.1.5 and Khovanov's results from [4]. □

4 Quadratic and linear-quadratic algebras and duality

4.1 Quadratic and linear-quadratic algebras

We now consider a method of describing algebras using explicit generators and relations. It will be important for the following sections where we relate the bordered Khovanov theory discussed above with Roberts' constructions in [11; 12]. The definitions and basic properties of quadratic and linear-quadratic algebras here all follow Polishchuk and Positselski [10], with some minor modifications.

Let $\mathcal{B}$ be a unital associative algebra over a ring R, where $R \cong \mathbb{Z}e_1 \times \cdots \times \mathbb{Z}e_k$ as above. We will not assume $\mathcal{B}$ is graded; however, we will assume $\mathcal{B}$ comes equipped with an *augmentation*, ie an algebra homomorphism ϵ from $\mathcal{B}$ to the coefficient ring R. The algebras of interest to us have a grading of some form, and R is the degree-zero summand. Such an algebra has a natural augmentation given by projection onto this summand.

Suppose $b_1, \dots, b_m$ is a set of multiplicative generators of $\mathcal{B}$, each in the kernel $\mathcal{B}_+$ of ϵ. We may assume that for each b_i, there is a unique idempotent e_j such that $e_j b_i = b_i$ and $e'_j b_i = 0$ for $j' \neq j$. Indeed, if $e_j b_i = 0$ for all j, then $b_i = 0$ so b_i is irrelevant as a generator, and if $e_{j_a} b_i$ were nonzero for multiple indices a, we could remove b_i from the list of generators and add each of the nonzero elements $e_{j_a} b_i$ to the list. So we may assume $e_j b_i \neq 0$ for exactly one j, and then $b_i = 1 b_i = (\sum_{j'} e_{j'}) b_i = e_j b_i$. The idempotent e_j will be called the left idempotent of b_i and denoted $e_L(b_i)$.

Similarly, we may further assume that for each b_i, there exists a unique right idempotent $e_R(b_i)$ such that $b_i e_R(b_i) = b_i$ and $b_i e_j = 0$ for $e_j \neq e_R(b_i)$.

Let V be the free $\mathbb{Z}$–module spanned by $\{b_1, \dots, b_m\}$. The assumptions above equip V with left and right module structures over R. The statement that the b_i generate $\mathcal{B}$ multiplicatively means that $\mathcal{B}$ is isomorphic to $T(V)/J$, where

$$T(V) = \bigoplus_{n \geq 0} T^n(V) = R \oplus V \oplus (V \otimes_R V) \oplus (V \otimes_R V \otimes_R V) \oplus \cdots$$

and J is the kernel of the natural map $T(V) \to \mathcal{B}$ sending a string of generators to their product in $\mathcal{B}$. As above, we may assume that each generator of the ideal J has unique left and right idempotents.

4.1.1 Definition The augmented algebra $\mathcal{B}$, with its choice of generators, is a *quadratic algebra* if the ideal of relations $J \subset T(V)$ is generated multiplicatively by its intersection with $T^2(V) = V \otimes_R V$. In other words,

$$J = T(V) \cdot I \cdot T(V),$$

where $I := J \cap (V \otimes_R V)$. Note that J always contains $T(V) \cdot I \cdot T(V)$, so $\mathcal{B}$ is a quadratic algebra if $J \subset T(V) \cdot I \cdot T(V)$.

4.1.2 Remark If $\mathcal{B}$ is a quadratic algebra, then $\mathcal{B}$ obtains a grading by word-length in the generators $\{b_1, \dots, b_m\}$.

4.1.3 Remark Let $\mathcal{B}$ be a quadratic algebra. At various points it will be helpful to work with the generators and relations of $\mathcal{B}$ more explicitly. Following Polishchuk and Positselski [10, Chapter 4.1], choose an ordering of the multiplicative generators $\{b_1, \dots, b_m\}$; we may assume that $b_i < b_j$ when $i < j$. Use this order to put a lexicographic ordering on monomials in these generators: the leftmost factor in a product is defined to be the most significant part.

Let Q denote the set of quadratic monomials in the b_i. Then Q can be naturally partitioned into two subsets Q_1 and Q_2: Q_1 consists of the monomials which cannot be written in $\mathcal{B}$ as sums of lesser monomials with respect to the lexicographic order, and Q_2 consists of the monomials which can. If $b_i b_j \in Q_2$, then

$$b_i b_j = \sum_{(i',j')<(i,j)} c_{i,j;i',j'} b_{i'} b_{j'}$$

and the coefficients $c_{i,j;i',j'}$ are uniquely determined if we require that $c_{i,j;i',j'} = 0$ for $b_{i'} b_{j'}$ in Q_2. By [10, Lemma 1.1 of Chapter 4.1], a set of generators for the quadratic relation ideal $I = J \cap T^2(V)$ of $\mathcal{B}$ is obtained by taking

$$I_{i,j} := b_i b_j - \sum_{(i',j')<(i,j)} c_{i,j;i',j'} b_{i'} b_{j'}$$

for all (i, j) such that $b_i b_j$ is in Q_2.

4.1.4 Definition [10, Chapter 5.1] The augmented algebra $\mathcal{B}$, with its choice of generators, is a *linear-quadratic algebra* if the ideal of relations $J \subset T(V)$ is generated multiplicatively by its intersection with $T^1(V) \oplus T^2(V)$. In other words, writing $J_2 := J \cap (V \oplus (V \otimes_R V))$, $\mathcal{B}$ is linear-quadratic if

$$J = T(V) \cdot J_2 \cdot T(V),$$

or equivalently

$$J \subset T(V) \cdot J_2 \cdot T(V).$$

We will furthermore assume that $J \cap V = 0$ so that there are no linear redundancies among the chosen generators. (As always, we assume that $V \subset \mathcal{B}_+$.)

4.1.5 Remark If $\mathcal{B}$ is a linear-quadratic algebra, we get a word-length filtration on $\mathcal{B}$ rather than a grading. An element of $\mathcal{B}$ has filtration level at most k if it is a sum of products of word-length at most k in the generators b_i.

The following definitions and propositions will be used in Section 4.3 to define quadratic duality for linear-quadratic algebras. They can all be found in [10, Chapter 5.1].

4.1.6 Definition Let $\mathcal{B}$ be a linear-quadratic algebra, so that $\mathcal{B} \cong T(V)/J$ with $J \subset T(V) \cdot J_2 \cdot T(V)$. The quadratic algebra $\mathcal{B}^{(0)}$ is defined as

$$\mathcal{B}^{(0)} := T(V)/(T(V) \cdot I \cdot T(V)),$$

where $I \subset T^2(V)$ is defined as the image of $J_2 \subset (V \oplus T^2(V))$ under the projection $(V \oplus T^2(V)) \to T^2(V)$ onto the second summand.

Every generator r of I is the image of some generator $v \oplus r$ of J_2, where $v \in V$. Furthermore, if $v \oplus r$ and $v' \oplus r$ were both in J_2 with $v \neq v'$, then $(v - v') \oplus 0$ would be a nonzero element of $J_2 \cap V$, contradicting the assumption that $J_2 \cap V = 0$. Thus, the following definition makes sense.

4.1.7 Definition The function $\varphi\colon I \to V$ is defined by sending a generator $r \in I$ to the unique element $\varphi(r)$ of V such that $\varphi(r) \oplus r$ is in J_2.

4.1.8 Proposition *The map φ respects the left and right R–actions on I and V.*

Proof Suppose e is the left idempotent of r (which exists without loss of generality). Then $e(\varphi(r) \oplus r)$ is in J_2, and $e(\varphi(r) \oplus r) = e\varphi(r) \oplus er = e\varphi(r) \oplus r$, so by the uniqueness above, $\varphi(r) = e\varphi(r)$. If e' is any idempotent not equal to e, then $e'(\varphi(r) \oplus r)$ is still in J_2, but now this expression equals $e'\varphi(r) \oplus 0$. Since $J_2 \cap V = 0$, we must have $e'\varphi(r) = 0$ for $e' \neq e$. Thus, φ respects the left R–action on I and V. The right action is analogous. □

Let φ^{12} denote $\varphi \otimes \mathrm{id}_V\colon I \otimes_R V \to V \otimes_R V$ and let φ^{23} denote $\mathrm{id}_V \otimes \varphi\colon V \otimes_R I \to V \otimes_R V$.

4.1.9 Proposition [10, Chapter 5.1, Proposition 1.1] *The map*

$$\varphi^{12} - \varphi^{23}\colon (V \otimes_R I) \cap (I \otimes_R V) \to (V \otimes_R V)$$

has image contained in I, and

$$\varphi \circ (\varphi^{12} - \varphi^{23}) = 0.$$

Proof The definition of φ implies that the image of the map

$$\varphi \oplus \iota\colon I \to (V \oplus (V \otimes_R V))$$

is contained in J_2, where ι denotes the inclusion map of I into $V \otimes_R V$. Thus, the map

$$(\varphi \oplus \iota) \otimes \mathrm{id}_V\colon I \otimes_R V \to (V \oplus (V \otimes_R V)) \otimes_R V = (V \otimes_R V) \oplus (V \otimes_R V \otimes_R V)$$

has image contained in J. On the other hand, $(\varphi \oplus \iota) \otimes \mathrm{id}_V$ is equal to the map

$$\varphi^{12} \oplus (\iota \otimes \mathrm{id}_V)\colon I \otimes_R V \to (V \otimes_R V) \oplus (V \otimes_R V \otimes_R V).$$

Thus, $\varphi^{12} \oplus (\iota \otimes \mathrm{id}_V)$ has image contained in J. By the same reasoning, $\varphi^{23} \oplus (\mathrm{id}_V \otimes \iota)$ has image contained in J as well.

If x is an element of $(V \otimes_R I) \cap (I \otimes_R V) \subset V^{\otimes 3}$, then we can apply $\varphi^{12} \oplus (\iota \otimes \mathrm{id}_V)$ and $\varphi^{23} \oplus (\mathrm{id}_V \otimes \iota)$ to x, producing two elements $\varphi^{12}(x) \oplus x$ and $\varphi^{23}(x) \oplus x$ of J. Subtracting, the x terms cancel and $\varphi^{12}(x) - \varphi^{23}(x)$ is also in J.

Since $\varphi^{12}(x) - \varphi^{23}(x)$ is an element of both J and $V \otimes_R V$, it is also an element of $J_2 = J \cap (V \oplus (V \otimes_R V))$. The corresponding element of I under the projection from J_2 to I is the same element $\varphi^{12}(x) - \varphi^{23}(x)$.

Hence we can conclude that $\varphi^{12} - \varphi^{23}$ has image contained in I, so it makes sense to postcompose this map with φ. Furthermore, the image of $\varphi^{12} - \varphi^{23}$ is contained not just in I but in J_2 and so $\varphi(\varphi^{12} - \varphi^{23}) = 0$. □

4.2 Khovanov's arc algebra as a linear-quadratic algebra

In this section we present a combinatorial result, Lemma 4.2.4, whose proof was found by Pálvölgyi [9] and independently by Potechin in an email correspondence. Besides being important for the constructions in Section 5 and Section 6, it will yield an explicit generators-and-relations description of H^n in Corollary 4.2.7. This description is not necessary, strictly speaking, for Section 5 and Section 6, but it may be of interest independently.

Let V be the free $\mathbb{Z}$–module spanned by the generators h_γ and h_α of β_{mult} as defined in Section 3.1. The idempotent ring $R = \mathcal{I}_n$ of H^n has both left and right actions on V. We may write H^n as $T(V)/J$ for some ideal J of $T(V)$.

4.2.1 Theorem (Theorem 1.0.2) *With the generators $\{h_\gamma, h_\alpha\}$ and the augmentation coming from its grading, H^n is a linear-quadratic algebra.*

This theorem will be proved using Lemma 4.2.4. We begin with some background.

Recall from Remark 3.0.2 that the elementary idempotents of H^n are in bijection with the set NC_n of noncrossing partitions of n points. In fact, NC_n has a natural partial ordering: suppose p and q are elements of NC_n. Then $p \le q$ if p is a refinement of q. In other words, $p \le q$ if each of the subsets comprising p is contained in one of the subsets comprising q (recall that p and q are collections of subsets of a set of n points).

As a poset, NC_n is a lattice: any two noncrossing partitions have a unique least upper bound and a unique greatest lower bound, although we will not make use of this property.

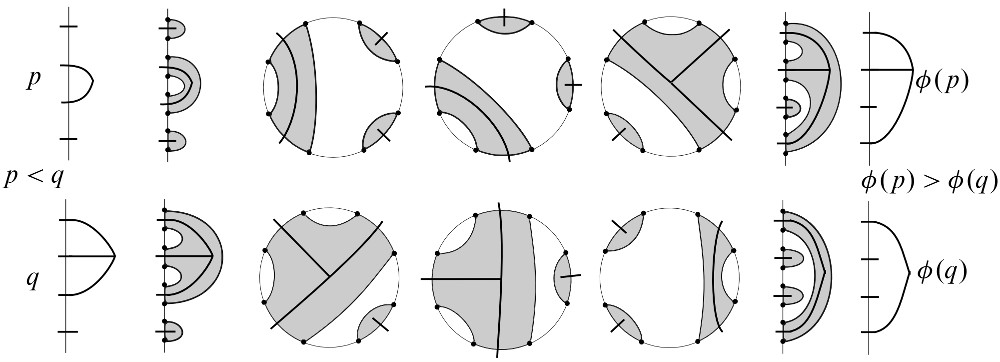

Figure 5: The order-reversing automorphism ϕ of NC_n (when read from left to right), or its inverse ϕ' (when read from right to left)

The dual of a partially ordered set is defined by reversing the order relations. It is a standard fact that the poset NC_n is self-dual.

4.2.2 Proposition NC_n *is order-isomorphic to the poset obtained by reversing all the order relations on* NC_n*.*

Proof We want to define a bijection $\phi\colon \mathrm{NC}_n \to \mathrm{NC}_n$ such that $p < q$ if and only if $\phi(p) > \phi(q)$. Let p be a noncrossing partition. Pick an embedding of acyclic graphs in the half-plane representing p; as in Remark 3.0.2, thicken these graphs to get planar surfaces embedded in the half-plane. Color the interiors of these surfaces black; then the half-plane is divided into black and white regions. Identify the half-plane with the disk and rotate the disk counterclockwise through an angle of π/n. Swap the colors of the regions and identify the disk back with the half plane. The skeleton of the new black region represents the noncrossing partition $\phi(p)$. This procedure is illustrated in Figure 5.

One may verify that ϕ, defined in this way, reverses the order relations. Finally, ϕ has an inverse whose definition is the same as for ϕ, except that the rotation is clockwise. □

Associated to the partial order on NC_n is a Hasse diagram G_n, which is a directed graph whose vertices are the elements of NC_n and which has an edge from p to q precisely when $p < q$ and there exists no vertex q' with $p < q' < q$.

We will view G_n as an undirected graph, ignoring the orientations on edges. For any two vertices p, q of G_n connected by an edge, there is a generator h_γ of H^n with left idempotent p and right idempotent q, and all the h_γ generators of H^n are of this form. Note that the existence of such h_γ does not depend on the ordering of p and q.

If $h_\gamma = (W(p)q, \text{all plus})$, where γ is a bridge on the crossingless matching p, then there is also a dual bridge $\gamma^\dagger$ on q; see Roberts [12, Definition 10]. Doing surgery on q along the bridge $\gamma^\dagger$ gives p, and the generator $h_{\gamma^\dagger} = (W(q)p,\ \text{all plus})$ has left idempotent q and right idempotent p.

Monomials in the generators h_γ either correspond to paths in G_n, or are zero for idempotent reasons. We will be especially concerned with paths of minimum length.

4.2.3 Definition Let p and q be vertices of G_n. The graph $G_{p,q}$ has one vertex for each minimal-length path, or geodesic, α from p to q in G_n. If α and β are two vertices of $G_{p,q}$, they are connected by an edge when α and β differ in exactly one vertex of G_n (viewing paths in G_n as sequences of vertices of G_n).

The proof of the following lemma was found by Dömötör Pálvölgyi and posted as an answer to a question on MathOverflow [9]; independently, another proof was found by Aaron Potechin and shared with the author privately in an email correspondence.

4.2.4 Lemma (Potechin [9]) *Let G_n denote the Hasse diagram of NC_n, viewed as an undirected graph. Let p, q be vertices of G_n and define $G_{p,q}$ as in Definition 4.2.3. Then $G_{p,q}$ is a connected graph.*

Proof First, note that as partitions of a set of n points, either q contains a singleton part or the dual $\phi(q)$ of q contains a singleton part. Indeed, out of the n points, consider a minimal pair of points which are matched in q. (We consider a pair to be minimal if there is no pair of points, also matched in q, nested inside the first pair.) If there are any points nested inside the minimal pair, then these points must be singletons by minimality, so q contains a singleton. On the other hand, if there are no points nested inside, then one can see from Figure 5 that $\phi(q)$ contains a singleton.

In the latter case, we can use Proposition 4.2.2 to reduce to the former case: if $G_{\phi(p),\phi(q)}$ is connected, then so is $G_{p,q}$. Thus, we may assume without loss of generality that q contains a singleton part, say $\{m\}$ where m is one of the n points on the line.

We will induct on both n and the distance between p and q; if this distance is 2 or less, or if $n \le 2$, there is nothing to prove.

Consider two minimal-length paths α and β from p to q in G_n. Since p is a partition of the n points on the line, the point m must be contained in one of the partitioning subsets which comprise p, say S. If S contains only m, then for each vertex along either α or β, the point m must be in a singleton set; otherwise α or β would not have minimal length (the length could be reduced by removing the steps that connect and disconnect m from the other points on the line). Hence we may ignore m and view α and β as paths in G_{n-1}. By induction, α may be modified one vertex at a time to produce β, and we may reintroduce the singleton point m without issue.

On the other hand, suppose S contains additional points as well as m. Then we may modify both α and β, one vertex at a time, to get paths α' and β' from p to q such that the first step of both α' and β' separates m from the other points in S. To do this, find the first step along α or β after which m is an isolated point and commute this step to the beginning of α or β by changing the path one vertex at a time.

Now let p' denote the partition p with the point m isolated from S. Both α' and β' start by moving from p to p' and then along a minimal-length path (say α'' or β'') from p' to q. Since the distance from p' to q is one less than the distance from p to q, we may conclude by induction that α'' may be modified one vertex at a time to obtain β''. Thus, the same is true for α' and β' and hence for α and β as well. □

4.2.5 Remark If p is the minimal element of NC_n with respect to the partial ordering, and q is the maximal element, then elements of $G_{p,q}$ are maximal chains in NC_n and Lemma 4.2.4 is a well-known result; see Bessis [2, Proposition 1.6.1], as well as Adin and Roichman [1] for more properties of $G_{p,q}$ in this case. Lemma 4.2.4 can be viewed as a generalization of this result to a setting in which p and q may not necessarily be comparable in the partial ordering.

Proof of Theorem 4.2.1 We want to show that $J \subset T(V) \cdot J_2 \cdot T(V)$. We start by exhibiting elements in the intersection $J_2 = J \cap (V \oplus (V \otimes_R V))$; we will let $\bar{J}_2$ denote the ideal generated by these elements. Recall that the notion of a bridge for a crossingless matching was defined in Roberts [12, Definition 8], and the dual bridge $\gamma^\dagger$ for a bridge γ was defined in [12, Definition 10].

(1) Whenever γ and η are two bridges which can be drawn without intersection on the same crossingless matching, the element $h_\gamma h_{\eta'} - h_\eta h_{\gamma'}$ is in $\bar{J}_2$, for the natural choices of η' and γ'.

(2) Whenever γ is a bridge and α is an arc such that h_γ and h_α have the same left idempotent, and neither of the endpoints of γ lies on the arc α, the element $h_\gamma h_{\alpha'} - h_\alpha h_\gamma$ is in $\bar{J}_2$ for the natural choice of α'. If one of the endpoints of the bridge γ lies on the arc α, then there are two natural choices for α'; for each of these choices, $h_\gamma h_{\alpha'} - h_\alpha h_\gamma$ is an element of $\bar{J}_2$.

(3) Whenever α_1 and α_2 are distinct arcs in the same crossingless matching, so that h_α and $h_{\alpha'}$ have the same left idempotent, the element $h_{\alpha_1} h_{\alpha_2} - h_{\alpha_2} h_{\alpha_1}$ is in $\bar{J}_2$. Furthermore, for every arc α, the element h_α^2 is in $\bar{J}_2$.

(4) Finally, if γ is any bridge, such that h_γ has left idempotent e_L and right idempotent e_R, the element $h_\gamma h_{\gamma^\dagger} - h_{\alpha_1} - h_{\alpha_2}$ is in $\bar{J}_2$, where α_1 and α_2 are the arcs containing the endpoints of γ.

To show that $J \subset T(V) \cdot J_2 \cdot T(V)$, it suffices to show that $J \subset T(V) \cdot \bar{J}_2 \cdot T(V)$, since $\bar{J}_2 \subset J_2$. Actually, $\bar{J}_2 = J_2$ (see Remark 4.2.6 below), but we will not need this fact in the current proof.

Let r be an arbitrary element of J. We may assume without loss of generality that r has a unique left idempotent e_L and right idempotent e_R. Since J is an ideal of the tensor algebra $T(V)$, r may be written as a linear combination of monomials in the generators h_γ and h_α. Let

$$r = \sum_i n_i (h_{i,1} \cdots h_{i,l_i}),$$

where $n_i \in \mathbb{Z}$ and each $h_{i,j}$ is one of the generators h_γ or h_α.

Consider one of the monomial summands $m_i = h_{i,1} \cdots h_{i,l_i}$ of r. After adding elements of $T(V) \cdot \bar{J}_2 \cdot T(V)$ to this monomial, we may assume that all the h_γ generators among the $h_{i,j}$ come before (ie with lower j than) the h_α generators. The necessary relation elements come from item (2) above. Let m'_i denote the monomial obtained by modifying m_i in this way.

Write m'_i as $m_{\gamma,i} \cdot m_{\alpha,i}$, where $m_{\gamma,i}$ is a product of h_γ generators and $m_{\alpha,i}$ is a product of h_α generators. Each $m_{\gamma,i}$ has left idempotent e_L and right idempotent e_R. We may view e_L and e_R as vertices of G_n, the undirected Hasse diagram of NC_n, and to the monomial $m_{\gamma,i}$ we may associate a path $p(m_{\gamma,i})$ in G_n from e_L to e_R.

We claim that we may further modify m'_i by adding elements of $T(V) \cdot \bar{J}_2 \cdot T(V)$ until $p(m_{\gamma,i})$ is a minimal-length path between e_L and e_R. Indeed, suppose $p(m_{\gamma,i})$ is a path of nonminimal length. Write $m_{\gamma,i} = h_{\gamma_1} \cdots h_{\gamma_k}$. Then there exists a minimal index $2 \le j \le k$ such that $h_{\gamma_1} \cdots h_{\gamma_{j-1}}$ corresponds to a path of minimal length in G_n but $h_{\gamma_1} \cdots h_{\gamma_j}$ does not.

Let $e_R(h_{\gamma_j})$ denote the right idempotent of h_{γ_j}. By assumption, the distance between e_L and $e_R(h_{\gamma_{j-1}})$ in G_n is $j-1$, but the distance between e_L and $e_R(h_{\gamma_j})$ is $j-2$ rather than j. Indeed, this distance must be less than j. It cannot be less than $j-2$, or the distance between e_L and $e_R(h_{\gamma_{j-1}})$ would be less than $j-1$. The distance between e_L and $e_R(h_{\gamma_j})$ also cannot be $j-1$, because each edge in G_n connects two noncrossing partitions whose number of parts differs by one modulo two. Hence this distance must be $j-2$.

Thus, there exists some monomial $h_{\gamma'_1} \cdots h_{\gamma'_{j-2}}$ corresponding to a path in G_n from e_L to $e_R(h_{\gamma_j})$. Appending $h_{\gamma_j^\dagger}$ to this monomial, we get $h_{\gamma'_1} \cdots h_{\gamma'_{j-2}} \cdot h_{\gamma_j^\dagger}$, which corresponds to a path of length $j-1$ in G_n between e_L and $e_R(h_{\gamma_{j-1}})$. By assumption, the distance between e_L and $e_R(h_{\gamma_{j-1}})$ is $j-1$, so $h_{\gamma'_1} \cdots h_{\gamma'_{j-2}} \cdot h_{\gamma_j^\dagger}$ corresponds to a minimal-length path in G_n.

We now have two minimal-length paths in G_n between e_L and $e_R(h_{\gamma_{j-1}})$, namely $\alpha = p(h_{\gamma_1} \cdots h_{\gamma_{j-1}})$ and $\beta = p(h_{\gamma'_1} \cdots h_{\gamma'_{j-2}} \cdot h_{\gamma_j^\dagger})$. By Lemma 4.2.4, we may modify α one vertex at a time to obtain β. Such modifications correspond, on the level of monomials, to adding relation terms obtained from item (1) above.

Thus, we may modify $m_{\gamma,i}$, which equals $h_{\gamma_1} \cdots h_{\gamma_j-1} \cdot h_{\gamma_j} \cdots h_{\gamma_k}$, by adding terms in $T(V) \cdot \bar{J}_2 \cdot T(V)$ to obtain $h_{\gamma'_1} \cdots h_{\gamma'_{j-2}} \cdot h_{\gamma_j^\dagger} \cdot h_{\gamma_j} \cdots h_{\gamma_k}$. Inside this monomial is $h_{\gamma_j^\dagger} \cdot h_{\gamma_j}$, which may be replaced with a sum of h_α terms using the relation terms in item (4) above. As before, these h_α terms may be commuted to the right side of m'_i using item (2).

After this modification, we have strictly reduced the length of $m_{\gamma,i}$ in the factorization of m'_i as $m_{\gamma,i} \cdot m_{\alpha,i}$. If the new $m_{\gamma,i}$ still does not represent a minimal-length path $p(m_{\gamma,i})$ in G_n, we can repeat the same procedure and eventually it will terminate.

At this point, we have shown that we can modify our original $r = \sum_i n_i(m_i)$ by adding terms in $T(V) \cdot \bar{J}_2 \cdot T(V)$, until each m_i is a monomial factorizable as $m_{\gamma,i} \cdot m_{\alpha,i}$ with $m_{\alpha,i}$ a monomial in the generators h_α and $m_{\gamma,i}$ a monomial in the generators h_γ representing a minimal-length path in G_n. The starting and ending points of all these paths are the same, namely the left and right idempotents of r. Thus, by Lemma 4.2.4 and the relations from item (1), we may do further modifications until all of the m_{γ_i} are the same monomial m_γ and we have

$$r = m_\gamma \sum_i n_i(m_{\alpha,i}) \quad \text{modulo } T(V) \cdot \bar{J}_2 \cdot T(V).$$

Let r' denote the right side of the above equality; r' is an element of $T(V)$ and we want to show that $r' = 0$ modulo $T(V) \cdot \bar{J}_2 \cdot T(V)$.

The monomial m_γ represents an element of H^n of the form $(W(a)b, \text{all plus})$, where $a := e_L$ is the left idempotent of r and $b := e_R$ is the right idempotent. The signs are all plus because m_γ corresponds to a path of minimal length. Indeed, by Proposition 3.1.1, $(W(a)b, \text{all plus})$ can be written as a product of h_γ generators. If m_γ represented a sum of basis elements $(W(a)b, \sigma)$ in H^n with σ not all plus, then the length of m_γ would be at least two plus the length of the product expansion of $(W(a)b, \text{all plus})$. This claim follows because the length of a monomial m_γ in the generators h_γ is equal to the grading of the corresponding element of H^n (assuming this element is nonzero). But then the product expansion of $(W(a)b, \text{all plus})$ would correspond to a shorter-length path from a to b than $p(m_\gamma)$, a contradiction. Finally, if m_γ were zero in H^n, then we could write $m_\gamma = \tilde{m}_\gamma \tilde{\tilde{m}}_\gamma$, where $\tilde{m}_\gamma$ is nonzero in H^n but becomes zero when multiplied on the right by the leftmost factor of $\tilde{\tilde{m}}_\gamma$. For this to be true, $\tilde{m}_\gamma$ must represent a sum of basis elements $(W(a)\tilde{b}, \sigma)$ with σ not all plus, where $\tilde{b}$ is the right idempotent of $\tilde{m}_\gamma$. By the above argument, we can obtain a shorter path than $p(\tilde{m}_\gamma)$ from a to $\tilde{b}$. Appending $p(\tilde{\tilde{m}}_\gamma)$, we get a shorter path from a to b than $p(m_\gamma)$, a contradiction.

Now we use the fact that $r \in J$, or in other words that $r = 0$ as an element of H^n. The same holds for r', since $T(V) \cdot \bar{J}_2 \cdot T(V)$ is a subset of J. The summand ${}_a(H^n)_b$ is a free abelian group with a basis element for every assignment of signs σ to the circles of $W(a)b$. Saying that $r' = 0$ in H^n means that the coefficient of r' on each of these basis elements is zero. In other words, for each assignment of signs σ to the circles of $W(a)b$, the sum of the terms $n_i m_\gamma m_{\alpha,i}$ of r' corresponding to σ is zero in H^n.

We will show that for a fixed σ, the terms $n_i m_\gamma m_{\alpha,i}$ such that $m_\gamma m_{\alpha,i}$ equals $(W(a)b, \sigma)$ in H^n actually sum to zero modulo the relation terms from items (2) and (3) above. There may also be some terms $m_\gamma m_{\alpha,i}$ which are already zero in H^n and thus which represent no basis element $(W(a)b, \sigma)$ of H^n. We will deal with these terms at the end.

Suppose $m_\gamma h_\alpha = m_\gamma h_{\alpha'} = (W(a)b, \sigma)$ in H^n, where m_γ corresponds to a minimal-length path; here α and α' are arcs in b which lie on the same circle in $W(a)b$, and σ assigns $-$ to this circle while assigning $+$ to all other circles of $W(a)b$. Then we may use relations from item (2) to write both $m_\gamma h_\alpha$ and $m_\gamma h_{\alpha'}$ as $h_{\tilde{\alpha}} m_\gamma$, where $\tilde{\alpha}$ is any arc in the left idempotent a of m_γ which, in $W(a)b$, lies in the same circle as α and α'. This generalization of the item (2) relations is true by induction on the length of γ.

Now, for a more general sum of terms $m_\gamma m_{\alpha,i}$ all representing $(W(a)b, \sigma)$ in H^n, we can use the above modifications to replace each of the monomials $m_{\alpha,i}$ with the same monomial m_α. We do this by picking, for example, $m_\alpha = m_{\alpha,i_1}$, and then for $i \neq i_1$, we move each factor of $m_{\alpha,i}$ to the left and back to the right so that it becomes identical to the factor appearing in m_{α,i_1}. After doing this for all i, we use relations from item (3) to replace each $m_{\alpha,i}$ with m_α.

For a fixed σ, let N_σ be the sum of the n_i such that $m_\gamma m_{\alpha,i}$ represents $(W(a)b, \sigma)$ in H^n. By the above paragraph, the sum of the terms $n_i m_\gamma m_{\alpha,i}$ of r' with $m_\gamma m_{\alpha,i}$ representing $(W(a)b, \sigma)$ in H^n is equivalent to $N_\sigma m_\gamma m_\alpha$ modulo $T(V) \cdot \bar{J}_2 \cdot T(V)$. We see that $N_\sigma m_\gamma m_\alpha = 0$ in H^n. But since $m_\gamma m_\alpha$ is the basis element of ${}_a(H^n)_b$ corresponding to σ, we can conclude that $N_\sigma = 0$. Thus, the sum of the terms $n_i m_\gamma m_{\alpha,i}$ of r' under consideration is equal to zero modulo $T(V) \cdot \bar{J}_2 \cdot T(V)$.

Finally, some of the terms $m_\gamma m_{\alpha,i}$ may not represent any $(W(a)b,\sigma)$ in H^n; this happens if and only if $m_\gamma m_{\alpha,i}$ is zero in H^n. In this case, by the above logic, we can use relations from items (2) and (3) to rearrange $m_\gamma m_{\alpha,i}$ until it has h_α^2 somewhere, for some generator h_α. Thus, these terms $m_\gamma m_{\alpha,i}$ are in $T(V)\cdot \bar{J}_2\cdot T(V)$ by item (3) above.

Starting with $r\in J$ above, we have successively modified r using linear-quadratic relations until we obtained zero. Hence

$$J\subset T(V)\cdot \bar{J}_2\cdot T(V)\subset T(V)\cdot J_2\cdot T(V)$$

and so H^n is a linear-quadratic algebra. □

4.2.6 Remark In fact, one can show by analyzing the grading possibilities case-by-case that the linear-quadratic relations listed above in (1)–(4) are a full set of generators for J_2. In other words, $\bar{J}_2 = J_2$.

We get a description of H^n in terms of generators and relations.

4.2.7 Corollary *Let V denote the free $\mathbb{Z}$–module spanned by the degree-1 generators h_γ and the degree-2 generators h_α of H^n, with left and right actions of $R=\mathcal{I}_n\cong\mathbb{Z}^{C_n}$ on V given by multiplication in H^n. Then*

$$H^n\cong T(V)/(T(V)\cdot J_2\cdot T(V)),$$

where the tensor products in $T(V)$ are over R, and $J_2=\bar{J}_2$ is generated by the explicit relations given above in items (1)–(4) of the proof of Theorem 4.2.1.

4.2.8 Remark All of the generators of J_2 listed in items (1)–(4) of the proof of Theorem 4.2.1 are homogeneous with respect to the intrinsic grading on H^n. Thus, Corollary 4.2.7 also gives us a description of H^n as a graded algebra. This grading differs from the word-length filtration which H^n acquires as a linear-quadratic algebra by Remark 4.1.5, even on the basic multiplicative generators: h_α has intrinsic degree 2 and word-length 1, while h_γ has intrinsic degree and word-length both equal to 1.

4.2.9 Remark Braden [3] gives a generators-and-relations description of a ring $A_{n,n}$ which has H^n as an idempotent truncation; see also Stroppel [13]. It would be interesting to compare Braden's generators and relations with the h_γ and h_α generators and relations discussed here; we have not tried to do this in any detail.

4.3 Quadratic duality

Next we discuss quadratic duality for quadratic and linear-quadratic algebras. The dual of a quadratic algebra $\mathcal{B}$ is another quadratic algebra $\mathcal{B}^!$. The dual of a linear-quadratic algebra $\mathcal{B}$ is a quadratic algebra $\mathcal{B}^!$ with a differential. Even if $\mathcal{B}$ is finitely generated over $\mathbb{Z}$, following Convention 2.1.1, the algebra $\mathcal{B}^!$ might be infinitely generated over $\mathbb{Z}$. Accordingly, Convention 2.1.1 will not be taken to hold for dual algebras $\mathcal{B}^!$ in general. However, $\mathcal{B}^!$ will still be generated multiplicatively by a finite set of elements.

We review the relevant definitions from Polishchuk and Positselski [10, Chapters 1 and 5]. We start with the case of quadratic algebras and then discuss the modification needed for linear-quadratic algebras. Let $\mathcal{B}$, R, b_i, V and J be defined as in Section 4.1.

Let V^* denote $\mathrm{Hom}_{\mathbb{Z}}(V,\mathbb{Z})$. Since V is a free $\mathbb{Z}$–module, V^* is free of the same rank as V. If b_i is a generator of V, let b_i^* denote the corresponding generator of V^*. We define left and right actions of R on V^* by declaring that b_i^* has the same left and right idempotents as b_i.

4.3.1 Definition Let $\mathcal{B}$ be a quadratic algebra and write $\mathcal{B}=T(V)/J$ as in Section 4.1, with $I := J\cap T^2(V)$. The *quadratic dual* $\mathcal{B}^!$ of $\mathcal{B}$ is defined to be

$$\mathcal{B}^! := T(V^*)/(T(V^*)\cdot I^{\perp}\cdot T(V^*)),$$

where $I^{\perp}$ is the submodule of $T^2(V^*)=V^*\otimes_R V^*$ annihilating I via the natural action of $V^*\otimes_R V^*$ on $V\otimes_R V$.

4.3.2 Remark Let Q_1, Q_2 and b_i be as defined as in Remark 4.1.3 above. We have a relation $I_{i,j}$ in I for every monomial b_ib_j in Q_2. If b_ib_j is in Q_1 rather than Q_2, consider instead the dual monomial $b_i^*b_j^*$ in $\mathcal{B}^!$. We can define a relation in $I^{\perp}$ by

$$I^!_{i,j} := b_i^*b_j^* + \sum_{(i',j')>(i,j)} c^!_{i,j;i',j'} b^*_{i'}b^*_{j'},$$

where $c^!_{i,j;i',j'}$ is only nonzero if $b_{i'}b_{j'}$ is in Q_2, in which case $c^!_{i,j;i',j'}$ is defined to be the coefficient $c_{i',j';i,j}$ of the $(i''=i,\ j''=j)$ term in the relation

$$I_{i',j'} = b_{i'}b_{j'} - \sum_{(i'',j'')<(i',j')} c_{i',j';i'',j''} b_{i''}b_{j''}.$$

The ideal $I^{\perp}$ is spanned by the relations $I^!_{i,j}$; like Remark 4.1.3, this follows from [10, Lemma 1.1 of Section 4.1].

We now extend quadratic duality to linear-quadratic algebras.

4.3.3 Definition [10, Chapter 5.4] Let $\mathcal{B}$ be a linear-quadratic algebra; recall that Definition 4.1.6 associates a quadratic algebra $\mathcal{B}^{(0)}$ to $\mathcal{B}$. The quadratic dual $\mathcal{B}^!$

of $\mathcal{B}$ is defined, as an algebra, to be the usual quadratic dual of $\mathcal{B}^{(0)}$. Since $\mathcal{B}^!$ is a quadratic algebra, it has a grading by word-length. We will interpret this grading as the homological grading for a differential μ_1 on $\mathcal{B}^!$.

We will first define μ_1 on the basis elements of V^* and extend to $\mathcal{B}^!$ with the Leibniz rule. We will use the map $\varphi\colon I \to V$ from Definition 4.1.7. Dualizing φ, we get $\varphi^*\colon V^* \to I^*$, where $I^* := \mathrm{Hom}_{\mathbb{Z}}(I, \mathbb{Z})$.

We claim that I^* is isomorphic to the degree-2 summand of $\mathcal{B}^!$. To see this, write the degree-2 summand of $\mathcal{B}^!$ as $T^2(V^*)/I^\perp = \mathrm{Hom}_{\mathbb{Z}}(V \otimes V, \mathbb{Z})/I^\perp$. There is a natural map Ξ from $\mathrm{Hom}_{\mathbb{Z}}(V \otimes V, \mathbb{Z})$ to I^* given by precomposing with the inclusion from I into $V \otimes V$. The map Ξ is surjective because any functional from I to $\mathbb{Z}$ may be extended to a functional from $V \otimes V$ to $\mathbb{Z}$. Indeed, using the conventions of Remark 4.1.3, the $\mathbb{Z}$–basis $\{I_{i,j} \mid b_i b_j \in Q_2\}$ for I may be extended to a $\mathbb{Z}$–basis $\{I_{i,j} \mid b_i b_j \in Q_2\} \cup Q_1$ for $V \otimes V$.

The kernel of Ξ, by definition, consists of those functionals on $V \otimes V$ which annihilate I. Thus, the kernel is the same as $I^\perp$. We can conclude that Ξ induces an isomorphism from the degree-2 summand of $\mathcal{B}^!$ to I^*.

Now, for a degree-1 element of $\mathcal{B}^!$, ie an element $v^* \in V^*$ dual to a basis element v of V, define $\mu_1(v^*)$ to be $\varphi^*(v^*)$. This is an element of I^* and thus a degree-2 element of $\mathcal{B}^!$.

We may extend μ_1 to a map from $\mathcal{B}^!$ to $\mathcal{B}^!$, homogeneous of degree $+1$, using the Leibniz rule

$$\mu_1(xy) = (-1)^{\deg x} \mu_1(x) y + x \mu_1(y).$$

Note that this Leibniz rule differs from the one used in Polishchuk and Positselski [10], to stay consistent with our earlier sign conventions.

4.3.4 Remark Suppose $\mathcal{B}$ is a linear-quadratic algebra with an intrinsic grading, whose augmentation map is induced from the grading. Suppose further that all the multiplicative generators b_i of $\mathcal{B}$ and the explicit generators $I_{i,j}$ of I from Remark 4.1.3 are homogeneous with respect to the intrinsic grading, and the map $\varphi\colon I \to V$ preserves intrinsic degree. For example, H^n satisfies these properties: the generators h_γ have intrinsic degree 1 and the generators h_α have intrinsic degree 2. Each term of each relation in items (1) and (4) of the proof of Theorem 4.2.1 has intrinsic degree 2. Those in item (2) have degree 3 and those in item (3) have degree 4. The map φ is only nonzero on relations from item (4), and sends elements of degree 2 to elements of degree 2.

With these assumptions, V^* has a natural intrinsic grading, namely the negative of the grading on V (so that the pairing of V^* with V is grading-preserving). The generators of $I^\perp$ are homogeneous with respect to this grading; this can be seen from Remark 4.3.2. Thus, $\mathcal{B}^!$ acquires an intrinsic grading. Since $\varphi\colon I \to V$ preserves intrinsic grading, so does $\varphi^*\colon V^* \to I^*$, and hence the differential μ_1 on $\mathcal{B}^!$ preserves intrinsic grading. The intrinsic grading on $\mathcal{B}^!$ is different from the homological grading, which μ_1 increases by one. In summary, $\mathcal{B}^!$ may be viewed as a differential bigraded algebra with a $(0,+1)$ differential.

4.4 The dual of Khovanov's arc algebra

Theorem 4.2.1, Definition 4.3.3 and Remark 4.3.4 together give us a differential bigraded algebra $(H^n)^!$, which we will call the dual of H^n.

4.4.1 Example When $n=1$, $H^n = H^1$ is the algebra $\mathbb{Z}[x]/x^2$ over the idempotent ring $\mathcal{I}_1 = \mathbb{Z}$. The generator x has intrinsic degree 2. Thus, the dual $(H^1)^!$ is $\mathbb{Z}[x^*]$, where x^* has bidegree $(-2,1)$. The differential on $(H^1)^!$ is zero and $(H^1)^!$ is not finitely generated over $\mathbb{Z}$.

In general, $(H^n)^!$ is never finitely generated over $\mathbb{Z}$, since arbitrary powers of any generator h_α^* will be nonzero in $(H^n)^!$.

4.5 Type *DD* bimodules

We may relate the duality discussed in Section 4.3 with the type DD bimodules encountered in bordered Heegaard Floer homology; see Lipshitz, Ozsváth and Thurston [6], especially Section 8. First, we give a definition of these bimodules over $\mathbb{Z}$; as in Section 2, we do not cover the most general possible case.

Let $\mathcal{B}$ and $\mathcal{B}'$ be differential bigraded algebras over an idempotent ring $R = \Pi_i(\mathbb{Z}e_i)$. The case $\mathcal{B}' = \mathcal{B}^!$ will be important, so we will not assume that $\mathcal{B}'$ is finitely generated over $\mathbb{Z}$.

Over $\mathbb{Z}/2\mathbb{Z}$, the following is equivalent to Definition 2.2.55 of Lipshitz, Ozsváth and Thurston [7].

4.5.1 Definition A *type DD bimodule* over $\mathcal{B}$ and $\mathcal{B}'$ is, first of all, a bigraded free $\mathbb{Z}$–module $\widehat{DD}$ with left and right actions of R, such that $\widehat{DD}$ admits a $\mathbb{Z}$–basis consisting of grading-homogeneous elements with unique left and right idempotents among the e_i. Furthermore, $\widehat{DD}$ must be equipped with an R–bilinear map

$$\delta\colon \widehat{DD} \to \mathcal{B} \otimes_R \widehat{DD} \otimes_R (\mathcal{B}')^{\mathrm{op}}$$

of degree $(0,+1)$, such that the type DD structure relations

$$(\mu_1 \otimes |\mathrm{id}| \otimes |\mathrm{id}|) \circ \delta + (\mathrm{id} \otimes |\mathrm{id}| \otimes \mu_1) \circ \delta + (\mu_2 \otimes \mathrm{id} \otimes \mu_2) \circ \sigma \circ (\mathrm{id} \otimes \delta \otimes \mathrm{id}) \circ \delta = 0$$

are satisfied, where μ_1 and μ_2 denote the differential and multiplication on $\mathcal{B}$ or $\mathcal{B}'$ as appropriate, and

$$\sigma\colon \mathcal{B} \otimes \mathcal{B} \otimes \widehat{DD} \otimes (\mathcal{B}')^{\mathrm{op}} \otimes (\mathcal{B}')^{\mathrm{op}} \to \mathcal{B} \otimes \mathcal{B} \otimes \widehat{DD} \otimes (\mathcal{B}')^{\mathrm{op}} \otimes (\mathcal{B}')^{\mathrm{op}}$$

is a sign flip which multiplies $b_1 \otimes b_2 \otimes x \otimes (b_3')^{\mathrm{op}} \otimes (b_4')^{\mathrm{op}}$ by $(-1)^{(\deg_h b_2)(\deg_h b_4')}$.

4.5.2 Remark The odd-seeming sign conventions reflect the fact that, while we write $\widehat{DD}$ with $\mathcal{B}$ on the left and $(\mathcal{B}')^{\mathrm{op}}$ on the right to make the notation more manageable, we really want to think of both $\mathcal{B}$ and $\mathcal{B}'$ being on the left of $\widehat{DD}$ when fixing sign conventions.

Of particular interest here are type DD bimodules with $\widehat{DD} = R$ as an R–bimodule. We will refer to these as *rank-one* DD bimodules, following the notation of [6, Section 8]. For a rank-one DD bimodule, we have

$$\mathcal{B} \otimes_R \widehat{DD} \otimes_R (\mathcal{B}')^{\mathrm{op}} \cong \mathcal{B} \otimes_R (\mathcal{B}')^{\mathrm{op}},$$

so we may rewrite the type DD structure relations as

$$(\mu_1 \otimes |\mathrm{id}|) \circ \delta + (\mathrm{id} \otimes \mu_1) \circ \delta + (\mu_2 \otimes \mu_2) \circ \sigma \circ (\mathrm{id} \otimes \delta \otimes \mathrm{id}) \circ \delta = 0,$$

where σ is now a map from $\mathcal{B} \otimes \mathcal{B} \otimes (\mathcal{B}')^{\mathrm{op}} \otimes (\mathcal{B}')^{\mathrm{op}}$ to itself.

When $\mathcal{B}$ is a linear-quadratic algebra with an intrinsic grading as in Remark 4.3.4, we can construct an associated rank-one DD bimodule over $\mathcal{B}$ and $\mathcal{B}^!$. Setting $\widehat{DD} = R$, we define $\delta\colon R \to \mathcal{B} \otimes_R (\mathcal{B}^!)^{\mathrm{op}}$ by

$$\delta(e) := \sum_i b_i \otimes (b_i^*)^{\mathrm{op}},$$

where e is one of the elementary idempotents and the sum runs over those multiplicative generators b_i of R which have left idempotent e. (These idempotent conditions will be implicit in what follows.) Note that δ has degree $(0,+1)$; it preserves the intrinsic grading, since the grading on $\mathcal{B}^!$ was defined to be the negative of that on $\mathcal{B}$, and it increases the homological grading by 1, since b_i has homological degree 0 while b_i^* has homological degree 1.

4.5.3 Proposition *The map δ, as defined above, satisfies the type DD structure relations.*

Proof First, let $e \in R$ be one of the elementary idempotents. Applying the term $(\mu_2 \otimes \mu_2) \circ \sigma \circ (\mathrm{id} \otimes \delta \otimes \mathrm{id}) \circ \delta$ to e, we get

$$(\mu_2 \otimes \mu_2) \circ \sigma \circ (\mathrm{id} \otimes \delta \otimes \mathrm{id}) \circ \delta(e) = \sum_{i,j} b_i b_j \otimes (b_j^*)^{\mathrm{op}} (b_i^*)^{\mathrm{op}},$$

where the sum runs over all pairs of multiplicative generators b_i, b_j of $\mathcal{B}$ with compatible idempotents, such that the left idempotent of b_i is e. Note that $\sigma = \mathrm{id}$ here, because the generators b_j all have homological degree zero.

In the notation of Remark 4.1.3, we may split the above sum as

$$\sum_{b_i b_j \in Q_1} b_i b_j \otimes (b_j^*)^{\mathrm{op}} (b_i^*)^{\mathrm{op}} + \sum_{b_i b_j \in Q_2} b_i b_j \otimes (b_j^*)^{\mathrm{op}} (b_i^*)^{\mathrm{op}}. \tag{4-1}$$

If $b_i b_j$ is in Q_1, then in $\mathcal{B}^!$, we may write $b_i^* b_j^*$ as

$$- \sum_{\substack{b_{i'} b_{j'} \in Q_2 \\ (i',j') > (i,j)}} c_{i',j';i,j} b_{i'}^* b_{j'}^*.$$

Thus,

$$\sum_{b_i b_j \in Q_1} b_i b_j \otimes (b_j^*)^{\mathrm{op}} (b_i^*)^{\mathrm{op}} = - \sum_{b_i b_j \in Q_1} b_i b_j \otimes \sum_{\substack{b_{i'} b_{j'} \in Q_2 \\ (i',j') > (i,j)}} c_{i',j';i,j} (b_{j'}^*)^{\mathrm{op}} (b_{i'}^*)^{\mathrm{op}}. \tag{4-2}$$

On the other hand, if $b_i b_j$ is in Q_2, then we may write $b_i b_j$ as

$$b_i b_j = \Bigg(\sum_{\substack{b_{i'} b_{j'} \in Q_1 \\ (i',j') < (i,j)}} c_{i,j;i',j'} b_{i'} b_{j'} \Bigg) - \varphi \Bigg(b_i b_j - \sum_{\substack{b_{i'} b_{j'} \in Q_1 \\ (i',j') < (i,j)}} c_{i,j;i',j'} b_{i'} b_{j'} \Bigg).$$

The expression $\varphi\big(b_i b_j - \sum_{b_{i'} b_{j'} \in Q_1, (i',j') < (i,j)} c_{i,j;i',j'} b_{i'} b_{j'}\big)$, or $\varphi(I_{i,j})$, denotes some linear combination of the multiplicative generators b_k of $\mathcal{B}$. Define integers $C_{i,j;k}$ by

$$\varphi \Bigg(b_i b_j - \sum_{\substack{b_{i'} b_{j'} \in Q_1 \\ (i',j') < (i,j)}} c_{i,j;i',j'} b_{i'} b_{j'} \Bigg) = \sum_k C_{i,j;k} b_k. \tag{4-3}$$

We have

$$\begin{aligned} &\sum_{b_i b_j \in Q_2} b_i b_j \otimes (b_j^*)^{\mathrm{op}} (b_i^*)^{\mathrm{op}} \\ &= \sum_{b_i b_j \in Q_2} \sum_{\substack{b_{i'} b_{j'} \in Q_1 \\ (i',j') < (i,j)}} c_{i,j;i',j'} b_{i'} b_{j'} \otimes (b_j^*)^{\mathrm{op}} (b_i^*)^{\mathrm{op}} - \sum_{b_i b_j \in Q_2} \Bigg(\sum_k C_{i,j;k} b_k \Bigg) \otimes (b_j^*)^{\mathrm{op}} (b_i^*)^{\mathrm{op}}. \end{aligned}$$

On the right side of this equation, the first term cancels with the first term

$$\sum_{b_i b_j \in Q_1} b_i b_j \otimes (b_j^*)^{\mathrm{op}} (b_i^*)^{\mathrm{op}}$$

of expression (4-1), by (4-2). Thus, we see that

$$(\mu_2 \otimes \mu_2) \circ \sigma \circ (\mathrm{id} \otimes \delta \otimes \mathrm{id}) \circ \delta(e) = -\sum_{b_i b_j \in Q_2} \left(\sum_k C_{i,j;k} b_k \right) \otimes (b_j^*)^{\mathrm{op}} (b_i^*)^{\mathrm{op}}.$$

Now we consider the terms $(\mu_1 \otimes |\mathrm{id}|) \circ \delta(e)$ and $(\mathrm{id} \otimes \mu_1) \circ \delta(e)$. The first of these is zero, because $\mathcal{B}$ has no differential. The second may be written as

$$(\mathrm{id} \otimes \mu_1) \circ \delta(e) = \sum_k b_k \otimes (\varphi^*(b_k^*))^{\mathrm{op}}.$$

To compute $\varphi^*(b_k^*)$ as an element of I^*, ie a homomorphism from I to $\mathbb{Z}$, use (4-3) above: this homomorphism sends the generator

$$I_{i,j} = b_i b_j - \sum_{\substack{b_{i'} b_{j'} \in Q_1 \\ (i',j') < (i,j)}} c_{i,j;i',j'} b_{i'} b_{j'}$$

of I to the coefficient $C_{i,j;k} \in \mathbb{Z}$.

We want to view $\varphi^*(b_k^*)$ as an element of $\mathcal{B}^!$ of homological degree 2. To do this, following Definition 4.3.3, we pick any extension of $\varphi^*(b_k^*)$ to a functional from $V \otimes_R V$ to $\mathbb{Z}$, or in other words an element of $V^* \otimes_R V^*$, and then consider this element modulo the ideal $I^\perp$. Since $\{I_{i,j} \mid b_i b_j \in Q_2\} \cup Q_1$ is a $\mathbb{Z}$–basis for $V \otimes_R V$, we may extend $\varphi^*(b_k^*)$ to $V \otimes_R V$ by defining it to be zero on any $b_{i'} b_{j'}$ in Q_1.

This extended $\varphi^*(b_k^*)$ sends $b_i b_j \in Q_2$ to $C_{i,j;k}$, since it sends $I_{i,j}$ to $C_{i,j;k}$ and sends every $b_{i'} b_{j'} \in Q_1$ to zero. Thus,

$$\varphi^*(b_k^*) = \sum_{b_i b_j \in Q_2} C_{i,j;k} b_i^* b_j^*.$$

We conclude that

$$(\mathrm{id} \otimes \mu_1) \circ \delta(e) = \sum_k b_k \otimes \sum_{b_i b_j \in Q_2} C_{i,j;k} (b_j^*)^{\mathrm{op}} (b_i^*)^{\mathrm{op}},$$

canceling the remaining term of $(\mu_2 \otimes \mu_2) \circ \sigma \circ (\mathrm{id} \otimes \delta \otimes \mathrm{id}) \circ \delta(e)$. This computation verifies that the type DD structure relations for δ are satisfied. □

We can also reverse the roles of $\mathcal{B}$ and $\mathcal{B}^!$: define $\delta'\colon R \to \mathcal{B}^! \otimes_R (\mathcal{B})^{\mathrm{op}}$ by

$$\delta'(e) := \sum_i b_i^* \otimes (b_i)^{\mathrm{op}},$$

where again the sum is over all multiplicative generators b_i with left idempotent e.

4.5.4 Proposition *The map δ' satisfies the type DD structure relations.*

Proof The proof is similar enough to the proof of Proposition 4.5.3 that we will omit it to save space. □

4.5.5 Definition The rank-one type DD bimodules constructed in Proposition 4.5.3 and Proposition 4.5.4 will be denoted ${}^{\mathcal{B}}K^{(\mathcal{B}^!)^{\mathrm{op}}}$ and ${}^{\mathcal{B}^!}K^{\mathcal{B}^{\mathrm{op}}}$ respectively.

4.5.6 Remark In [6, Section 8], Lipshitz, Ozsváth and Thurston define a notion of Koszul duality in the language of DD bimodules: two algebras $\mathcal{B}$ and $\mathcal{B}'$ are Koszul dual if there exists a rank-one DD bimodule over $\mathcal{B}$ and $\mathcal{B}'$ which is quasi-invertible (and such that the algebra outputs of the DD operation δ lie in the kernel of the augmentation maps on $\mathcal{B}$ and $\mathcal{B}'$; this technical condition is satisfied for all the bimodules we consider). We will not define the notion of quasi-invertibility precisely here; see [6], although they use $\mathbb{Z}/2\mathbb{Z}$ coefficients.

By Proposition 4.5.3, we get a type DD bimodule over H^n and $(H^n)^!$; Proposition 4.5.4 gives us a type DD bimodule over $(H^n)^!$ and H^n. It would be interesting to know whether these bimodules are quasi-invertible; if they were, then $(H^n)^!$ could be regarded as the Koszul dual of H^n in this generalized sense.

However, bordered Floer homology has even stronger duality properties: Theorem 13 of [6] asserts that the bordered surface algebra $\mathcal{A}(\mathcal{Z}, i)$ is Koszul dual to both $\mathcal{A}(\mathcal{Z}, -i)$ and $\mathcal{A}(\mathcal{Z}_*, i)$, where $\mathcal{Z}$ is a pointed matched circle and $\mathcal{Z}_*$ is another pointed matched circle constructed from $\mathcal{Z}$. This situation contrasts with that of H^n, where the quadratic dual algebra is infinitely generated and thus much larger than H^n itself. Below, we will see that Roberts' construction is able to avoid this issue.

5 Khovanov's algebra and Roberts' algebra

In this section we begin to discuss Roberts' bordered theory for Khovanov homology from [11; 12]. Roberts' bordered theory uses a differential bigraded algebra which is denoted $\mathcal{B}\Gamma_n$. This algebra is generated by some right-pointing generators $\overrightarrow{e}$ and left-pointing generators $\overleftarrow{e}$, modulo some explicitly given relations. The differential on $\mathcal{B}\Gamma_n$ is zero on all the right-pointing generators $\overrightarrow{e}$.

We start by defining an algebra $\mathcal{B}_R(H^n)$ using the structure of H^n, with its additive basis $\beta = \{(W(a)b, \sigma)\}$ and set of multiplicative generators $\beta_{\text{mult}} = \{h_\gamma, h_\alpha\}$. In Proposition 5.1.9, we show that $\mathcal{B}_R(H^n)$ is isomorphic to the subalgebra $\mathcal{B}_R\Gamma_n$ of $\mathcal{B}\Gamma_n$ generated by the right-pointing elements.

The subalgebra $\mathcal{B}_R\Gamma_n$ is a linear-quadratic algebra. Its quadratic dual, as defined in Section 4.3, is closely related to the subalgebra $\mathcal{B}_L\Gamma_n$ of $\mathcal{B}\Gamma_n$ generated by the left-pointing elements. More precisely, in Definition 5.2.6 we define a mirroring operation on algebras over $\mathcal{I}_\beta$, and in Proposition 5.2.8, we show that $\mathcal{B}_L\Gamma_n$ is a quotient of $m(\mathcal{B}_R(H^n)^!)$, the mirroring of the quadratic dual of $\mathcal{B}_R(H^n) \cong \mathcal{B}_R\Gamma_n$, by a few explicitly given extra relations.

Finally, in Section 5.3, we take a suitably defined product of $m(\mathcal{B}_R(H^n))^!$ and $\mathcal{B}_R(H_n)$, obtaining an algebra whose quotient by the same extra relations as above is $\mathcal{B}\Gamma_n$.

5.1 Right side of Roberts' algebra

As in Section 3.1, let β denote the $\mathbb{Z}$–basis $\{(W(a)b, \sigma)\}$ of H^n. As at the beginning of Section 3, let $\mathcal{I}_n$ denote the idempotent ring of H^n. The space $\mathrm{Hom}_{\mathcal{I}_n}(H^n, H^n)$ of left $\mathcal{I}_n$–module maps from H^n to itself is a free $\mathbb{Z}$–module. A $\mathbb{Z}$–basis for $\mathrm{Hom}_{\mathcal{I}_n}(H^n, H^n)$ has generators $e(h_1, h_2)$ for each pair $h_1 \in \beta, h_2 \in \beta$ such that h_1 and h_2 have the same left idempotent. Here, $e(h_1, h_2)$ is the homomorphism that sends h_1 to h_2 and sends all other basis elements in β to zero.

Note that $\mathrm{Hom}_{\mathcal{I}_n}(H^n, H^n)$ has the structure of a ring, with multiplication given by composition. We will define a grading on $\mathrm{Hom}_{\mathcal{I}_n}(H^n, H^n)$ which differs from the usual one by a factor of $-\frac{1}{2}$.

5.1.1 Definition Let $e(h_1, h_2)$ be a generator of $\mathrm{Hom}_{\mathcal{I}_n}(H^n, H^n)$. The *degree* of $e(h_1, h_2)$ is defined to be $\frac{1}{2}(\deg h_1 - \deg h_2)$.

5.1.2 Remark This choice of grading has the advantage that it agrees with Roberts' choice, but it can also be justified on its own grounds. The factor of -1 comes from the fact that Khovanov, in [4], replaces the usual q–grading by its negative, to make H^n positively rather than negatively graded. We will see below (in the proof of Proposition 6.4.1 in particular) why the factor of $\frac{1}{2}$ is reasonable. Note that while this grading is now a $\frac{1}{2}\mathbb{Z}$–grading rather than a $\mathbb{Z}$–grading, it will always function as an intrinsic grading rather than a homological grading. Thus, it will have no effect on signs and we are free to use a $\frac{1}{2}\mathbb{Z}$–grading if desired.

The elements $e(h, h) \in \mathrm{Hom}_{\mathcal{I}_n}(H^n, H^n)$, for $h \in \beta$, generate a subring which is isomorphic to a direct product of copies of $\mathbb{Z}$. We will denote this subring by $\mathcal{I}_\beta$; note that $\mathcal{I}_\beta$ is isomorphic to the idempotent ring of $\mathcal{B}\Gamma_n$, and hence of $\mathcal{B}_R\Gamma_n$ as well.

5.1.3 Definition Let $\mathcal{B}_R(H^n)$ denote the smallest subring of $\mathrm{Hom}_{\mathcal{I}_n}(H^n, H^n)$ containing $\mathcal{I}_\beta$ and containing every $e(h_1, h_2)$ such that h_2 occurs as a nonzero term in the β–expansion of $h_1 \cdot h$, for some h in the set of multiplicative generators β_{mult}. We may view $\mathcal{B}_R(H^n)$ as an algebra over $\mathcal{I}_\beta$.

The algebra $\mathcal{B}_R(H^n)$ inherits an intrinsic grading from the grading on $\mathrm{Hom}_{\mathcal{I}_n}(H^n, H^n)$ defined in Definition 5.1.1. The degree-0 summand of $\mathcal{B}_R(H^n)$ is its idempotent ring $\mathcal{I}_\beta$. The multiplicative generators $e(h_1, h_2)$ such that h_2 occurs as a nonzero term in the basis expansion of $h_1 \cdot h_\gamma$, for some γ, have degree $-\frac{1}{2}$, since h_γ has intrinsic degree 1. Those such that h_2 occurs as a nonzero term in the expansion of some $h_1 \cdot h_\alpha$ have degree -1, since h_α has degree 2.

A natural set of multiplicative generators for $\mathcal{B}_R(H^n)$ as an algebra over $\mathcal{I}_\beta$ is given in its definition, namely the elements $e(h_1, h_2)$ such that h_2 occurs as a nonzero term in the β–expansion of $h_1 \cdot h$, for some h in the set of multiplicative generators β_{mult}. If $h = h_\gamma$, the corresponding element of $\mathcal{B}_R(H^n)$ will be denoted $b_{\gamma;h_1,h_2}$. If $h = h_\alpha$, the corresponding element of $\mathcal{B}_R(H^n)$ will be denoted $b_{C;h_1,h_2}$, where if $h_1 = (W(a)b, \sigma)$, then C is the circle in $W(a)b$ containing α. Note that for a fixed h_1, all arcs α' which lie on the same circle C as α in $W(a)b$ yield the same generator $b_{C;h_1,h_2}$ of $\mathcal{B}_R(H^n)$.

5.1.4 Remark We use notation with subscripts, such as $b_{\gamma;h_1,h_2}$ or $b_{C;h_1,h_2}$, to refer to elements of $\mathcal{B}_R(H^n)$. We also use b, without any subscripts, to refer to a crossingless matching. Below, if γ is a bridge on b, we will let $b(\gamma)$ denote the crossingless matching obtained by surgery on b along γ.

There are no linear relations among the generators $b_{\gamma;h_1,h_2}$ and $b_{C;h_1,h_2}$. The generators are homogeneous with intrinsic degree $-\frac{1}{2}$ or -1, so they are in the kernel of the augmentation map on $\mathcal{B}_R(H^n)$ (which is the projection onto the degree-0 summand). The left idempotent of each generator of the form $b_{\gamma;h_1,h_2}$ and $b_{C;h_1,h_2}$ is $e(h_1, h_1)$; the right idempotent is $e(h_2, h_2)$. For compactness of notation, we will identify each elementary idempotent $e(h, h) \in \mathcal{I}_\beta$ with the corresponding element $h \in \beta$. Thus, we say that the left idempotent of $b_{\gamma;h_1,h_2}$ and $b_{C;h_1,h_2}$ is h_1 and the right idempotent is h_2.

We will show that $\mathcal{B}_R(H^n)$, with the set of generators $b_{\gamma;h_1,h_2}$ and $b_{C;h_1,h_2}$, is a linear-quadratic algebra. The proof will closely follow that of Theorem 4.2.1.

Let V be the free $\mathbb{Z}$–module spanned by the generators of $\mathcal{B}_R(H^n)$; as discussed above, the idempotent ring $\mathcal{I}_\beta$ has left and right actions on V. We may write $\mathcal{B}_R(H^n) = T(V)/J$ for some ideal J of $T(V)$. Let $J_2 := J \cap (T^1(V) \oplus T^2(V))$. We identify a set of generators for J_2.

5.1.5 Proposition *The ideal J_2 of linear-quadratic relations of $\mathcal{B}_R(H^n)$ is generated by the following relations:*

(1) *Suppose γ and η are two bridges which can be drawn without intersection on the same crossingless matching b; let η' denote the bridge on $b(\gamma)$ corresponding to η, where $b(\gamma)$ is the crossingless matching resulting from surgery on γ. Define γ' similarly. For any choice of (a,σ), let $h_1 = (W(a)b,\sigma) \in \beta$. If the generators $b_{\gamma;h_1,h_2}$ and $b_{\eta';h_2,h_3}$ exist in $\mathcal{B}_R(H^n)$ for some h_2, h_3 in β, we get a relation*

$$b_{\gamma;h_1,h_2} b_{\eta';h_2,h_3} - b_{\eta;h_1,\tilde{h}_2} b_{\gamma';\tilde{h}_2,h_3} \in J_2,$$

where $\tilde{h}_2$ is any element of β such that $b_{\eta;h_1,\tilde{h}_2}$ and $b_{\gamma';\tilde{h}_2,h_3}$ exist in $\mathcal{B}_R(H^n)$.

(2) *Suppose γ is a bridge on a crossingless matching b. For any choice of (a,σ), let $h_1 = (W(a)b,\sigma) \in \beta$. Let C be any circle in $W(a)b$. Let C' be any circle in $W(a)(b(\gamma))$ which corresponds to C under surgery on γ; if the endpoints of γ do not both lie on C, then C' is unique, and otherwise there are two choices for C'. If the generators $b_{\gamma;h_1,h_2}$ and $b_{C';h_2,h_3}$ exist in $\mathcal{B}_R(H^n)$ for some h_2, h_3 in β, we get a relation*

$$b_{\gamma;h_1,h_2} b_{C';h_2,h_3} - b_{C;h_1,\tilde{h}_2} b_{\gamma;\tilde{h}_2,h_3} \in J_2,$$

where $\tilde{h}_2$ is any element of β such that $b_{C;h_1,\tilde{h}_2}$ and $b_{\gamma;\tilde{h}_2,h_3}$ exist; note that $\tilde{h}_2$ is uniquely determined by C and h_1.

(3) *For any choice of (a,b,σ), let $h_1 = (W(a)b,\sigma) \in \beta$. Let C_1 and C_2 be two circles in $W(a)b$. If the generators $b_{C_1;h_1,h_2}$ and b_{C_2,h_2,h_3} exist in $\mathcal{B}_R(H^n)$ for some h_2, h_3 in β, we get a relation*

$$b_{C_1;h_1,h_2} b_{C_2;h_2,h_3} - b_{C_2;h_1;\tilde{h}_2} b_{C_1;\tilde{h}_2,h_3} \in J_2,$$

where $\tilde{h}_2$ is any element of β such that $b_{C_2;h_1;\tilde{h}_2}$ and $b_{C_1;\tilde{h}_2,h_3}$ exist. As above, $\tilde{h}_2$ is uniquely determined by C_2 and h_1.

(4) *Finally, suppose γ is any bridge on a crossingless matching b. Recall from Section 4.2 or Roberts* [12, Definition 10] *that γ has a dual bridge $\gamma^\dagger$. For any choice of (a,σ), let $h_1 = (W(a)b,\sigma) \in \beta$. If the generators $b_{\gamma;h_1,h_2}$ and $b_{\gamma^\dagger;h_2,h_3}$ exist in $\mathcal{B}_R(H^n)$ for some h_2, h_3 in β, then h_3 differs from h_1 by switching the sign of one circle of $W(a)b$ from plus to minus. Let C denote this circle. We get a relation*

$$b_{\gamma;h_1,h_2} b_{\gamma^\dagger;h_2,h_3} - b_{C;h_1,h_3} \in J_2.$$

Proof Since $\mathcal{B}_R(H^n)$ is an intrinsically graded algebra, if we have a relation in J_2, then each of its grading-homogeneous parts must also be in J_2. Thus, we may analyze

J_2 one degree at a time. Since the generators of $\mathcal{B}_R(H^n)$ have intrinsic degree $-\frac{1}{2}$ or -1, and we are trying to identify the linear-quadratic relations among them, we may assume these relations have intrinsic degree -1, $-\frac{3}{2}$ or -2. The case of intrinsic degree $-\frac{1}{2}$ is excluded since any such relation would be a linear dependency among the generators of $\mathcal{B}_R(H^n)$.

The relations of intrinsic degree -1 may be sums of quadratic monomials in the degree $-\frac{1}{2}$ generators $b_{\gamma;h_1,h_2}$ of $\mathcal{B}_R(H^n)$ and linear monomials in the degree -1 generators $b_{C;h_1,h_2}$. Analyzing the possible cases, we get the relations of items (1) and (4) above.

The relations of intrinsic degree $-\frac{3}{2}$ are sums of quadratic monomials, each involving one degree $-\frac{1}{2}$ generator $b_{\gamma;h_1,h_2}$ and one degree -1 generator $b_{C;h_1,h_2}$. These relations are generated by the relations of item (2) above.

Finally, the relations of degree -2 are sums of quadratic monomials in the degree -2 generators $b_{C;h_1,h_2}$. They are generated by the relations of item (3) above. □

5.1.6 Remark As in Section 4.2, Proposition 5.1.5 is not actually needed to prove Proposition 5.1.7. We could instead introduce $\bar{J}_2$, generated by the relations in Proposition 5.1.5, and show that $J \subset T(V) \cdot \bar{J}_2 \cdot T(V)$.

5.1.7 Proposition *With J and J_2 defined as above, we have*

$$J = T(V) \cdot J_2 \cdot T(V).$$

Thus, $\mathcal{B}_R(H^n)$ is a linear-quadratic algebra.

Proof We want to show that $J \subset T(V) \cdot J_2 \cdot T(V)$. As in Theorem 4.2.1 above, it suffices to show that for a general element r of J, one may successively add to r elements of the ideal generated by the relation elements listed in items (1)–(4) of Proposition 5.1.5, until one obtains zero.

Let r be an arbitrary element of J. We may assume without loss of generality that r has a unique left idempotent and right idempotent. Since J is an ideal of the tensor algebra $T(V)$, r may be written as a linear combination of monomials in the generators $b_{\gamma;h_1,h_2}$ and $b_{C;h_1,h_2}$. Let

$$r = \sum_i n_i (b_{i,1} \cdots b_{i,l_i}),$$

where $n_i \in \mathbb{Z}$ and each $b_{i,j}$ is one of the generators $b_{\gamma;h_1,h_2}$ or $b_{C;h_1,h_2}$.

Consider one of the monomial summands $m_i = b_{i,1} \cdots b_{i,l_i}$ of r. After adding elements of $T(V) \cdot J_2 \cdot T(V)$ to this monomial, we may assume that all the $b_{\gamma;h_1,h_2}$ generators among the $b_{i,j}$ come before (ie with lower j than) the $b_{C;h_1,h_2}$ generators. The necessary relations come from item (2) of Proposition 5.1.5. Let m'_i denote the monomial obtained by modifying m_i in this way.

Write m'_i as $m_{\gamma,i} \cdot m_{C,i}$, where $m_{\gamma,i}$ is a product of $b_{\gamma;h_1,h_2}$ generators and $m_{C,i}$ is a product of $b_{C;h_1,h_2}$ generators. Let $h \in \beta$ be the left idempotent of $m_{\gamma,i}$ and let $h'_i \in \beta$ be the right idempotent of $m_{\gamma,i}$. Note that h does not depend on i, since h is the left idempotent of our original relation term r.

Viewing h and h'_i as elements of H^n, let $e \in \mathcal{I}_n$ denote the right idempotent of h. Let $e' \in \mathcal{I}_n$ denote the right idempotent of h'_i, which does not depend on i since the monomial $m_{C,i}$ is a product of b_C generators. As in Theorem 4.2.1, e and e' are vertices of G_n, the undirected Hasse diagram of NC_n. To the monomial $m_{\gamma,i}$, we can associate a path $p(m_{\gamma,i})$ from e to e' in G_n.

We claim that we may further modify m'_i such that $p(m_{\gamma,i})$ is a minimal-length path between e and e' as vertices of G_n. Indeed, suppose $m_{\gamma,i}$ corresponds to a path of nonminimal length between e and e'. Write $m_{\gamma,i} = b_{\gamma_1;h_1,h_2} \cdots b_{\gamma_k;h_k,h_{k+1}}$. Then there exists a minimal index $2 \le j \le k$ such that $b_{\gamma_1;h_1,h_2} \cdots b_{\gamma_{j-1};h_{j-1},h_j}$ corresponds to a path ϕ of minimal length in G_n but $b_{\gamma_1;h_1,h_2} \cdots b_{\gamma_j;h_j,h_{j+1}}$ does not.

Let $e_R(h_j) \in \mathcal{I}_n$ denote the right idempotent of h_j. Then $e_R(h_j)$ is a vertex of G_n and the distance in G_n between e and $e_R(h_j)$ is $j-1$. However, the distance between e and $e_R(h_{j+1})$ is $j-2$ rather than j; the argument is the same as in the proof of Theorem 4.2.1. Thus, there exists a path $\tilde{\psi}$ in G_n, of length $j-2$, from e to $e_R(h_{j+1})$. Appending $e_R(h_j)$ to the end of the path $\tilde{\psi}$, we get a path ψ in G_n, of length $j-1$, between e and $e_R(h_j)$. By assumption, ψ is a minimal-length path.

We now have two minimal-length paths ϕ and ψ between e and $e_R(h_j)$. The path ϕ corresponds to the monomial $b_{\gamma_1;h_1,h_2} \cdots b_{\gamma_{j-1};h_{j-1},h_j}$. The path ψ corresponds to a monomial $b_{\gamma'_1;h_1,h'_2} \cdots b_{\gamma'_{j-2};h'_{j-2},h'_{j-1}} \cdot b_{\gamma_j^\dagger;h'_{j-1},h_j}$, and we have $e_R(h'_{j-1}) = e_R(h_{j+1})$. By Lemma 4.2.4, we may modify ϕ one vertex at a time to obtain ψ. Such modifications can be mirrored on the level of monomials by adding relation terms obtained from item (1) of Proposition 5.1.5. Thus, we may modify $m_{\gamma,i}$, which equals

$$b_{\gamma_1;h_1,h_2} \cdots b_{\gamma_{j-1};h_{j-1},h_j} \cdot b_{\gamma_j;h_j,h_{j+1}} \cdots b_{\gamma_k;h_k,h_{k+1}},$$

by adding terms in $T(V) \cdot J_2 \cdot T(V)$ to obtain

$$b_{\gamma'_1;h_1,h'_2} \cdots b_{\gamma'_{j-2};h'_{j-2},h'_{j-1}} \cdot b_{\gamma_j^\dagger;h'_{j-1},h_j} \cdot b_{\gamma_j;h_j,h_{j+1}} \cdots b_{\gamma_k;h_k,h_{k+1}}.$$

Inside this monomial is $b_{\gamma_j^\dagger;h'_{j-1},h_j} \cdot b_{\gamma_j;h_j,h_{j+1}}$, which may be replaced with a $b_{C;h'_{j-1},h_{j+1}}$ term using the relation terms in item (4) of Proposition 5.1.5. As before, this b_C term may be commuted to the right side of m'_i.

After this modification, we have strictly reduced the length of $m_{\gamma,i}$ in the factorization of m_i' as $m_{\gamma,i} \cdot m_{C,i}$. If the new $m_{\gamma,i}$ still does not represent a minimal-length path in G_n, we can repeat the same procedure and eventually it will terminate.

At this point, we have shown that we can modify our original $r = \sum_i n_i(m_i)$ by adding terms in $T(V) \cdot J_2 \cdot T(V)$, until each m_i is a monomial factorizable as $m_{\gamma,i} \cdot m_{C,i}$ with $m_{\gamma,i}$ representing a minimal-length path in G_n. The starting and ending vertices of all these paths are the same. Thus, by Lemma 4.2.4 and the relations from item (1) of Proposition 5.1.5, we may do further modifications until all of the $m_{\gamma,i}$ are the same monomial m_γ and we have

$$r = m_\gamma \sum_i n_i(m_{C,i}) \quad \text{modulo } T(V) \cdot J_2 \cdot T(V).$$

Since r was assumed to have unique left and right idempotents in $\mathcal{I}_\beta$, the set of circles C involved in each term $m_{C,i}$ of the above expression must be the same. Thus, using relations from item (3) of Proposition 5.1.5, we may rewrite each $m_{C,i}$ as the same monomial m_C. Then

$$r = N \cdot m_\gamma m_C \quad \text{modulo } T(V) \cdot J_2 \cdot T(V),$$

where $N = \sum n_i$.

Finally, we use the fact that $r \in J$, or in other words that $r = 0$ as an element of $\mathcal{B}_R(H^n)$. This condition implies that $N \cdot m_\gamma m_C$ must also be in J, since it differs from r by an element of $T(V) \cdot J_2 \cdot T(V)$ which is contained in J.

Note that $\mathcal{B}_R(H^n)$ is a subring of $\mathrm{Hom}_{\mathcal{I}_n}(H^n, H^n)$; the element $m_\gamma m_C$ may be identified with the left R–linear map from H^n to H^n which sends e to e' and sends all other elements of β to zero, where e and e' here are the left and right idempotents of $m_\gamma m_C$. If $N \cdot m_\gamma m_C$ is zero in $\mathcal{B}_R(H^n)$, then it is zero in $\mathrm{Hom}_{\mathcal{I}_n}(H^n, H^n)$, implying that N must be zero.

In other words, starting with $r \in J$ above, we have shown that $r = 0$ modulo $T(V) \cdot J_2 \cdot T(V)$. Hence $J \subset T(V) \cdot J_2 \cdot T(V)$, so $\mathcal{B}_R(H^n)$ is a linear-quadratic algebra. □

Now we can see that $\mathcal{B}_R(H^n)$ is isomorphic to $\mathcal{B}_R\Gamma_n$. First, we define the latter algebra more precisely.

5.1.8 Definition Let $\mathcal{B}\Gamma_n$ be Roberts' algebra from [11; 12]. Let $\mathcal{B}_R\Gamma_n$ be the subalgebra of $\mathcal{B}\Gamma_n$ spanned over $\mathcal{I}_\beta$ by those generators $\overrightarrow{e}$ with right pointing arrows ($\mathcal{B}\Gamma_n$ also has some generators $\overleftarrow{e}$ with left pointing arrows). The subalgebra $\mathcal{B}_R\Gamma_n$

has no differential. It inherits a bigrading from $\mathcal{B}\Gamma_n$ (see [12, Definition 19]); the homological grading is identically zero on $\mathcal{B}_R\Gamma_n$.

5.1.9 Proposition *$\mathcal{B}_R(H^n) \cong \mathcal{B}_R\Gamma_n$ as bigraded algebras over $\mathcal{I}_\beta$.*

Proof An examination of the subset of Roberts' algebra relations in [12] which involve only right-pointing generators shows that they correspond with the relations listed in Proposition 5.1.5 under the (bigrading-preserving) identification of the generators b of $\mathcal{B}_R(H^n)$ with the generators $\overrightarrow{e}$ of $\mathcal{B}_R\Gamma_n$. Thus, this proposition follows from Proposition 5.1.7. □

5.2 Left side of Roberts' algebra

5.2.1 Definition Let $\mathcal{B}_L\Gamma_n$ be the subalgebra of $\mathcal{B}\Gamma_n$ spanned over $\mathcal{I}_\beta$ by those generators $\overleftarrow{e}$ with left pointing arrows. The bigrading and differential on $\mathcal{B}\Gamma_n$ give us a bigrading and differential on $\mathcal{B}_L\Gamma_n$.

We will see that $\mathcal{B}_L\Gamma_n$ may be identified, after a mirroring operation defined in Definition 5.2.6, with a quotient of the quadratic dual $(\mathcal{B}_R(H^n))^!$ of $\mathcal{B}_R(H^n)$ by a few explicitly given extra relations.

First, we analyze the dual algebra $(\mathcal{B}_R(H^n))^!$. As an algebra, it is the quadratic dual of $(\mathcal{B}_R(H^n))^{(0)}$. We may write $\mathcal{B}_R(H^n)$ as $T(V)/J$, where if

$$J_2 := J \cap (T^1(V) \oplus T^2(V))$$

then we have

$$J = T(V) \cdot J_2 \cdot T(V).$$

Let I denote the image of J_2 under the projection map $T^1(V) \oplus T^2(V) \to T^2(V)$ onto the second summand. Then

$$(\mathcal{B}_R(H^n))^{(0)} \cong T(V)/I,$$
$$(\mathcal{B}_R(H^n))^! \cong T(V^*)/I^\perp.$$

The ideal J_2 is generated explicitly by the relations listed in Proposition 5.1.5. We may discard the linear parts of these relations, and keep the quadratic parts, to get a set of generators for I. These generators have a simple form: if r is a generating relation in I, then r is either a single quadratic monomial or a difference of two quadratic monomials.

Define a graph G whose vertices are all quadratic monomials appearing with nonzero coefficient in some relation $r \in I$. Two monomials v and $\tilde{v}$ are connected by an edge in G if $v - \tilde{v}$ is in I.

5.2.2 Proposition *The graph G is a disjoint union of isolated points, line segments (two points connected by an edge and disconnected from the rest of G), triangles (three points, all connected, and disconnected from the rest of G), and tetrahedra (four points, all connected, and disconnected from the rest of G).*

Proof Define a graph G' with the same vertices as G; two monomials v and $\tilde{v}$ are connected by an edge in G' if $v-\tilde{v}$ is the quadratic part of one of the explicit relations (1)–(4) listed in Proposition 5.1.5. We will determine the structure of G' by looking at the quadratic parts of the relations (1)–(4); G may be obtained from G' by replacing each connected component of G' with a complete graph on the same number of vertices.

First, the isolated points in G' are quadratic monomials of the form

$$b_{\gamma;h_1,h_2} b_{\gamma^\dagger,h_2,h_3};$$

these are the quadratic parts of relations from item (4) of Proposition 5.1.5.

Next we look at relations from item (1) of Proposition 5.1.5, which have no linear parts and are already quadratic (the same applies to items (2) and (3); only the relations from item (4) have linear parts). Some of the line segments in G' come from relation terms

$$b_{\gamma;h_1,h_2} b_{\eta';h_2,h_3} - b_{\eta;h_1,\tilde{h}_2} b_{\gamma';\tilde{h}_2,h_3},$$

where γ and η are two bridges which can be drawn on the same crossingless matching without intersection, such that $\eta \in B_d(L,\gamma)$ in the notation of [12, Proposition 11]. Roberts' L corresponds to our $W(a)b$.

Other line segments in G' come from the same relations when $\eta \in B_o(L,\gamma)$, in every case except when γ splits a plus-labeled circle and η' joins the two newly formed circles into a new minus-labeled circle. For notations like $\mathcal{B}_o(L,\gamma)$ and $\mathcal{B}_d(L,\gamma)$, see Roberts [12, Proposition 11].

Line segments in G' also come from relations $b_{\gamma;h_1,h_2} b_{C;h_2,h_3} - b_{C;h_1,\tilde{h}_2} b_{\gamma;\tilde{h}_2,h_3}$ of item (2) of Proposition 5.1.5 when the circle C is disjoint from the support of γ and from relations $b_{C_1;h_1,h_2} b_{C_2;h_2,h_3} - b_{C_2;h_1;\tilde{h}_2} b_{C_1;\tilde{h}_2,h_3}$ of item (3) of Proposition 5.1.5, where C_1 and C_2 are two distinct circles labeled $+$ in h_1.

The remaining relations from item (2) of Proposition 5.1.5 give configurations of three vertices in G' connected by two edges. We get triangles in G which connect triples

$$\{b_{\gamma;h_1,h_2} b_{C;h_2,h_3}, b_{C_1;h_1,\tilde{h}_2} b_{\gamma;\tilde{h}_2,h_3}, b_{C_2;h_1,\tilde{\tilde{h}}_2} b_{\gamma;\tilde{\tilde{h}}_2,h_3}\},$$
$$\{b_{C;h_1,h_2} b_{\gamma;h_2,h_3}, b_{\gamma;h_1,\tilde{h}_2} b_{C_1;\tilde{h}_2,h_3}, b_{\gamma;h_1,\tilde{\tilde{h}}_2} b_{C_2;\tilde{\tilde{h}}_2,h_3}\}$$

when the circle C is not disjoint from the support of γ. The rest of the triangles in G connect triples

$$\{b_{\gamma_1;h_1,h_2}b_{*;h_2,h_3}, b_{\gamma_2;h_1,\tilde{h}_2}b_{*;\tilde{h}_2,h_3}, b_{\gamma_3;h_1,\tilde{\tilde{h}}_2}b_{*;\tilde{\tilde{h}}_2,h_3}\}$$

when $\gamma_i \in B_s(L, \gamma_j)$ for $1 \le i, j \le 3$, using relations from item (1) of Proposition 5.1.5.

We have accounted for the quadratic parts of all relations from items (2)–(4) of Proposition 5.1.5, as well as most of the relations from item (1). The remaining relations from item (1) give rise to squares of four vertices in G' and thus to tetrahedra in G. These four-vertex components exist whenever we have two bridges γ and η, with $\eta \in B_o(L, \gamma)$, such that γ splits a plus-labeled circle and η' joins the newly formed circles into a minus-labeled circle. In such cases, we have four quadratic monomials which are all equal modulo the relation terms in I (and thus are connected in G). These can be written as $b_{\gamma;h_1,h_2}b_{\eta';h_2,h_3}$, $b_{\gamma;h_1,\tilde{h}_2}b_{\eta';\tilde{h}_2,h_3}$, $b_{\eta;h_1,\tilde{\tilde{h}}_2}b_{\gamma';\tilde{\tilde{h}}_2,h_3}$ and $b_{\eta;h_1,\tilde{\tilde{\tilde{h}}}_2}b_{\gamma';\tilde{\tilde{\tilde{h}}}_2,h_3}$. □

Order the set of generators $b_{\gamma;h_1,h_2}$ and $b_{C;h_1,h_2}$ of $\mathcal{B}_R(H^n)$ such that the $b_{\gamma;h_1,h_2}$ come before the $b_{C;h_1,h_2}$ in the ordering. Using G, the generators of I may be summarized as follows: for every connected component of G, there exists a minimal vertex v. For all other vertices $\tilde{v}$ in the same component of v, there exists a relation $\tilde{v}-v$ in I, and if v is a singleton, then v is also a relation in I. These relations are a set of generators for I as in Remark 4.1.3.

We may use the reasoning of Remark 4.3.2 to identify a set of generators for $I^{\perp}$. For any quadratic monomial in the generators b which does not appear as a vertex of G, the corresponding monomial in the generators b^* is an element of I^*. Isolated points of G do not give generators of I^*. For every line segment in G between vertices v and $\tilde{v}$, let v^* and $(\tilde{v})^*$ denote the corresponding monomials in the generators b^*. Then $v^* + (\tilde{v})^*$ is an element of I^*. For every triangle in G with a minimal vertex v and two nonminimal vertices $\tilde{v}$ and $\tilde{\tilde{v}}$, let v^*, $(\tilde{v})^*$ and $(\tilde{\tilde{v}})^*$ denote the corresponding monomials in the generators b^*. Then

$$v^* + (\tilde{v})^* + (\tilde{\tilde{v}})^*$$

is an element of $I^{\perp}$. Finally, for every tetrahedron in G with a minimal vertex v and three nonminimal vertices $\tilde{v}$, $\tilde{\tilde{v}}$ and $\tilde{\tilde{\tilde{v}}}$, let v^*, $(\tilde{v})^*$, $(\tilde{\tilde{v}})^*$ and $(\tilde{\tilde{\tilde{v}}})^*$ denote the corresponding monomials in the generators b^*. Then

$$v^* + (\tilde{v})^* + (\tilde{\tilde{v}})^* + (\tilde{\tilde{\tilde{v}}})^*$$

is an element of $I^{\perp}$. The above-listed elements generate $I^{\perp}$.

We may also compute the action of the map φ on the generators of I, using the relations from item (4) of Proposition 5.1.5. For every generator of I of the form $b_{\gamma;h_1,h_2}b_{\gamma^\dagger;h_2,h_3} - b_{C;h_1,h_3}$, we have

$$\varphi(b_{\gamma;h_1,h_2}b_{\gamma^\dagger;h_2,h_3}) = -b_{C;h_1,h_3}.$$

The map φ sends all other generators of I to zero. Thus, dualizing φ, we have

$$\varphi^*(b^*_{C;h_1,h_3}) = -\sum_i b^*_{\gamma_i;h_1,h_{2,i}} b^*_{\gamma_i^\dagger;h_{2,i},h_3},$$

where the sum runs over all bridges γ_i on the right crossingless matching of h_1 which have an endpoint on C, as well as all compatible $h_{2,i}$.

Finally, $(\mathcal{B}_R(H^n))^!$ is bigraded; the generators $b^*_{\gamma;h_1,h_2}$ have degree $\left(\frac{1}{2},1\right)$ since $b_{\gamma;h_1,h_2}$ has degree $\left(-\frac{1}{2},0\right)$, and the generators $b^*_{C;h_1,h_2}$ have degree $(1,1)$ since $b_{C;h_1,h_2}$ has degree $(-1,0)$. Here, the first index denotes the intrinsic degree, and the second index denotes the homological degree (this is the reverse of Roberts' convention). The generators of $\mathcal{B}_R(H^n)$ are all placed in homological degree 0.

5.2.3 Remark While the quadratic dual of an algebra which is finitely generated over $\mathbb{Z}$ (like $\mathcal{B}_R\Gamma_n$) may in general be infinitely generated over $\mathbb{Z}$, the algebra $(\mathcal{B}_R(H^n))^!$ is finitely generated over $\mathbb{Z}$. In fact, the relations on the algebra are irrelevant for this property: $T(V^*)$ is already finitely generated over $\mathbb{Z}$, since the structure of the idempotents only allows monomials of a certain length in the generators of V^* to be nonzero. The same is true for $T(V)$.

We can now relate $(\mathcal{B}_R(H^n))^!$ with $\mathcal{B}_L\Gamma_n$. To do this, we need to define a mirroring operation for modules and bimodules over the idempotent ring $\mathcal{I}_\beta$ of $(\mathcal{B}_R(H^n))^!$ and $\mathcal{B}_L\Gamma_n$.

5.2.4 Definition Let X be any module or bimodule over the idempotent ring $\mathcal{I}_\beta$. The *mirror* of X, denoted $m(X)$, is the module or bimodule whose actions of $\mathcal{I}_\beta$ are the actions on X, precomposed with the map from $\mathcal{I}_\beta$ to $\mathcal{I}_\beta$ which mirrors each elementary idempotent across the line $\{0\}\times\mathbb{R}$. Note that $m(m(X)) = X$.

5.2.5 Example Suppose X is a left module over $\mathcal{I}_\beta$. Let $x \in X$ and let $m(x)$ denote the corresponding element of $m(X)$. Let $h = (W(a)b,\sigma) \in \mathcal{I}_\beta$; then

$$h \cdot m(x) := m(m(h)\cdot x),$$

where $m(h)$ is $(W(b)a, m(\sigma))$ and $m(\sigma)$ is the same labeling of circles as σ, mirrored across $\{0\}\times\mathbb{R}$. See Figure 6 for an illustration.

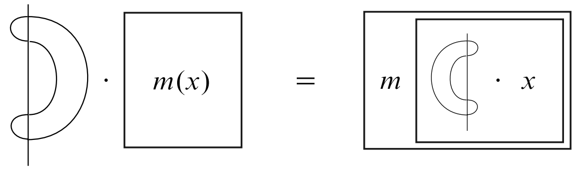

Figure 6: The mirror of a left module over $\mathcal{I}_\beta$

We will have definitions related to Definition 5.2.4 for $\mathcal{I}_\beta$–modules and bimodules with more structure. Here, we are concerned with algebras.

5.2.6 Definition Let $\mathcal{B}$ be a differential bigraded algebra over the idempotent ring $\mathcal{I}_\beta$. The *mirror* of $\mathcal{B}$, denoted $m(\mathcal{B})$, is the same differential bigraded ring as $\mathcal{B}$. As an algebra, the left and right actions of $\mathcal{I}_\beta$ are mirrored as in Definition 5.2.4. The map from $\mathcal{B}$ to $m(\mathcal{B})$ sending $b \in \mathcal{B}$ to $m(b) \in m(\mathcal{B})$ is an isomorphism of rings (but not of algebras); its inverse is the analogously defined map from $m(\mathcal{B})$ to $m(m(\mathcal{B})) = \mathcal{B}$. To avoid confusion with other uses of the letter m, we will refer to both of these mirroring maps as mirr: we have

$$\mathrm{mirr}\colon \mathcal{B} \to m(\mathcal{B}) \quad \text{and} \quad \mathrm{mirr}\colon m(\mathcal{B}) \to m(m(\mathcal{B})) = \mathcal{B}.$$

5.2.7 Remark The mirroring operation for algebras commutes with quadratic duality: if $\mathcal{B}$ is a linear-quadratic algebra over $\mathcal{I}_\beta$, then

$$m(\mathcal{B}^!) = (m(\mathcal{B}))^!.$$

Thus, we can write either of these algebras as $m(\mathcal{B})^!$. Mirroring also commutes with taking the opposite algebra: we have $m(\mathcal{B}^{\mathrm{op}}) = (m(\mathcal{B}))^{\mathrm{op}}$, so we can write either of these algebras as $m(\mathcal{B})^{\mathrm{op}}$.

5.2.8 Proposition *$\mathcal{B}_L\Gamma_n$ is isomorphic to the quotient of $m(\mathcal{B}_R(H^n))^!$ by the following extra relations. Let the graph G be defined as above; for each tetrahedron in G, the only relation in $m(\mathcal{B}_R(H^n))^!$ involving the vertices of the tetrahedron is that the sum of all its vertices is zero. The algebra $\mathcal{B}_L\Gamma_n$ imposes extra relations among the vertices of each tetrahedron. Recall that tetrahedra in G arise when we have two bridges γ and η, with $\eta \in B_o(L,\gamma)$, such that γ splits a plus-labeled circle and η' joins the newly formed circles into a minus-labeled circle. The vertices of the corresponding tetrahedron are, following the discussion above:*

- $a := m(b^*_{\gamma;m(h_1),m(h_2)})m(b^*_{\eta';m(h_2),m(h_3)})$,
- $b := m(b^*_{\gamma;m(h_1),m(\tilde{h}_2)})m(b^*_{\eta';m(\tilde{h}_2),m(h_3)})$,
- $c := m(b^*_{\eta;m(h_1),m(\tilde{\tilde{h}}_2)})m(b^*_{\gamma';m(\tilde{\tilde{h}}_2),m(h_3)})$,
- $d := m(b^*_{\eta;m(h_1),m(\tilde{\tilde{\tilde{h}}}_2))})m(b^*_{\gamma';m(\tilde{\tilde{\tilde{h}}}_2),m(h_3)})$.

Whereas the algebra $m(\mathcal{B}_R(H^n))^!$ *imposes only the relation* $a+b+c+d=0$*, the algebra* $\mathcal{B}_L\Gamma_n$ *imposes the relations*

- $a+c=0$,
- $a+d=0$,
- $b+c=0$,
- $b+d=0$ *(this relation also follows from the previous three).*

From these relations, $a+b+c+d=0$ *may be deduced, as well as relations for the two remaining edges of the tetrahedron:*

- $a-b=0$,
- $c-d=0$.

Proof Consider the map from $m(\mathcal{B}_R(H^n))^!$ to $\mathcal{B}_L\Gamma_n$ sending $m(b^*_{\gamma;m(h_1),m(h_2)})$ to $\overleftarrow{e}_{\gamma;h_1,h_2}$ and sending $m(b^*_{C;m(h_1),m(h_2)})$ to $\overleftarrow{e}_{C;h_1,h_2}$. By examining the subset of Roberts' relations from [12] involving only left-pointing generators, and comparing with the relations for $m(\mathcal{B}_R(H^n))^!$ above, we see that this is a well-defined surjective bigrading-preserving map whose kernel is generated by the extra relations listed in the statement of this proposition. These extra anticommutation relations can be found in Roberts' algebra as a subset of the relations (21), case (2) [12, page 98].

After mirroring, the formula above for φ^* agrees with Roberts' definition [12, Proposition 25], of the differential on $\mathcal{B}\Gamma_n$ (or equivalently on $\mathcal{B}_L\Gamma_n$, since the differential of any right-pointing generator $\overrightarrow{e}$ of $\mathcal{B}\Gamma_n$ is zero). Since both the differential on $m(\mathcal{B}_R(H^n))^!$ and the differential on $\mathcal{B}_L\Gamma_n$ are defined by the same formula on the degree-1 generators and extended formally to the full algebras by the Leibniz rule, we can conclude that the differential on $\mathcal{B}_L\Gamma_n$ agrees with the differential on $m(\mathcal{B}_R(H^n))^!$ after quotienting the latter algebra by the extra relations. □

5.3 The full algebra

In the following, $\mathcal{B}$ will denote $\mathcal{B}_R(H^n) \cong \mathcal{B}_R\Gamma_n$ unless otherwise specified.

The goal of this section is to construct a product algebra $m(\mathcal{B})^! \odot \mathcal{B}$ and identify $\mathcal{B}\Gamma_n$ with a quotient of $m(\mathcal{B})^! \odot \mathcal{B}$. We also want to construct a rank-one DD bimodule for $m(\mathcal{B})^! \odot \mathcal{B}$ using rank-one DD bimodules for $m(\mathcal{B})^!$ and $\mathcal{B}$.

By Proposition 4.5.3 and Proposition 4.5.4, we have rank-one type DD bimodules which we may refer to as ${}^{\mathcal{B}}K^{(\mathcal{B}^!)^{\mathrm{op}}}$ and ${}^{\mathcal{B}^!}K^{\mathcal{B}^{\mathrm{op}}}$. Like in Definition 5.2.6 above, we can extend the mirroring operation of Definition 5.2.4 to these bimodules.

5.3.1 Definition Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be differential bigraded algebras over the idempotent ring $\mathcal{I}_\beta$ and let (K,δ) be a type DD bimodule over $\mathcal{B}_1$ and $\mathcal{B}_2$. The *mirrored DD bimodule* $(m(K),\delta')$ is defined as follows: as an $(\mathcal{I}_\beta,\mathcal{I}_\beta)$–bimodule, $m(K)$ is the mirror of K as defined in Definition 5.2.4. As in Definition 5.2.6, denote the natural map from K to $m(K)$ or $m(K)$ to K by mirr. The DD operation on $m(K)$ is

$$\delta' := m(K) \xrightarrow{\text{mirr}} K \xrightarrow{\delta} \mathcal{B}_1 \otimes K \otimes (\mathcal{B}_2)^{\text{op}} \xrightarrow{\text{mirr}\otimes\text{mirr}\otimes\text{mirr}} m(\mathcal{B}_1)\otimes m(K)\otimes m(\mathcal{B}_2)^{\text{op}}.$$

Applying Definition 5.3.1 to the bimodules ${}^{\mathcal{B}}K^{(\mathcal{B}^!)^{\text{op}}}$ and ${}^{\mathcal{B}^!}K^{\mathcal{B}^{\text{op}}}$, we get rank-one DD bimodules which we will denote ${}^{m(\mathcal{B})}K^{m(\mathcal{B}^!)^{\text{op}}}$ and ${}^{m(\mathcal{B})^!}K^{m(\mathcal{B})^{\text{op}}}$.

We will focus on the DD bimodules ${}^{\mathcal{B}}K^{(\mathcal{B}^!)^{\text{op}}}$ and ${}^{m(\mathcal{B})^!}K^{m(\mathcal{B})^{\text{op}}}$. Let δ_1 and δ_2 denote the corresponding maps $\delta_1\colon \mathcal{I}_\beta \to \mathcal{B}\otimes_{\mathcal{I}_\beta}(\mathcal{B}^!)^{\text{op}}$ and $\delta_2\colon \mathcal{I}_\beta \to m(\mathcal{B})^!\otimes_{\mathcal{I}_\beta} m(\mathcal{B})^{\text{op}}$.

The set of multiplicative generators of $m(\mathcal{B})^! \odot \mathcal{B}$ will be the union of the generator sets of $\mathcal{B}$ and $m(\mathcal{B})^!$; there will be inclusion maps from $\mathcal{B}$ and $m(\mathcal{B})^!$ into $m(\mathcal{B})^!\odot\mathcal{B}$. Similarly, there will be inclusion maps from $m(\mathcal{B})$ and $\mathcal{B}^! = m(m(\mathcal{B}))^!$ into $m(m(\mathcal{B})^!\odot\mathcal{B})$, and thus maps from $m(\mathcal{B})^{\text{op}}$ and $(\mathcal{B}^!)^{\text{op}}$ into $m(m(\mathcal{B})^!\odot\mathcal{B})^{\text{op}}$. The algebra $m(\mathcal{B})^!\odot\mathcal{B}$ will be defined such that, when δ_1 and δ_2 are postcomposed with these inclusion maps, their sum

$$\delta_1+\delta_2\colon \mathcal{I}_\beta \to (m(\mathcal{B})^!\odot\mathcal{B})\otimes_{\mathcal{I}_\beta}(m(m(\mathcal{B})^!\odot\mathcal{B}))^{\text{op}}$$

satisfies the type DD structure relations.

Let $V_{\mathcal{B}}$ (respectively $V_{m(\mathcal{B})^!}$) denote the free $\mathbb{Z}$–module spanned by the multiplicative generators of $\mathcal{B}$ (respectively $m(\mathcal{B})^!$). Then $V_{\mathcal{B}}$ and $V_{m(\mathcal{B})^!}$ have left and right actions of $\mathcal{I}_\beta$, and we may write $\mathcal{B} = T(V_{\mathcal{B}})/J_{\mathcal{B}}$ and $m(\mathcal{B})^!$ as $T(V_{m(\mathcal{B})^!})/J_{m(\mathcal{B})^!}$.

Define V_{full}, as a bigraded free $\mathbb{Z}$–module, to be $V_{\mathcal{B}}\oplus V_{m(\mathcal{B})^!}$. The actions of $\mathcal{I}_\beta$ on the summands of V_{full} give V_{full} an $\mathcal{I}_\beta$–bimodule structure.

We will define $m(\mathcal{B})^!\odot\mathcal{B}$ to be $T(V_{\text{full}})/J_{\text{full}}$, for some ideal J_{full} of $T(V_{\text{full}})$. We will define J_{full} with an explicit set of linear-quadratic generators, which will agree with Roberts' relations involving both left-pointing and right-pointing elements of $\mathcal{B}\Gamma_n$.

We can start by analyzing $T^1(V_{\text{full}})\oplus T^2(V_{\text{full}})$, which is equal to

$$\begin{aligned}&(V_{\mathcal{B}}\oplus V_{m(\mathcal{B})^!})\oplus((V_{\mathcal{B}}\oplus V_{m(\mathcal{B})^!})\otimes(V_{\mathcal{B}}\oplus V_{m(\mathcal{B})^!}))\\ &\quad= V_{\mathcal{B}}\oplus V_{m(\mathcal{B})^!}\oplus(V_{\mathcal{B}}\otimes V_{\mathcal{B}})\oplus(V_{\mathcal{B}}\otimes V_{m(\mathcal{B})^!})\oplus(V_{m(\mathcal{B})^!}\otimes V_{\mathcal{B}})\oplus(V_{m(\mathcal{B})^!}\otimes V_{m(\mathcal{B})^!}).\end{aligned}$$

Thus, $T^1(V_{\text{full}})\oplus T^2(V_{\text{full}})$ is the direct sum of

$$T^1(V_{\mathcal{B}})\oplus T^2(V_{\mathcal{B}}),\quad T^1(V_{m(\mathcal{B})^!})\oplus T^2(V_{m(\mathcal{B})^!}),$$

and two more summands

$$V_{\mathcal{B}} \otimes V_{m(\mathcal{B})^!}, \quad V_{m(\mathcal{B})^!} \otimes V_{\mathcal{B}}.$$

The ideal J_{full} will be generated multiplicatively by

$$J_{\mathcal{B}} \cap (T^1(V_{\mathcal{B}}) \oplus T^2(V_{\mathcal{B}})), \quad J_{m(\mathcal{B})^!} \cap (T^1(V_{m(\mathcal{B})^!}) \oplus T^2(V_{m(\mathcal{B})^!})),$$

and some extra relations

$$J_{\text{extra}} \subset (V_{\mathcal{B}} \otimes V_{m(\mathcal{B})^!}) \oplus (V_{m(\mathcal{B})^!} \otimes V_{\mathcal{B}}).$$

5.3.2 Definition $J_{\text{extra}} \subset (V_{\mathcal{B}} \otimes V_{m(\mathcal{B})^!}) \oplus (V_{m(\mathcal{B})^!} \otimes V_{\mathcal{B}})$ is defined additively by the following relations:

(1) For two bridge generators $b_{\gamma;h_1,h_2}$ and $m(b^*_{\eta';m(h_2),m(h_3)})$, the "commutation relation"

$$b_{\gamma;h_1,h_2} m(b^*_{\eta';m(h_2),m(h_3)}) - m(b^*_{\eta;m(h_1),m(\tilde{h}_2)}) b_{\gamma';\tilde{h}_2,h_3}$$

is in J_{extra} for any $\tilde{h}_2 \in \beta$ such that $m(b^*_{\eta;m(h_1),m(\tilde{h}_2)})$ and $b_{\gamma';\tilde{h}_2,h_3}$ exist. The bridges γ' and η are uniquely determined. We will call such relations *commutation relations* even though they do not exactly express that two elements commute.

(2) Any generator $b_{C;h_1,h_2}$ of $\mathcal{B} = \mathcal{B}_R(H^n)$ for $n > 1$ can be written as a product of bridge generators $b_\gamma b_{\gamma^\dagger}$. Thus, by the commutation relations above, $b_{C;h_1,h_2}$ must also commute with bridge generators $m(b^*_{\eta;m(h_2),m(h_3)})$: the relation

$$b_{C;h_1,h_2} m(b^*_{\eta;m(h_2),m(h_3)}) - m(b^*_{\eta;m(h_1),m(\tilde{h}_2)}) b_{C;\tilde{h}_2,h_3}$$

must be in J_{extra} for any $\tilde{h}_2 \in \beta$ such that $m(b^*_{\eta;m(h_1),m(\tilde{h}_2)})$ and $b_{C;\tilde{h}_2,h_3}$ exist.

(3) For a bridge generator $b_{\gamma;h_1,h_2}$ and a decoration generator $m(b^*_{C;m(h_2),m(h_3)})$ in which the circle C is disjoint from the circles involved in surgery on γ, we also impose commutation relations:

$$b_{\gamma;h_1,h_2} m(b^*_{C;m(h_2),m(h_3)}) - m(b^*_{C;m(h_1),m(\tilde{h}_2)}) b_{\gamma;\tilde{h}_2,h_3}$$

must be in J_{extra}, for the unique $\tilde{h}_2 \in \beta$ such that $m(b^*_{C;m(h_1),m(\tilde{h}_2)})$ and $b_{\gamma;\tilde{h}_2,h_3}$ exist.

(4) For two disjoint circles C and C', we again have commutation relations:

$$b_{C;h_1,h_2} m(b^*_{C';m(h_2),m(h_3)}) - m(b^*_{C';m(h_1),m(\tilde{h}_2)}) b_{C;\tilde{h}_2,h_3}$$

must be in J_{extra}, for the unique $\tilde{h}_2 \in \beta$ such that $m(b^*_{C';m(h_1),m(\tilde{h}_2)})$ and $b_{C;\tilde{h}_2,h_3}$ exist.

(5) Finally, let γ be a bridge and let C be one of the circles involved in surgery on γ. When γ joins C' and C'' to form C,

$$b_{\gamma;h_1,h_2} m(b^*_{C;m(h_2),m(h_3)}) - m(b^*_{C';m(h_1),m(\tilde{h}_2)}) b_{\gamma;\tilde{h}_2,h_3} - m(b^*_{C'';m(h_1),m(\tilde{\tilde{h}}_2)}) b_{\gamma;\tilde{\tilde{h}}_2,h_3}$$

is in J_{extra} for the unique $\tilde{h}_2 \in \beta$ and $\tilde{\tilde{h}}_2 \in \beta$ such that the relevant generators exist, and when γ splits C to form C' and C'',

$$b_{\gamma;h_1,\tilde{h}_2} m(b^*_{C';m(\tilde{h}_2),m(h_3)}) + b_{\gamma;h_1,\tilde{\tilde{h}}_2} m(b^*_{C'';m(\tilde{\tilde{h}}_2),m(h_3)}) - m(b^*_{C;m(h_1),m(h_2)}) b_{\gamma;h_2,h_3}$$

is in J_{extra} for the unique $\tilde{h}_2 \in \beta$ and $\tilde{\tilde{h}}_2 \in \beta$ such that the relevant generators exist.

5.3.3 Definition The ideal J_{full} is defined by

$$\begin{aligned} J_{\mathrm{full}} := T(V_{\mathrm{full}}) \cdot \big((J_{\mathcal{B}} \cap (T^1(V_{\mathcal{B}}) \oplus T^2(V_{\mathcal{B}}))) \\ \oplus \big(J_{m(\mathcal{B})^!} \cap (T^1(V_{m(\mathcal{B})^!}) \oplus T^2(V_{m(\mathcal{B})^!})) \big) \oplus J_{\mathrm{extra}} \big) \cdot T(V_{\mathrm{full}}). \end{aligned}$$

The differential bigraded algebra $m(\mathcal{B})^! \odot \mathcal{B}$ is defined by

$$m(\mathcal{B})^! \odot \mathcal{B} := T(V_{\mathrm{full}}) / J_{\mathrm{full}},$$

with a differential induced from the differential on $m(\mathcal{B})^!$. The differential of any generator of $\mathcal{B}$ is declared to be zero.

5.3.4 Corollary *$\mathcal{B}\Gamma_n$, as a differential bigraded algebra, is the quotient of the algebra $m(\mathcal{B})^! \odot \mathcal{B}$ by the same additional relations as specified in Proposition 5.2.8. These relations involve only quadratic monomials with two generators of $m(\mathcal{B})^!$.*

Proof The relations in J_{extra} are modeled on Roberts' relations for $\mathcal{B}\Gamma_n$ in [12] involving quadratic monomials with one left-pointing and one right-pointing generator; see [12, Section 2.3]. □

5.3.5 Remark The relations in Definition 5.3.2 were chosen to match Roberts' quadratic relations involving a left-pointing and a right-pointing generator. A more general formulation of the product operation $\odot$ would be desirable. The best we can do now is to say that the relations in J_{extra} have an additional motivation beyond lining up with Roberts' relations: with these relations, $\delta_1 + \delta_2$ defines a valid rank-one DD bimodule over the product algebra, as we see below in Proposition 5.3.6.

From the definition of the product algebra $m(\mathcal{B})^! \odot \mathcal{B}$, there are natural inclusion maps of $\mathcal{B}$ and $m(\mathcal{B})^!$ into the product. Also, by Definition 5.2.6, we have a mirror algebra $m(m(\mathcal{B})^! \odot \mathcal{B})$; both $m(\mathcal{B})$ and $\mathcal{B}^! = m(m(\mathcal{B})^!)$ have natural inclusion maps into $m(m(\mathcal{B})^! \odot \mathcal{B})$.

We may view the type DD map

$$\delta_1\colon \mathcal{I}_\beta \to \mathcal{B} \otimes_{\mathcal{I}_\beta} (\mathcal{B}^!)^{\mathrm{op}} = \mathcal{B} \otimes_{\mathcal{I}_\beta} m(m(\mathcal{B})^!)^{\mathrm{op}}$$

as a map from $\mathcal{I}_\beta$ to $(m(\mathcal{B})^! \odot \mathcal{B}) \otimes_{\mathcal{I}_\beta} (m(m(\mathcal{B})^! \odot \mathcal{B}))^{\mathrm{op}}$, using the inclusion maps from $\mathcal{B}$ into $m(\mathcal{B})^! \odot \mathcal{B}$ and from $m(m(\mathcal{B})^!)^{\mathrm{op}}$ into $(m(m(\mathcal{B})^! \odot \mathcal{B}))^{\mathrm{op}}$. Similarly, we may view

$$\delta_2\colon \mathcal{I}_\beta \to m(\mathcal{B})^! \otimes (m(\mathcal{B}))^{\mathrm{op}}$$

as a map from $\mathcal{I}_\beta$ to $(m(\mathcal{B})^! \odot \mathcal{B}) \otimes_{\mathcal{I}_\beta} (m(m(\mathcal{B})^! \odot \mathcal{B}))^{\mathrm{op}}$ using the inclusion maps from $m(\mathcal{B})^!$ into $m(\mathcal{B})^! \odot \mathcal{B}$ and from $m(\mathcal{B})^{\mathrm{op}}$ into $(m(m(\mathcal{B})^! \odot \mathcal{B}))^{\mathrm{op}}$.

5.3.6 Proposition *The map $\delta_1 + \delta_2\colon \mathcal{I}_\beta \to (m(\mathcal{B})^! \odot \mathcal{B}) \otimes_{\mathcal{I}_\beta} (m(m(\mathcal{B})^! \odot \mathcal{B}))^{\mathrm{op}}$ satisfies the type DD structure relations.*

Proof Many of the type DD structure terms cancel since δ_1 and δ_2 individually satisfy the type DD relations. In particular, all terms of type $(\mu_1 \otimes |\mathrm{id}|) \circ \delta$ and $(\mathrm{id} \otimes \mu_1) \circ \delta$ are accounted for and we are left with terms of type $(\mu_2 \otimes \mu_2) \circ \sigma \circ (\mathrm{id} \otimes \delta \otimes \mathrm{id}) \circ \delta$.

The remaining terms of type $(\mu_2 \otimes \mu_2) \circ \sigma \circ (\mathrm{id} \otimes \delta \otimes \mathrm{id}) \circ \delta$ are those in which one of the applications of δ uses δ_1 and the other uses δ_2. These are

$$-b_i \cdot m(b_j^*) \otimes m(b_j)^{\mathrm{op}} \cdot (b_i^*)^{\mathrm{op}} = -b_i \cdot m(b_j^*) \otimes m(m(b_i^*) \cdot b_j)^{\mathrm{op}}, \tag{5-1}$$

referred to as terms of type (5-1), as well as

$$m(b_j^*) \cdot b_i \otimes (b_i^*)^{\mathrm{op}} \cdot m(b_j)^{\mathrm{op}} = m(b_j^*) \cdot b_i \otimes m(b_j \cdot m(b_i^*))^{\mathrm{op}}, \tag{5-2}$$

referred to as terms of type (5-2), where b_i and b_j^* run over all generators of $\mathcal{B}$ and $\mathcal{B}^!$, respectively, with compatible idempotents. Note that the negative signs in the terms of type (5-1) come from the sign-flip operator σ.

The commutation relations among the relations defining $m(\mathcal{B})^! \odot \mathcal{B}$ ensure that all the above terms cancel, except for potentially two sets of terms.

The first set S_1 of terms includes those terms $-b_i \cdot m(b_j^*) \otimes m(m(b_i^*) \cdot b_j)^{\mathrm{op}}$ of type (5-1) in which both b_i and b_j are among the generators b_γ, and such that the product $b_i \cdot m(b_j^*)$ corresponds to splitting a circle C on the right into two circles C_1 and C_2 and then joining C_1 to C_2 again on the left to produce a new circle C_3. It also includes those terms $m(b_j^*) \cdot b_i \otimes m(b_j \cdot m(b_i^*))^{\mathrm{op}}$ of type (5-2) with b_i and b_j

among the generators b_γ such that $m(b_j^*)\cdot b_i$ corresponds to splitting a circle C on the left into two circles C_1 and C_2 and then joining C_1 and C_2 again on the right to produce a new circle C_3.

The second set S_2 consists of those terms of type (5-1) or (5-2) in which b_i is one of the generators b_γ and b_j is one of the generators b_C, where C is one of the circles involved in surgery on γ.

For all terms except those in S_1 and S_2, commutation relations may be applied to the type (5-1) term $-b_i\cdot m(b_j^*)\otimes m(m(b_i^*)\cdot b_j)^{\mathrm{op}}$ uniquely to cancel with a unique corresponding type (5-2) term $m(b_{j'}^*)\cdot b_{i'}\otimes m(b_{j'}\cdot m(b_{i'}^*))^{\mathrm{op}}$.

First, we show that the terms in S_1 sum to zero. If the type (5-1) term $-b_i\cdot m(b_j^*)\otimes m(m(b_i^*)\cdot b_j)^{\mathrm{op}}$ is in S_1, then there are two terms $m(b_{j'}^*)\cdot b_{i'}\otimes m(b_{j'}\cdot m(b_{i'}^*))^{\mathrm{op}}$ and $m(b_{j''}^*)\cdot b_{i''}\otimes m(b_{j''}\cdot m(b_{i''}^*))^{\mathrm{op}}$ of type (5-2) in S_1 which have the same left and right idempotents as $-b_i\cdot m(b_j^*)\otimes m(m(b_i^*)\cdot b_j)^{\mathrm{op}}$. By the commutation relations, both these terms are equal to $b_i\cdot m(b_j^*)\otimes m(m(b_i^*)\cdot b_j)^{\mathrm{op}}$.

Furthermore, there is one other type (5-1) term $-b_{i'''}\cdot m(b_{j'''}^*)\otimes m(m(b_{i'''}^*)\cdot b_{j'''})^{\mathrm{op}}$ of S_1 which is equal to both $-m(b_{j'}^*)\cdot b_{i'}\otimes m(b_{j'}\cdot m(b_{i'}^*))^{\mathrm{op}}$ and $-m(b_{j''}^*)\cdot b_{i''}\otimes m(b_{j''}\cdot m(b_{i''}^*))^{\mathrm{op}}$ by the commutation relations. Hence it is equal to $-b_i\cdot m(b_j^*)\otimes m(m(b_i^*)\cdot b_j)^{\mathrm{op}}$ as well. These four terms are the only terms in S_1 with the same idempotents as $-b_i\cdot m(b_j^*)\otimes m(m(b_i^*)\cdot b_j)^{\mathrm{op}}$, and their sum is zero. Thus, the terms in S_1 sum to zero.

Now we show that the terms in S_2 sum to zero. If b_i is a b_γ generator and b_j is a b_C generator with C involved in surgery on γ, then suppose first that γ joins two circles C_1 and C_2 to produce C. By item (5) of Definition 5.3.2, we have $b_i\cdot m(b_j^*)=m(b_j'^*)\cdot b_i'+m(b_j''^*)\cdot b_i''$, where b_j' and b_j'' are the generators b_{C_1} and b_{C_2}, and b_i' and b_i'' are the appropriate b_γ generators.

Thus, if $-b_i\cdot m(b_j^*)\otimes m(m(b_i^*)\cdot b_j)^{\mathrm{op}}$ is the corresponding term of type (5-1), we have

$$\begin{aligned}
-b_i\cdot m(b_j^*)&\otimes m(m(b_i^*)\cdot b_j)^{\mathrm{op}}\\
&=-m(b_j'^*)\cdot b_i'\otimes m(m(b_i^*)\cdot b_j)^{\mathrm{op}}-m(b_j''^*)\cdot b_i''\otimes m(m(b_i^*)\cdot b_j)^{\mathrm{op}}\\
&=-m(b_j'^*)\cdot b_i'\otimes m(b_j'\cdot m(b_i'^*))^{\mathrm{op}}-m(b_j''^*)\cdot b_i''\otimes m(b_j''\cdot m(b_i''^*))^{\mathrm{op}},
\end{aligned}$$

where in the last step we use commutation relations from item (2) of Definition 5.3.2. The two resulting terms cancel the two relevant terms of type (5-2). The case when γ splits a circle, rather than joining two circles, is analogous, so the terms in S_2 sum to zero.

Hence all the relation terms cancel and $(\delta_1 + \delta_2)$ satisfies the type DD structure relations. $\square$

A general property of type D structures and type DD bimodules over an algebra $\mathcal{B}$ is they give induced type D or DD structures over any quotient of $\mathcal{B}$.

5.3.7 Proposition *Let $\mathcal{B}$ be a differential bigraded algebra; let J be a bigrading-homogeneous ideal of $\mathcal{B}$ which is preserved by the differential on $\mathcal{B}$. Let $\pi\colon \mathcal{B} \to \mathcal{B}/J$ denote the quotient projection map. Let $(\widehat{D}, \delta)$ be a type D structure over $\mathcal{B}$; then $\widehat{D}$ descends to a type D structure over $\mathcal{B}/J$, with structure operation*

$$\widehat{D} \xrightarrow{\delta} \mathcal{B} \otimes \widehat{D} \xrightarrow{\pi \otimes \mathrm{id}} (\mathcal{B}/J) \otimes \widehat{D}.$$

Similarly, if $\mathcal{B}'$ and J' are another algebra and ideal satisfying the same conditions as $\mathcal{B}$ and J, and $(\widehat{DD}, \delta)$ is a type DD bimodule over $\mathcal{B}$ and $\mathcal{B}'$, then $\widehat{DD}$ descends to a type DD bimodule over $\mathcal{B}/J$ and $\mathcal{B}'/J'$, with structure operation

$$\widehat{DD} \xrightarrow{\delta} \mathcal{B} \otimes \widehat{DD} \otimes (\mathcal{B}')^{\mathrm{op}} \xrightarrow{\pi \otimes \mathrm{id} \otimes (\pi')^{\mathrm{op}}} (\mathcal{B}/J) \otimes \widehat{DD} \otimes (\mathcal{B}'/J')^{\mathrm{op}}.$$

Proof This is a simple consequence of the type D and type DD structure relations. It is also a special case of induction of scalars for type D structures as defined by Lipshitz, Ozsváth and Thurston [7, Section 2.4.2]. $\square$

We know that $\mathcal{B}\Gamma_n$ is a quotient of $m(\mathcal{B})^! \odot \mathcal{B}$, and it follows that $m(\mathcal{B}\Gamma_n)^{\mathrm{op}}$ is a quotient of $m(m(\mathcal{B})^! \odot \mathcal{B})^{\mathrm{op}}$.

5.3.8 Corollary *The map $\mathcal{I}_\beta \to \mathcal{B}\Gamma_n \otimes_{\mathcal{I}_\beta} m(\mathcal{B}\Gamma_n)^{\mathrm{op}}$ obtained by postcomposing $\delta_1 + \delta_2$ with the tensor product of the quotient projections from $m(\mathcal{B})^! \odot \mathcal{B}$ and $m(m(\mathcal{B})^! \odot \mathcal{B})^{\mathrm{op}}$ onto $\mathcal{B}\Gamma_n$ and $m(\mathcal{B}\Gamma_n)^{\mathrm{op}}$ satisfies the type DD structure relations.*

Thus, we have rank-one type DD bimodules

$${}^{m(\mathcal{B})^! \odot \mathcal{B}}K^{m(m(\mathcal{B})^! \odot \mathcal{B})^{\mathrm{op}}} \quad \text{and} \quad {}^{\mathcal{B}\Gamma_n}K^{m(\mathcal{B}\Gamma_n)^{\mathrm{op}}}.$$

5.3.9 Conjecture *Either or both of the DD bimodules ${}^{m(\mathcal{B})^! \odot \mathcal{B}}K^{m(m(\mathcal{B})^! \odot \mathcal{B})^{\mathrm{op}}}$ and ${}^{\mathcal{B}\Gamma_n}K^{m(\mathcal{B}\Gamma_n)^{\mathrm{op}}}$ are quasi-invertible. Hence, either or both of the algebras $m(\mathcal{B})^! \odot \mathcal{B}$ and $\mathcal{B}\Gamma_n$ are Koszul dual to their mirrors, $m(m(\mathcal{B})^! \odot \mathcal{B})$ and $m(\mathcal{B}\Gamma_n)$, in the generalized sense of Lipshitz, Ozsváth and Thurston* [6].

A proof of the above conjecture would provide a nice parallel between Roberts' theory and bordered Floer homology. In bordered Floer homology, the rank-one DD bimodule corresponding to the identity Heegaard diagram of a pointed matched circle has a quasi-inverse, namely the type AA bimodule associated to this diagram. See [6] for more on the Koszul duality structure of bordered Floer homology, including an additional duality relating the algebras of a pointed matched circle $\mathcal{Z}$ and a dual pointed matched circle $\mathcal{Z}_*$.

6 Khovanov's modules and Roberts' modules

In this section, we relate Roberts' type D and type A structures from [11; 12] to the type D and type A structures over H^n from Section 3, or equivalently to Khovanov's dg modules $[T]^{\mathrm{Kh}}$ which contain the same information. In Section 6.2, we show that given a chain complex of projective graded right H^n–modules satisfying an algebraic condition C_{module} defined below in Definition 6.1.1, we may construct a differential bigraded right module over $m(\mathcal{B})^! \odot \mathcal{B}$. Applied to Khovanov's tangle complex $[T]^{\mathrm{Kh}}$, which satisfies C_{module}, this module over $m(\mathcal{B})^! \odot \mathcal{B}$ descends to a module over $\mathcal{B}\Gamma_n$; in other words, the relations of Proposition 5.2.8 act as zero on the $m(\mathcal{B})^! \odot \mathcal{B}$–module. The resulting $\mathcal{B}\Gamma_n$–module agrees with Roberts' type A structure.

In Section 6.3, given a chain complex of projective graded left H^n–modules satisfying the algebraic condition C_{module} for left modules defined in Definition 6.3.3, we construct a type D structure over $m(\mathcal{B})^! \odot \mathcal{B}$. We do this by, first, reflecting the chain complex of left H^n–modules to get a complex of right modules (this operation will be defined in Definition 6.3.1). Then we take the associated type A structure over $m(\mathcal{B})^! \odot \mathcal{B}$, tensor with ${}^{m(\mathcal{B})^! \odot \mathcal{B}}K^{m(m(\mathcal{B})^! \odot \mathcal{B})^{\mathrm{op}}}$, the DD bimodule from the end of Section 5.3, to get a type D structure over $m(m(\mathcal{B})^! \odot \mathcal{B})$, and finally mirror this type D structure to get a type D structure over $m(\mathcal{B})^! \odot \mathcal{B}$. We may quotient the algebra outputs of this type D structure by the relations from Proposition 5.2.8 to get a type D structure over $\mathcal{B}\Gamma_n$, which agrees with the one constructed by Roberts when one starts with the complex $[T]^{\mathrm{Kh}}$.

Given two chain complexes, one of projective graded left H^n–modules and one of projective graded right H^n–modules, their tensor product over H^n is a chain complex with an additional grading, or equivalently a differential bigraded $\mathbb{Z}$–module. In Section 6.4, we show that this tensor product agrees with the box tensor product of the type D and type A structures over $m(\mathcal{B})^! \odot \mathcal{B}$ associated to the two complexes in Section 6.3 and Section 6.2, assuming these complexes satisfy C_{module}. The type A structure is always an ordinary right $m(\mathcal{B})^! \odot \mathcal{B}$–module; if it descends to a $\mathcal{B}\Gamma_n$–module, then the box tensor products of the type A and type D structures over $m(\mathcal{B})^! \odot \mathcal{B}$ and $\mathcal{B}\Gamma_n$ agree. Applying these constructions to Khovanov's chain complexes $[T]^{\mathrm{Kh}}$ of H^n–modules, we get an alternate proof that the pairing of Roberts' type D and type A structures recovers the original Khovanov complex.

In Section 6.5, we show that chain homotopy equivalences of complexes of H^n–modules, satisfying appropriate algebraic conditions, give $\mathcal{A}_\infty$–homotopy equivalences of the corresponding type A structures over $m(\mathcal{B})^! \odot \mathcal{B}$. Reidemeister moves on tangle diagrams yield chain homotopy equivalences of complexes of H^n–modules, as shown by Khovanov in [4]. These Reidemeister-move homotopy equivalences satisfy the right conditions, and the $\mathcal{A}_\infty$–homotopy equivalences associated to them descend to $\mathcal{A}_\infty$–homotopy equivalences of type A structures over $\mathcal{B}\Gamma_n$. This reasoning yields an alternate proof that Roberts' type A structures are tangle invariants up to $\mathcal{A}_\infty$–homotopy equivalence.

In Section 6.6, we do the same for the type D structures over $m(\mathcal{B})^! \odot \mathcal{B}$. All homotopy equivalences of type D structures over $m(\mathcal{B})^! \odot \mathcal{B}$ descend to homotopy equivalences of type D structures over the quotient $\mathcal{B}\Gamma_n$. Thus, we obtain an alternate proof that Roberts' type D structures are tangle invariants up to homotopy equivalence.

6.1 Preliminaries

Let M be a differential bigraded projective right H^n–module, or equivalently a chain complex of projective graded H^n–modules by Proposition 2.2.4 or a right type D structure over H^n by the appropriate analogue of Proposition 2.2.5. Recall that in accordance with Convention 2.1.1, such an M is assumed to be finitely generated over $\mathbb{Z}$. Let $\{x_i \mid i \in S\}$ be bigrading-homogeneous elements of M, where S is some finite index set, such that

$$M \cong \bigoplus_i x_i H^n$$

as right H^n–modules and each summand $x_i H^n$ is isomorphic to eH^n for some elementary idempotent e of H^n via an isomorphism sending x_i to e. The idempotent e associated to x_i will be denoted $e(x_i)$.

We will use notation from Section 3.1: β will denote the usual $\mathbb{Z}$–basis of H^n and β_{mult} will denote the subset of β consisting of the multiplicative generators h_γ and h_α of H^n. We will further subdivide β_{mult} into β_γ, consisting of h_γ generators, and β_α, consisting of h_α generators. Recall that for $h \in \beta$, $e_L(h) \in \mathcal{I}_n$ is the left idempotent of h and $e_R(h)$ is the right idempotent of h.

The module M has a $\mathbb{Z}$–basis given by

$$\{x_i \cdot h_1 \mid i \in S, h_1 \in \beta, e_L(h_1) = e(x_i)\}.$$

We define integer coefficients $c_{i,j}$ and $\tilde{c}_{i,j;h'}$ for $i, j \in S$ and $h' \in \beta$ by expanding out the differential of each x_i:

$$d(x_i) = \sum_{\substack{j \in S \\ \deg x_j = \deg x_i + (0,1)}} c_{i,j} x_j + \sum_{\substack{j \in S,\, h' \in \beta,\, \deg h' \neq 0 \\ \deg x_j + \deg h' = \deg x_i + (0,1)}} \tilde{c}_{i,j;h'} x_j \cdot h'.$$

6.1.1 Definition M, together with the set of generators $\{x_i\}$, satisfies the algebraic condition C_{module} if $\tilde{c}_{i,j;h'} = 0$ unless $h' \in \beta_{\text{mult}}$.

This condition is satisfied for Khovanov's tangle complexes $[T]^{\text{Kh}}$; the natural choice of $\{x_i\}$ was described in Definition 3.1.2. By slight abuse of notation, we will speak of M satisfying C_{module}, but the choice of $\{x_i\}$ was necessary to define the coefficients $c_{i,j}$ and $\tilde{c}_{i,j;h'}$.

For any M satisfying C_{module}, we can write $d(x_i)$ as

$$\begin{aligned} d(x_i) = \sum_{\substack{j \in S \\ \deg x_j = \deg x_i + (0,1)}} c_{i,j} x_j &+ \sum_{\substack{j \in S, h' \in \beta_\gamma \\ \deg x_j = \deg x_i + (-1,1)}} \tilde{c}_{i,j;h'} x_j \cdot h' \\ &+ \sum_{\substack{j \in S, h' \in \beta_\alpha \\ \deg x_j = \deg x_i + (-2,1)}} \tilde{c}_{i,j;h'} x_j \cdot h'. \end{aligned}$$

This is an expansion of $d(x_i)$ in the $\mathbb{Z}$–basis of M.

Thus, if $x_i \cdot h_1$ is a basis element of M, we have

$$\begin{aligned} d(x_i \cdot h_1) = \sum_{\substack{j \in S \\ \deg x_j = \deg x_i + (0,1)}} c_{i,j} x_j \cdot h_1 &+ \sum_{\substack{j \in S, h' \in \beta_\gamma \\ \deg x_j = \deg x_i + (-1,1)}} \tilde{c}_{i,j;h'} x_j \cdot h' h_1 \\ &+ \sum_{\substack{j \in S, h' \in \beta_\alpha \\ \deg x_j = \deg x_i + (-2,1)}} \tilde{c}_{i,j;h'} x_j \cdot h' h_1. \end{aligned}$$

However, this is not necessarily a basis expansion of $d(x_i \cdot h_1)$, because the elements $h' h_1 \in H^n$ are not necessarily elements of the basis β. Instead, we may define integer coefficients $\tilde{\tilde{c}}$ by

$$h' h_1 := \sum_{h_2 \in \beta} \tilde{\tilde{c}}_{h' h_1; h_2} h_2$$

and thus

$$d(x_i \cdot h_1) = \sum_{\substack{j \in S \\ \deg x_j = \deg x_i + (0,1)}} c_{i,j} x_j \cdot h_1 + \sum_{\substack{j \in S, h' \in \beta_\gamma, h_2 \in \beta \\ \deg x_j = \deg x_i + (-1,1)}} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h' h_1; h_2} x_j \cdot h_2$$
$$+ \sum_{\substack{j \in S, h' \in \beta_\alpha, h_2 \in \beta \\ \deg x_j = \deg x_i + (-2,1)}} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h' h_1; h_2} x_j \cdot h_2.$$

This is a basis expansion of $d(x_i \cdot h_1)$ in the $\mathbb{Z}$–basis of M.

6.1.2 Proposition *Suppose M satisfies C_{module}. The equation $d^2 = 0$ on M gives rise to the following five sets of equations involving the coefficients $c_{i,j}$, $\tilde{c}_{i,j,h'}$ and $\tilde{\tilde{c}}_{h' h_1; h_2}$:*

(1) *For all x_i and x_k with $\deg x_k = \deg x_i + (0,2)$, we have*
$$\sum_j c_{i,j} c_{j,k} = 0.$$

(2) *For all $x_i \cdot h_1$ and $x_k \cdot h_3$ with $\deg x_k = \deg x_i + (-1,2)$, we have*
$$\sum_{j, h' \in \beta_\gamma} (\tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h' h_1; h_3} c_{j,k} + c_{i,j} \tilde{c}_{j,k;h'} \tilde{\tilde{c}}_{h' h_1; h_3}) = 0.$$

(3) *For all $x_i \cdot h_1$ and $x_k \cdot h_3$ with $\deg x_k = \deg x_i + (-2,2)$, we have*
$$\sum_{j, h' \in \beta_\alpha} (\tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h' h_1; h_3} c_{j,k} + c_{i,j} \tilde{c}_{j,k;h'} \tilde{\tilde{c}}_{h' h_1; h_3})$$
$$+ \sum_{\substack{j, h' \in \beta_\gamma, \\ h'' \in \beta_\gamma, h_2 \in \beta}} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h' h_1; h_2} \tilde{c}_{j,k;h''} \tilde{\tilde{c}}_{h'' h_2; h_3} = 0.$$

(4) *For all $x_i \cdot h_1$ and $x_k \cdot h_3$ with $\deg x_k = \deg x_i + (-3,2)$, we have*
$$\sum_{\substack{j, h' \in \beta_\gamma, \\ h'' \in \beta_\alpha, h_2 \in \beta}} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h' h_1; h_2} \tilde{c}_{j,k;h''} \tilde{\tilde{c}}_{h'' h_2; h_3} + \sum_{\substack{j, h' \in \beta_\alpha, \\ h'' \in \beta_\gamma, h_2 \in \beta}} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h' h_1; h_2} \tilde{c}_{j,k;h''} \tilde{\tilde{c}}_{h'' h_2; h_3} = 0.$$

(5) *For all $x_i \cdot h_1$ and $x_k \cdot h_3$ with $\deg x_k = \deg x_i + (-4,2)$, we have*
$$\sum_{\substack{j, h' \in \beta_\alpha, \\ h'' \in \beta_\alpha, h_2 \in \beta}} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h' h_1; h_2} \tilde{c}_{j,k;h''} \tilde{\tilde{c}}_{h'' h_2; h_3} = 0.$$

Proof This follows from writing out $d^2(x_i \cdot h_1)$ as a sum of basis elements $x_k \cdot h_3$, using the above basis expansion for $d(x_i \cdot h_1)$, and then grouping the $x_k \cdot h_3$ according to the intrinsic degree of x_k relative to x_i. □

6.1.3 Example Suppose $M = [T]^{\mathrm{Kh}}$, where T is an oriented tangle diagram in $\mathbb{R}_{\le 0} \times \mathbb{R}$ with an ordering of its crossings. We will analyze the generators x_i and coefficients $\tilde{c}_{i,j;h'}$ and $c_{i,j}$. To specify a generator x_i of $[T]^{\mathrm{Kh}}$, we first specify a resolution ρ_i of all crossings of T; we can view ρ_i as a function from the set of crossings to the two-element set $\{0, 1\}$. If T_{ρ_i} denotes the diagram T with the crossings resolved according to ρ_i, then T_{ρ_i} consists of a left crossingless matching of $2n$ points together with some free circles contained in $\mathbb{R}_{<0} \times \mathbb{R}$. The remaining data needed to specify x_i are a choice of $+$ (plus) or $-$ (minus) on each free circle; the crossingless-matching part of T_{ρ_i} is closed up symmetrically and all resulting circles are labeled plus. Then $[T]^{\mathrm{Kh}}$ has a $\mathbb{Z}$–basis consisting of elements $x_i \cdot h$, where the left crossingless matching of h agrees with the matching obtained from T_{ρ_i} by discarding the free circles.

Let S denote the set of x_i specified above. The basis expansion defining $c_{i,j}$ and $\tilde{c}_{i,j;h'}$ is

$$d(x_i) = \sum_{\substack{j \in S \\ \deg x_j = \deg x_i + (0,1)}} c_{i,j} x_j + \sum_{\substack{j \in S, h' \in \beta_\gamma \\ \deg x_j = \deg x_i + (-1,1)}} \tilde{c}_{i,j;h'} x_j \cdot h' + \sum_{\substack{j \in S, h' \in \beta_\alpha \\ \deg x_j = \deg x_i + (-2,1)}} \tilde{c}_{i,j;h'} x_j \cdot h'.$$

The coefficients $c_{i,j}$ and $\tilde{c}_{i,j;h'}$ can only be nonzero when the resolution ρ_j of x_j differs from the resolution ρ_i of x_i only at one crossing, to which ρ_i assigns 0 and ρ_j assigns 1. Let $\#_1(i, j)$ denote the number of 1–resolutions of crossings in x_i among those crossings that, in the ordering on crossings, occur earlier than the crossing being changed when going from x_i to x_j.

Changing the crossing to get from x_i to x_j has several possible effects:

(1) The crossing change could join two free circles or split a free circle. In this case, $c_{i,j}$ is $(-1)^{\#_1(i,j)}$ and all $\tilde{c}_{i,j;h'}$ are zero.

(2) The crossing change could join a free circle in x_i, labeled $+$, with an arc of x_i. Alternatively, it could split a new free circle, labeled $-$ in x_j, off an arc of x_i. In both these cases, $c_{i,j}$ is $(-1)^{\#_1(i,j)}$ and all $\tilde{c}_{i,j;h'}$ are zero.

(3) The crossing change could join a free circle in x_i, labeled $-$, with an arc α of x_i. Alternatively, it could split a new free circle, labeled $+$ in x_j, off an arc α of x_i. In both these cases, $c_{i,j}$ is zero and $\tilde{c}_{i,j;h'}$ is only nonzero for one value of h'. If a denotes the crossingless matching of x_i (or of x_j), then when $h' = (W(a)a,\ \text{minus on } \alpha)$, we have $\tilde{c}_{i,j;h'} = (-1)^{\#_1(i,j)}$; all other $\tilde{c}_{i,j;h'}$ are zero.

(4) Finally, the crossing change could surger two arcs of x_i, changing the crossingless matching. Again, $c_{i,j}=0$ and $\tilde{c}_{i,j;h'}$ is nonzero for a unique h'. Let a_i and a_j denote the crossingless matchings of x_i and x_j respectively. Then if $h'=(W(a_j)a_i, \text{all plus})$, we have $\tilde{c}_{i,j;h'}=(-1)^{\#_1(i,j)}$; all other $\tilde{c}_{i,j;h'}$ are zero.

Note that $[T]^{\mathrm{Kh}}$ satisfies C_{module} with the elements $x_i=x_i\cdot 1$ as generators.

6.2 Type A structures

As in Section 5.3, let $\mathcal{B}$ denote $\mathcal{B}_R(H^n)\cong\mathcal{B}_R\Gamma_n$. Recall that $\mathcal{B}_L\Gamma_n$ is the quotient of $m(\mathcal{B})^!$ by the extra relations listed in Proposition 5.2.8.

Let M be a differential bigraded projective right H^n module as at the beginning of Section 6.1; assume that M satisfies the algebraic condition C_{module} of Definition 6.1.1 for a set of generators $\{x_i \mid i\in S\}$. We first define a type A structure $\widehat{A}(M)_{m(\mathcal{B})^!}$ over $m(\mathcal{B})^!$. Then we formally extend $\widehat{A}(M)$ to a type A structure $\widehat{A}(M)_{m(\mathcal{B})^!\odot\mathcal{B}}$ over $m(\mathcal{B})^!\odot\mathcal{B}$.

6.2.1 Definition As a $\mathbb{Z}$–module, $\widehat{A}(M)$ is defined to be M. A $\mathbb{Z}$–basis for M is given by $\{x_i\cdot h_1 \mid i\in S, h_1\in\beta, e_L(h_1)=e(x_i)\}$, where $\{x_i\mid i\in S\}$ is the designated set of generators of M. For $\widehat{A}(M)$, we label the same basis elements as

$$\{X_{x_i\cdot h_1}\mid i\in S, h_1\in\beta, e_L(h_1)=e(x_i)\}.$$

The idempotent ring of $m(\mathcal{B})^!$ is $\mathcal{I}_\beta$; let $h_2\in\beta$ be an elementary idempotent of $m(\mathcal{B})^!$. Multiplying $X_{x_i\cdot h_1}$ by h_2 gives $X_{x_i\cdot h_1}$ if $h_2=h_1$ and zero otherwise.

Suppose that the generator x_i has bigrading (j,k) as an element of M, and h has grading j' (or bigrading $(j',0)$) as an element of H^n. Then, as an element of $\widehat{A}(M)$, the bigrading of $X_{x_i\cdot h_1}$ is defined to be

$$\deg_{\widehat{A}(M)}(X_{x_i\cdot h}) := \left(-j-\tfrac{1}{2}j', k\right).$$

The algebra $m(\mathcal{B})^!$ acts on $\widehat{A}(M)$ on the right; we will use m_2 to denote this action (not to be confused with m here, which means *mirror*). Let $m(b^*_{*;m(h_1),m(h_2)})$ denote either $m(b^*_{\gamma;m(h_1),m(h_2)})$ or $m(b^*_{C;m(h_1),m(h_2)})$. We define

$$m_2(X_{x_i\cdot h_1}, m(b^*_{*;m(h_1),m(h_2)})) := \sum_j \sum_{h'\in\beta_{\text{mult}}} \tilde{c}_{i,j;h'}\tilde{\tilde{c}}_{h'h_1;h_2} X_{x_j\cdot h_2}.$$

If $m(b^*_{*;m(h_1),m(h_2)})=m(b^*_{\gamma;m(h_1),m(h_2)})$, then $\tilde{\tilde{c}}_{h'h_1;h_2}$ is only nonzero for one value of h', namely $h'=(W(a')a, \text{all plus})$, where $h_1=(W(a)b,\sigma)$ and $h_2=(W(a')b,\sigma')$.

For this value of h', $\tilde{\tilde{c}}_{h'h_1;h_2}$ is 1. Thus,

$$m_2(X_{x_i\cdot h_1}, m(b^*_{\gamma;m(h_1),m(h_2)})) = \sum_j \tilde{c}_{i,j;h'} X_{x_j\cdot h_2}. \tag{6-1}$$

If $m(b^*_{*;m(h_1),m(h_2)}) = m(b^*_{C;m(h_1),m(h_2)})$, then $\tilde{\tilde{c}}_{h'h_1;h_2}$ will be nonzero for any h'_α which equals $(W(a)a,$ minus on $\alpha)$, where $h_1 = (W(a)b, \sigma)$ and α is any arc of a which is part of the circle C in $W(a)a$. For h' equal to one of the h'_α, $\tilde{\tilde{c}}_{h'h_1;h_2}$ is 1, and for all other h', $\tilde{\tilde{c}}_{h'h_1;h_2}$ is zero. Thus,

$$m_2(X_{x_i\cdot h_1}, m(b^*_{C;m(h_1),m(h_2)})) = \sum_j \sum_{\text{left arcs } \alpha \text{ of } C} \tilde{c}_{i,j;h'_\alpha} X_{x_j\cdot h_2}. \tag{6-2}$$

Note that m_2 is bigrading-preserving; this follows from the degree conditions on x_i and x_j in the basis expansion of $d(x_i \cdot h_1)$ given in Section 6.1 above.

We then extend m_2 to an action of $m(\mathcal{B})^!$ on $\widehat{A}(M)$ by imposing the associativity relation

$$m_2 \circ (\mathrm{id} \otimes \mu_2) := m_2 \circ (m_2 \otimes \mathrm{id}),$$

where μ_2 is the algebra multiplication on $m(\mathcal{B})^!$. Below we will verify that this algebra action is well-defined. Finally, $\widehat{A}(M)$ has a differential m_1 given by

$$m_1(X_{x_i\cdot h_1}) = \sum_j c_{i,j} X_{x_j\cdot h_1}.$$

6.2.2 Proposition *The action of $m(\mathcal{B})^!$ on $\widehat{A}(M)$ is well-defined and associative. Thus, $\widehat{A}(M)$ is a right module over $m(\mathcal{B})^!$.*

Proof The action is associative by definition, once we show that it is well-defined. We may write $\mathcal{B}^!$ as $T(V^*_{\mathcal{B}})/I^\perp$; thus,

$$m(\mathcal{B})^! = T(m(V^*_{\mathcal{B}}))/m(\mathcal{I}^\perp),$$

where the mirrors of the $\mathcal{I}_\beta$–bimodules $V^*_{\mathcal{B}}$ and $\mathcal{I}^\perp$ are defined as in Definition 5.2.4. Now, Definition 6.2.1, extended by associativity, gives us a map

$$\widehat{A}(M) \otimes_{\mathcal{I}_\beta} T(m(V^*_{\mathcal{B}})) \to \widehat{A}(M).$$

We want to show that if $m(r^*)$ is a generator of $m(I^\perp)$, then multiplying any $X_{x_i\cdot h_1}$ by $m(r^*)$ gives zero.

The generators $m(r^*)$ of $m(I^\perp)$ are quadratic in the $m(b^*_{*;m(h_1),m(h_2)})$ and they have intrinsic degree either 1, $\frac{3}{2}$ or 2. For those $m(r^*)$ of intrinsic degree 2, the equations in item (5) of Proposition 6.1.2 above imply that $m(r^*)$ acts as zero on any $X_{x_i\cdot h_1}$.

For those $m(r^*)$ of intrinsic degree $\frac{3}{2}$, the equations in item (4) of Proposition 6.1.2 similarly imply that $m(r^*)$ acts as zero on $\widehat{A}(M)$.

The generators $m(r^*)$ of $m(I^\perp)$ which have intrinsic degree 1 are sums of either one, two, three or four terms $m(b^*_\gamma)m(b^*_{\gamma'})$ with all coefficients $+1$. For a fixed $m(r^*)$, let $m(h_1)\in\mathcal{I}_\beta$ denote its left idempotent and let $m(h_3)\in\mathcal{I}_\beta$ denote its right idempotent. The element h_3 of β has degree 2 more than h_1, as elements of H^n with its intrinsic grading, and h_3 differs from h_1 by two surgeries on its left crossingless matching. In particular, the left crossingless matchings of h_1 and h_3 are different; this follows from inspection of the generators $m(r^*)$ of intrinsic degree 1 which actually appear in $m(I^\perp)$. Monomials of the form $m(b^*_\gamma)m(b^*_{\gamma^\dagger})$ do not appear as terms of these generators.

For any generators of $\widehat{A}(M)$ of the form $X_{x_i\cdot h_1}$ and $X_{x_k\cdot h_3}$, where h_1 and h_3 are as above, with $\deg x_k=\deg x_i+(-2,2)$ as elements of M, the equations from item (3) of Proposition 6.1.2 become

$$\sum_{\substack{j,h'\in\beta_\gamma,\\ h''\in\beta_\gamma,h_2\in\beta}} \tilde{c}_{i,j;h'}\tilde{\tilde{c}}_{h'h_1;h_2}\tilde{c}_{j,k;h''}\tilde{\tilde{c}}_{h''h_2;h_3}=0;$$

the terms involving $h'\in\beta_\alpha$ vanish for these choices of h_1 and h_3. These equations imply that all generators $m(r^*)$ of $m(I^\perp)$ of intrinsic degree 1 act as zero on $\widehat{A}(M)$. Thus, the algebra action m_2 of $m(\mathcal{B})^!$ on $\widehat{A}(M)$ is well-defined. □

6.2.3 Proposition *The differential m_1 on $\widehat{A}(M)$ satisfies $m_1^2=0$, and the Leibniz rule*

$$m_1\circ m_2=m_2\circ(m_1\otimes|\mathrm{id}|)+m_2\circ(\mathrm{id}\otimes\mu_1)$$

is satisfied, where μ_1 is the differential on $m(\mathcal{B})^!$. Thus, $\widehat{A}(M)$ is a differential bigraded right module over $m(\mathcal{B})^!$ and hence a type A structure over $m(\mathcal{B})^!$.

Proof First, $m_1^2=0$ by the equations in item (1) of Proposition 6.1.2.

We want to show that the Leibniz rule is satisfied for $\widehat{A}(M)$. Since the action of $m(\mathcal{B})^!$ on $\widehat{A}(M)$ is associative, and μ_1 satisfies its own Leibniz rule, it suffices to show that

$$m_1\circ m_2(X_{x_i\cdot h_1},m(b^*_{\gamma;m(h_1),m(h_3)}))=-m_2\circ(m_1(X_{x_i\cdot h_1})\otimes m(b^*_{\gamma;m(h_1),m(h_3)}))$$

and

$$\begin{aligned}&m_1\circ m_2(X_{x_i\cdot h_1},m(b^*_{C;m(h_1),m(h_3)}))\\&=-m_2\circ(m_1(X_{x_i\cdot h_1})\otimes m(b^*_{C;m(h_1),m(h_3)}))+m_2(X_{x_i\cdot h_1}\otimes\mu_1(m(b^*_{C;m(h_1),m(h_3)}))).\end{aligned}$$

Note that in the relation involving $m(b^*_\gamma)$, the μ_1 term vanishes.

The first of these two equations follows from item (2) of Proposition 6.1.2. For the second equation, note that for a fixed $h_1 \in \beta$, the only h_3 such that $m(b^*_{C;m(h_1),m(h_3)})$ is a generator of $m(\mathcal{B})^!$ are those $h_3 \in \beta$ which differ only from h_1 by changing the sign of one circle C from plus to minus. We have

$$\mu_1(m(b^*_{C;m(h_1),m(h_3)})) = -\sum_{h_2 \in \beta} m(b^*_{\gamma;m(h_1),m(h_2)}) m(b^*_{\gamma^\dagger;m(h_2),m(h_3)}),$$

where the sum is implicitly over those $h_2 \in \beta$ such that the generators $m(b^*_{\gamma;m(h_1),m(h_2)})$ and $m(b^*_{\gamma^\dagger;m(h_2),m(h_3)})$ exist.

For such h_1 and h_3, consider generators $X_{x_i \cdot h_1}$ and $X_{x_k \cdot h_3}$ such that $\deg x_k = \deg x_i + (-2, 2)$; these are the only $X_{x_k \cdot h_3}$ which may appear in the basis expansion of the left or right side of the second equation above. Applying the equations in item (3) of Proposition 6.1.2 to $X_{x_i \cdot h}$ and $X_{x_k \cdot h_3}$, we see that the second equation above holds. Thus, the Leibniz rule on $\widehat{A}(M)_{m(\mathcal{B})^!}$ is satisfied. □

Now we formally add actions of $\mathcal{B}$ to $\widehat{A}(M)$, to make it a type A structure over $m(\mathcal{B})^! \odot \mathcal{B}$ rather than just over $m(\mathcal{B})^!$.

6.2.4 Definition The type A operation m_2 on $\widehat{A}(M)^{m(\mathcal{B})^! \odot \mathcal{B}}$ is defined as in Definition 6.2.1 on the generators of $m(\mathcal{B})^!$. On the generators of $\mathcal{B}$, it is defined by

$$m_2(X_{x_i \cdot h_1}, b_{\gamma;h_1,h_2}) := X_{x_i \cdot h_2} \quad \text{and} \quad m_2(X_{x_i \cdot h_1}, b_{C;h_1,h_2}) := X_{x_i \cdot h_2}.$$

Note that these actions are bigrading-preserving. To define the action of an arbitrary element of $m(\mathcal{B})^! \odot \mathcal{B}$ on $\widehat{A}(M)$, we impose associativity of the action. Below we check that this definition respects the relations on $m(\mathcal{B})^! \odot \mathcal{B}$.

6.2.5 Proposition *The action m_2 of $m(\mathcal{B})^! \odot \mathcal{B}$ on $\widehat{A}(M)$ is well-defined; with this action and the differential $m_1\colon \widehat{A}(M) \to \widehat{A}(M)$ from Definition 6.2.1, $\widehat{A}(M)$ is a differential bigraded module (hence type A structure) over $m(\mathcal{B})^! \odot \mathcal{B}$.*

Proof First we need to check that

$$m_2\colon \widehat{A}(M) \otimes_{\mathcal{I}_\beta} (m(\mathcal{B})^! \odot \mathcal{B}) \to \widehat{A}(M)$$

is well-defined. Recall that the relation ideal J_{full} of $m(\mathcal{B})^! \odot \mathcal{B}$ was defined to be

$$\begin{aligned} J_{\text{full}} := T(V_{\text{full}}) \cdot \big((J_{\mathcal{B}} \cap (T^1(V_{\mathcal{B}}) \oplus T^2(V_{\mathcal{B}}))) \\ \oplus (J_{m(\mathcal{B})^!} \cap (T^1(V_{m(\mathcal{B})^!}) \oplus T^2(V_{m(\mathcal{B})^!}))) \oplus J_{\text{extra}} \big) \cdot T(V_{\text{full}}). \end{aligned}$$

By Proposition 6.2.2, the generators of $J_{m(\mathcal{B})^!} \cap (T^1(V_{m(\mathcal{B})^!}) \oplus T^2(V_{m(\mathcal{B})^!}))$ act as zero on $\widehat{A}(M)$. It is immediate from the definition of the action of $\mathcal{B}$ on $\widehat{A}(M)$ that the generators of $J_{\mathcal{B}} \cap (T^1(V_{\mathcal{B}}) \oplus T^2(V_{\mathcal{B}}))$ act as zero on $\widehat{A}(M)$.

Thus, to show that m_2 is well-defined, it remains to show that the generators of J_{extra} act as zero on $\widehat{A}(M)$. These generators are listed in items (1)–(5) of Definition 5.3.2.

- Consider a relation

$$b_{\gamma;h_1,h_2} m(b^*_{\eta';m(h_2),m(h_3)}) - m(b^*_{\eta;m(h_1),m(\tilde{h}_2)}) b_{\gamma';\tilde{h}_2,h_3}$$

from item (1) of Definition 5.3.2. Write $h_1 = (W(a_1)b_1, \sigma_1)$ and let $X_{x_i \cdot h_1}$ be a generator of $\widehat{A}(M)$. Multiplying $X_{x_i \cdot h_1}$ by $b_{\gamma;h_1,h_2}$, we get $X_{x_i \cdot h_2}$ where $h_2 = (W(a_1)b_2, \sigma_2)$. Multiplying $X_{x_i \cdot h_2}$ by $m(b^*_{\eta';m(h_2),m(h_3)})$, with $h_3 = (W(a_2)b_2, \sigma_3)$, by (6-1) we get

$$\sum_j \tilde{c}_{i,j;h'} X_{x_j \cdot h_3},$$

where $h' \in \beta_{\mathrm{mult}}$ is $(W(a_2)a_1, \text{all plus})$.

On the other hand, if we first multiply $X_{x_i \cdot h_1}$ by $m(b^*_{\eta;m(h_1),m(\tilde{h}_2)})$, by (6-1) we get

$$\sum_j \tilde{c}_{i,j;h'} X_{x_j \cdot \tilde{h}_2},$$

where h' is also $(W(a_2)a_1, \text{all plus})$.

If we multiply this result by $b_{\gamma';\tilde{h}_2,h_3}$, we get

$$\sum_j \tilde{c}_{i,j;h'} X_{x_j \cdot h_3}.$$

Thus, generators of J_{extra} from item (1) of Definition 5.3.2 act as zero on $\widehat{A}(M)$.

- For relations

$$b_{C;h_1,h_2} m(b^*_{\eta;m(h_2),m(h_3)}) - m(b^*_{\eta;m(h_1),m(\tilde{h}_2)}) b_{C;\tilde{h}_2,h_3}$$

from item (2) of Definition 5.3.2, the argument is essentially the same. If $h_1 = (W(a_1)b_1, \sigma_1)$ and $\tilde{h}_2 = (W(a_2)b_1, \sigma_2')$, then h' is again $(W(a_2)a_1, \text{ all plus })$. We still use (6-1).

- Consider a relation

$$b_{\gamma;h_1,h_2} m(b^*_{C;m(h_2),m(h_3)}) - m(b^*_{C;m(h_1),m(\tilde{h}_2)}) b_{\gamma;\tilde{h}_2,h_3}$$

from item (3) of Definition 5.3.2. Let $X_{x_i \cdot h_1}$ be a generator of $\widehat{A}(M)$ and write $h_1 = (W(a_1)b_1, \sigma_1)$. Multiplying $X_{x_i \cdot h_1}$ by $b_{\gamma;h_1,h_2}$, we get $X_{x_i \cdot h_2}$ where $h_2 = (W(a_1)b_2, \sigma_2)$. Multiplying $X_{x_i \cdot h_2}$ by $m(b^*_{C;m(h_2),m(h_3)})$, with $h_3 = (W(a_1)b_2, \sigma_3)$, by (6-2) we get

$$\sum_j \sum_{\text{left arcs } \alpha \text{ of } C} \tilde{c}_{i,j;h'_\alpha} X_{x_j \cdot h_3},$$

where h'_α is $(W(a_1)a_1,\ \text{minus on } \alpha)$.

On the other hand, if we first multiply $X_{x_i \cdot h_1}$ by $m(b^*_{C;m(h_1),m(\tilde{h}_2)})$, by (6-2) we get

$$\sum_j \sum_{\text{left arcs } \alpha \text{ of } C} \tilde{c}_{i,j;h'_\alpha} X_{x_j \cdot \tilde{h}_2},$$

where h'_α is again $(W(a_1)a_1,\ \text{minus on } \alpha)$.

If we multiply this result by $b_{\gamma;\tilde{h}_2,h_3}$, we get

$$\sum_j \sum_{\text{left arcs } \alpha \text{ of } C} \tilde{c}_{i,j;h'_\alpha} X_{x_j \cdot h_3}.$$

Thus, generators of J_{extra} from item (3) of Definition 5.3.2 act as zero on $\widehat{A}(M)$.

- For relations

$$b_{C;h_1,h_2} m(b^*_{C';m(h_2),m(h_3)}) - m(b^*_{C';m(h_1),m(\tilde{h}_2)}) b_{C;\tilde{h}_2,h_3}$$

from item (4) of Definition 5.3.2, the argument is the same as for relations from item (3).

- Finally, consider a relation

$$b_{\gamma;h_1,h_2} m(b^*_{C;m(h_2),m(h_3)}) - m(b^*_{C';m(h_1),m(h'_2)}) b_{\gamma;h'_2,h_3} - m(b^*_{C'';m(h_1),m(\tilde{\tilde{h}}_2)}) b_{\gamma;\tilde{\tilde{h}}_2,h_3}$$

from item (5) of Definition 5.3.2, in the case where γ joins two circles C' and C'' to produce C. Write $h_1 = (W(a_1)b_1, \sigma_1)$ and let $X_{x_i \cdot h_1}$ be a generator of $\widehat{A}(M)$. Multiplying $X_{x_i \cdot h_1}$ by $b_{\gamma;h_1,h_2}$, we get $X_{x_i \cdot h_2}$ where $h_2 = (W(a_1)b_2, \sigma_2)$. Multiplying $X_{x_i \cdot h_2}$ by $m(b^*_{C;m(h_2),m(h_3)})$, with $h_3 = (W(a_1)b_2, \sigma_3)$, by (6-2) we get

$$\sum_j \sum_{\text{left arcs } \alpha \text{ of } C} \tilde{c}_{i,j;h'_\alpha} X_{x_j \cdot h_3},$$

where h'_α is $(W(a_1)a_1,\ \text{minus on } \alpha)$.

On the other hand, if we first multiply $X_{x_i \cdot h_1}$ by $m(b^*_{C';m(h_1),m(\tilde{h}_2)})$, we get

$$\sum_j \sum_{\text{left arcs } \alpha \text{ of } C'} \tilde{c}_{i,j;h'_\alpha} X_{x_j \cdot \tilde{h}_2},$$

where h'_α is again $(W(a_1)a_1,\ \text{minus on } \alpha)$. Multiplying by $b_{\gamma';\tilde{h}_2,h_3}$, we get

$$\sum_j \sum_{\text{left arcs } \alpha \text{ of } C'} \tilde{c}_{i,j;h'_\alpha} X_{x_j \cdot h_3}.$$

Finally, if we multiply $X_{x_i \cdot h_1}$ by $m(b^*_{C'';m(h_1),m(\tilde{\tilde{h}}_2)})$, we get

$$\sum_j \sum_{\text{left arcs } \alpha \text{ of } C''} \tilde{c}_{i,j;h'_\alpha} X_{x_j \cdot \tilde{\tilde{h}}_2}.$$

Multiplying this result by $b_{\gamma;\tilde{\tilde{h}}_2,h_3}$, we get

$$\sum_j \sum_{\text{left arcs } \alpha \text{ of } C''} \tilde{c}_{i,j;h'_\alpha} X_{x_j \cdot h_3}.$$

Now, since γ was assumed to join the circles C' and C'' to produce C via a bridge on the right side of $\{0\} \times \mathbb{R}$, the set of left arcs α of C is the disjoint union of the sets of left arcs of C' and C''. Thus, the relation

$$b_{\gamma;h_1,h_2} m(b^*_{C;m(h_2),m(h_3)}) - m(b^*_{C';m(h_1),m(\tilde{h}_2)}) b_{\gamma;\tilde{h}_2,h_3} - m(b^*_{C'';m(h_1),m(\tilde{\tilde{h}}_2)}) b_{\gamma;\tilde{\tilde{h}}_2,h_3}$$

acts as zero on $\widehat{A}(M)$. For relations of the form

$$b_{\gamma;h_1,\tilde{h}_2} m(b^*_{C';m(\tilde{h}_2),m(h_3)}) + b_{\gamma;h_1,\tilde{\tilde{h}}_2} m(b^*_{C'';m(\tilde{\tilde{h}}_2),m(h_3)}) \\ - m(b^*_{C;m(h_1),m(h_2)}) b_{\gamma;h_2,h_3},$$

where γ splits C into C' and C'', the argument is analogous. Thus, generators of J_{extra} from item (5) of Definition 5.3.2 act as zero on $\widehat{A}(M)$.

At this point, we have shown that the action m_2 of $m(\mathcal{B})^! \odot \mathcal{B}$ on $\widehat{A}(M)$ is well-defined. It is associative by definition. To show that the Leibniz rule is satisfied, it suffices by associativity to check it on the generators of $\mathcal{B}$ and of $m(\mathcal{B})^!$, and we have already done this for the generators of $m(\mathcal{B})^!$ in Proposition 6.2.2.

Let $X_{x_i \cdot h_1}$ be a generator of $\widehat{A}(M)$ and let $b_{\gamma;h_1,h_2}$ be a generator of $\mathcal{B}$. Then

$$m_1 \circ m_2(X_{x_i \cdot h_1}, b_{\gamma;h_1,h_2}) = m_1(X_{x_i \cdot h_2}) = \sum_j c_{i,j} X_{x_j \cdot h_2},$$

while

$$m_2(m_1 \otimes |\mathrm{id}|)(X_{x_i \cdot h_1}, b_{\gamma;h_1,h_2}) = m_2\Bigl(\sum_j c_{i,j} X_{x_j \cdot h_1}, b_{\gamma;h_1,h_2}\Bigr) = \sum_j c_{i,j} X_{x_j \cdot h_2},$$

and the $m_2 \circ (\mathrm{id} \otimes \mu_1)$ term is zero because $\mu_1 = 0$ on $\mathcal{B}$. The argument is unchanged for generators $b_{C;h_1,h_2}$. Thus, the Leibniz rule

$$m_1 \circ m_2 = m_2 \circ (m_1 \otimes |\mathrm{id}|) + m_2 \circ (\mathrm{id} \otimes \mu_1)$$

holds and $\widehat{A}(M)$ is a differential bigraded right module over $m(\mathcal{B})^! \odot \mathcal{B}$. □

Now we will consider the case $M = [T]^{\mathrm{Kh}}$, where T is a tangle diagram in $\mathbb{R}_{\le 0} \times \mathbb{R}$. Recall from Section 6.1 that $[T]^{\mathrm{Kh}}$ satisfies C_{module}, so we get a type A structure $\widehat{A}([T]^{\mathrm{Kh}})_{m(\mathcal{B})^! \odot \mathcal{B}}$. This type A structure descends to a type A structure over the quotient algebra $\mathcal{B}\Gamma_n$.

6.2.6 Proposition *The extra relations from Proposition 5.2.8 act as zero on the $m(\mathcal{B})^! \odot \mathcal{B}$–module $\widehat{A}([T]^{\mathrm{Kh}})$ defined above in Definition 6.2.4. Thus, $\widehat{A}([T]^{\mathrm{Kh}})$ descends to a differential bigraded right module over the quotient algebra $\mathcal{B}\Gamma_n$ of $m(\mathcal{B})^! \odot \mathcal{B}$ by these relations.*

Proof Since the relations from Proposition 5.2.8 involve only quadratic monomials in the generators $m(b^*_{\gamma;m(h_1),m(h_2)})$ of $m(\mathcal{B})^!$, with no generators from $\mathcal{B}$ appearing, it suffices to show that these relations act as zero on the $m(\mathcal{B})^!$–module $\widehat{A}([T]^{\mathrm{Kh}})$ defined in Definition 6.2.1.

Consider a tetrahedron in the graph G of Proposition 5.2.8, with vertices a, b, c and d as labeled in that proposition. We will show that the relation term $a + c$ acts as zero on $\widehat{A}([T]^{\mathrm{Kh}})$; the proofs for the remaining extra relation terms are exactly analogous.

We may write out

$$\begin{aligned} a &= m(b^*_{\gamma;m(h_1),m(h_2)}) m(b^*_{\eta';m(h_2),m(h_3)}), \\ c &= m(b^*_{\eta;m(h_1),m(\tilde{\tilde{h}}_2)}) m(b^*_{\gamma';m(\tilde{\tilde{h}}_2),m(h_3)}), \end{aligned}$$

as in Proposition 5.2.8. Suppose we have two generators of $\widehat{A}([T]^{\mathrm{Kh}})$ of the form $X_{x_{00} \cdot h_1}$ and $X_{x_{11} \cdot h_3}$. Here, T may have more than two crossings, but for two designated crossings, x_{00} has the zero-resolution at both and x_{11} has the one-resolution at both (and x_{00} and x_{11} agree at all other crossings). We assume that changing x_{00} to x_{10} has the effect of surgery on γ, while changing x_{00} to x_{01} has the effect of surgery on η.

To show that $a + c$ acts as zero on $\widehat{A}([T]^{\mathrm{Kh}})$, it suffices to show $m_2(X_{x_{00} \cdot h_1}, a + c)$ has zero coefficient on $X_{x_{11} \cdot h_3}$. We can compute the coefficient of $m_2(X_{x_{00} \cdot h_1}, a)$

and $m_2(X_{x_{00}\cdot h_1}, c)$ on $X_{x_{11}\cdot h_3}$ using associativity and (6-1). We have, ignoring terms which do not contribute to a coefficient on $X_{x_{11}\cdot h_3}$,

$$\begin{aligned} m_2(X_{x_{00}\cdot h_1}, a) &= \tilde{c}_{00,10;h'} m_2(X_{x_{10}\cdot h_2}, m(b^*_{\eta';m(h_2),m(h_3)})) \\ &= \tilde{c}_{00,10;h'} \tilde{c}_{10,11;h''} X_{x_{11}\cdot h_3} \end{aligned}$$

for two uniquely determined elements h' and h'' of β_γ. Similarly, the coefficient of $m_2(X_{x_{00}\cdot h_1}, c)$ on $X_{x_{11}\cdot h_3}$ is $\tilde{c}_{00,01;h'''}\tilde{c}_{01,11;h''''}$, for two further uniquely defined elements h''' and h'''' of β_γ.

By item (4) of Example 6.1.3, we have the following coefficients:

- $\tilde{c}_{00,10;h'} = (-1)^{\#_1(00,10)}$,
- $\tilde{c}_{10,11;h''} = (-1)^{\#_1(10,11)}$,
- $\tilde{c}_{00,01;h'''} = (-1)^{\#_1(00,01)}$,
- $\tilde{c}_{01,11;h''''} = (-1)^{\#_1(01,11)}$.

Recall that $\#_1(i,j)$ denotes the number of 1–resolutions of crossings in x_i among those crossings which occur earlier than the changed crossing (going from x_i to x_j) in the ordering on crossings of T (which is implicitly assumed, as usual, to be part of the choice of T).

Since

$$(-1)^{\#_1(00,10)}(-1)^{\#_1(10,11)} + (-1)^{\#_1(00,01)}(-1)^{\#_1(01,11)} = 0,$$

we can conclude that the coefficient of $m_2(X_{x_{00}\cdot h_1}, a+c)$ on $X_{x_{11}\cdot h_3}$ is zero, for all possible pairs $X_{x_{00}\cdot h_1}$ and $X_{x_{11}\cdot h_3}$. Thus, the extra relation terms of Proposition 5.2.8 act as zero on $\widehat{A}([T]^{\mathrm{Kh}})$. □

6.2.7 Proposition *Roberts' type A structure from* [11] *agrees with* $\widehat{A}([T]^{\mathrm{Kh}})$*, the module over* $\mathcal{B}\Gamma_n$ *constructed in Proposition 6.2.6.*

Proof First, $\widehat{A}([T]^{\mathrm{Kh}})$ has the same $\mathbb{Z}$–basis, with the same bigradings and action of the idempotent ring $\mathcal{I}_\beta$, as Roberts' type A structure. We can use the data of Example 6.1.3 to check that the differentials m_1 agree and that the algebra actions m_2 agree under the identification of $\mathcal{B}\Gamma_n$ with a quotient of $m(\mathcal{B})^! \odot \mathcal{B}$.

For the differentials, we have $m_1(X_{x_i\cdot h}) = \sum_j (-1)^{\#_1(i,j)} X_{x_j\cdot h}$, where the sum is over those x_j related to x_i by crossing changes from items (1) or (2) of Example 6.1.3. This formula also gives Roberts' differential $m_1 = d_{\mathrm{APS}}$ as specified in [11, Section 3.3].

It suffices to check that the algebra actions m_2 agree when multiplying by the generators of $\mathcal{B}$ and $m(\mathcal{B})^!$. First, for a generator $X_{x_i \cdot h_1}$ of $\widehat{A}([T]^{\mathrm{Kh}})$ and a generator $b_{\gamma;h_1,h_2}$ of $\mathcal{B}$, we have

$$m_2(X_{x_i \cdot h_1}, b_{\gamma;h_1,h_2}) = X_{x_i \cdot h_2},$$

agreeing with Roberts' definition of the action of $\overrightarrow{e}_{\gamma;h_1,h_2}$ in item 5 of his definition of m_2 [11, Section 4]. Similarly, our action of $b_{C;h_1,h_2}$ is the same as Roberts' action of $\overrightarrow{e}_{C;h_1,h_2}$, defined in item 2 of his definition of m_2.

For a generator $m(b^*_{\gamma;m(h_1),m(h_2)})$ of $m(\mathcal{B})^!$, we have

$$m_2(X_{x_i \cdot h_1}, m(b^*_{\gamma;m(h_1),m(h_2)})) = \sum_j \tilde{c}_{i,j;h'} X_{x_j \cdot h_2},$$

where if $h_1 = (W(a_1)b_1, \sigma_1)$ and $h_2 = (W(a_2)b_1, \sigma_2)$, then $h' = (W(a_2)a_1, \text{all plus})$, and the coefficient $\tilde{c}_{i,j;h'}$ equals zero or $(-1)^{\#_1(i,j)}$ according to Example 6.1.3(4). Thus,

$$m_2(X_{x_i \cdot h_1}, m(b^*_{\gamma;m(h_1),m(h_2)})) = \sum_j (-1)^{\#_1(i,j)} X_{x_j \cdot h_2},$$

where the sum is over the subset of j making $\tilde{c}_{i,j;h'}$ nonzero. This algebra action agrees with the action of $\overleftarrow{e}_{\gamma;h_1,h_2}$ as defined in item 4 of Roberts' definition of m_2.

Finally, for a generator $m(b^*_{C;m(h_1),m(h_2)})$ of $m(\mathcal{B})^!$, we have

$$m_2(X_{x_i \cdot h_1}, m(b^*_{C;m(h_1),m(h_2)})) = \sum_j \sum_{\text{left arcs } \alpha \text{ of } C} \tilde{c}_{i,j;h'_\alpha} X_{x_j \cdot h_2},$$

where if $h_1 = (W(a)b, \sigma_1)$ and $h_2 = (W(a)b, \sigma_2)$, then $h'_\alpha = (W(a)a, \text{ minus on } \alpha)$, and the coefficient $\tilde{c}_{i,j;h'}$ equals zero or $(-1)^{\#_1(i,j)}$ according to item (3) above. Thus,

$$m_2(X_{x_i \cdot h_1}, m(b^*_{C;m(h_1),m(h_2)})) = \sum_j (-1)^{\#_1(i,j)} X_{x_j \cdot h_2},$$

where the sum is over the subset of j making some $\tilde{c}_{i,j;h'_\alpha}$ nonzero (note that, given such j, the element h'_α is uniquely determined). This algebra action agrees with the action of $\overleftarrow{e}_{C;h_1,h_2}$ as defined in item 3 of Roberts' definition of m_2. □

6.3 Type D structures

Given a chain complex M of projective graded right H^n–modules satisfying $\mathcal{C}_{\mathrm{module}}$, in Definition 6.2.4 we defined a type A structure $\widehat{A}(M)$ over $m(\mathcal{B})^! \odot \mathcal{B}$. For a tangle diagram T, $\widehat{A}([T]^{\mathrm{Kh}})$ descends to a type A structure over $\mathcal{B}\Gamma_n$ by Proposition 6.2.6, which agrees with Roberts' type A structure by Proposition 6.2.7.

Now suppose N is a chain complex of graded projective left H^n–modules. We want to define condition C_{module} for left modules, and for N satisfying this condition, we want to define a type D structure $\widehat{D}(N)$ over $m(\mathcal{B})^! \odot \mathcal{B}$.

To do this, we will first define an operation called *reflection* which, when applied to N, yields a complex of right modules.

6.3.1 Definition Let N be a chain complex of projective graded left H^n–modules. Viewing N as a differential bigraded projective left H^n–module, write

$$N = \bigoplus_i H^n \cdot y_i[j_i, k_i],$$

with H^n acting by left multiplication. Then we may define a differential bigraded projective right H^n–module $r(N)$, called the *reflection* of N, as

$$r(N) := \bigoplus_i r(y_i)[j_i, k_i] \cdot H^n,$$

where the $r(y_i)$ are formal reflections of the y_i, with the same idempotents as the y_i, and H^n acts by right multiplication. Let refl denote the map from N to $r(N)$ such that

$$\text{refl}(h \cdot y_i) = r(y_i) \cdot m(h),$$

where $m(h)$ is defined as in Example 5.2.5. The inverse of refl: $N \to r(N)$ is refl: $r(N) \to r(r(N)) = N$ (using a simple generalization of the above definition which reflects right modules to left modules rather than left modules to right modules). If m_1 denotes the differential on N, then the differential on $r(N)$ is $\text{refl} \circ m_1 \circ \text{refl}$.

6.3.2 Remark Although the geometric content of both Definition 6.3.1 and (variants of) Definition 5.2.4 is just the reflection across the line $\{0\} \times \mathbb{R}$, the algebraic consequences of this reflection are different in Definition 6.3.1. Whereas in Definition 5.2.4, left modules remain left modules and right modules remain right modules under mirroring, in Definition 6.3.1 left modules are sent to right modules and vice-versa.

6.3.3 Definition A chain complex N of projective graded left H^n–modules satisfies the condition C_{module} for a generating set $\{y_i\}$ if and only if $r(N)$ satisfies the condition C_{module} as defined in Definition 6.1.1 for the generating set $\{r(y_i)\}$.

If N satisfies C_{module}, then we can take the box tensor product of $\widehat{A}(r(N))$ and the type DD bimodule ${}^{m(\mathcal{B})^! \odot \mathcal{B}}K^{m(m(\mathcal{B})^! \odot \mathcal{B})^{\text{op}}}$ to get a (left) type D structure over $m(m(\mathcal{B})^! \odot \mathcal{B})$. Below we define this tensor product precisely. It is a slight modification of Definition 2.3.5; we will not give the definition in the fullest possible generality. See [7, Definition 2.3.9] for a more general definition using $\mathbb{Z}/2\mathbb{Z}$ coefficients.

6.3.4 Definition Let $\mathcal{B}$ be a differential bigraded algebra over an idempotent ring $\mathcal{I}$. Let $\widehat{A}$ be a differential bigraded right module over $\mathcal{B}$. Assume $\widehat{A}$ is free as a $\mathbb{Z}$–module, with a $\mathbb{Z}$–basis consisting of elements which are grading-homogeneous and have a unique right idempotent.

Let $\mathcal{B}'$ be another differential bigraded algebra over $\mathcal{I}$ and let $\widehat{DD}$ be a rank-one type DD bimodule over $\mathcal{B}$ and $\mathcal{B}'$ with DD operation $\delta_{DD}\colon \mathcal{I} \to \mathcal{B} \otimes_{\mathcal{I}} (\mathcal{B}')^{\mathrm{op}}$. The type D structure $\widehat{A} \boxtimes \widehat{DD}$ over $\mathcal{B}'$, as a $\mathbb{Z}$–module, is

$$\widehat{A} \boxtimes \widehat{DD} := \widehat{A} \otimes_{\mathcal{I}} \widehat{DD} = \widehat{A} \otimes_{\mathcal{I}} \mathcal{I} = \widehat{A}.$$

The idempotent ring $\mathcal{I}$ has a right action on $\widehat{A}$, which we will view instead as a left action (since $\mathcal{I}$ is commutative, we may view right actions as left actions and vice-versa). Since $\widehat{DD}$ is a rank-one DD bimodule, the left and right actions of $\mathcal{I}$ on $\widehat{DD}$ are the same. There is a bigrading on $\widehat{A} \boxtimes \widehat{DD}$ inherited from that on $\widehat{A}$ (recall that $\widehat{DD}$ is contained in bigrading $(0,0)$).

The type D operation $\delta^{\boxtimes}\colon \widehat{A} \boxtimes \widehat{DD} \to \mathcal{B}' \otimes_{\mathcal{I}} (\widehat{A} \boxtimes \widehat{DD})$ is defined by

$$\delta^{\boxtimes} := 1 \otimes m_1 + \xi \circ (m_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD})\colon \widehat{A} \to \mathcal{B}' \otimes_{\mathcal{I}} \widehat{A},$$

where m_1 and m_2 are the type A operations on $\widehat{A}$, and $\xi\colon \widehat{A} \otimes_{\mathcal{I}} (\mathcal{B}')^{\mathrm{op}} \to \mathcal{B}' \otimes_{\mathcal{I}} \widehat{A}$ is defined by

$$\xi(X \otimes (b')^{\mathrm{op}}) := (-1)^{(\deg_h X)(\deg_h b')} b' \otimes X.$$

More precisely, the second summand is the composition

$$\widehat{A} \xrightarrow{\mathrm{id}\otimes\delta_{DD}} \widehat{A} \otimes_{\mathcal{I}} \mathcal{B} \otimes_{\mathcal{I}} (\mathcal{B}')^{\mathrm{op}} \xrightarrow{m_2\otimes\mathrm{id}} \widehat{A} \otimes_{\mathcal{I}} (\mathcal{B}')^{\mathrm{op}} \xrightarrow{\xi} \mathcal{B}' \otimes_{\mathcal{I}} \widehat{A}.$$

The map $\delta^{\boxtimes}$ has bidegree $(0,+1)$.

6.3.5 Proposition *$(\widehat{A} \boxtimes \widehat{DD}, \delta^{\boxtimes})$ is a well-defined type D structure over $\mathcal{B}'$.*

Proof First, since $\widehat{A}$ was assumed to have a $\mathbb{Z}$–basis $\{X_i\}$, with each X_i grading-homogeneous and having unique idempotents, the same is true for $\widehat{A} \boxtimes \widehat{DD} \cong \widehat{A}$.

To verify the type D structure relations, we must show that

$$(\mu_1 \otimes |\mathrm{id}|) \circ \delta^{\boxtimes} + (\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta^{\boxtimes}) \circ \delta^{\boxtimes} = 0.$$

Substituting in the definition of $\delta^{\boxtimes}$ and simplifying some terms, we want to show that

$$\begin{aligned}(6\text{-}3)\quad &(\mu_1\otimes|\mathrm{id}|)\circ\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta_{DD})\\
&+(\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\xi)\circ(\mathrm{id}\otimes m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\mathrm{id}\otimes\delta_{DD})\circ(1\otimes m_1)\\
&+(\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes 1\otimes m_1)\circ\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta_{DD})\\
&+(\mu_2\otimes\mathrm{id})\circ\big(\mathrm{id}\otimes(\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta_{DD}))\big)\\
&\qquad\circ(\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta_{DD}))=0.\end{aligned}$$

We claim that we may rewrite the final term on the left side of (6-3) as

$$(6\text{-}4)\quad \xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\mu_2\otimes\mu_2)\circ(\mathrm{id}\otimes\sigma)\circ(\mathrm{id}\otimes\mathrm{id}\otimes\delta_{DD}\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta_{DD}),$$

where

$$\sigma\colon \mathcal{B}\otimes\mathcal{B}\otimes(\mathcal{B}')^{\mathrm{op}}\otimes(\mathcal{B}')^{\mathrm{op}}\to\mathcal{B}\otimes\mathcal{B}\otimes(\mathcal{B}')^{\mathrm{op}}\otimes(\mathcal{B}')^{\mathrm{op}}$$

was defined in Definition 4.5.1.

To verify that term (6-4) is equal to the final term of (6-3), let X be a generator of $\widehat{A}$. Write $\delta_{DD}(1)=\sum_i b_i\otimes(b_i')^{\mathrm{op}}$. We have

$$\begin{aligned}&(\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes(\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta_{DD})))\circ(\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta_{DD}))(X)\\
&\qquad=\sum_{i,j}(-1)^{(\deg_h b_i')(\deg_h(Xb_i))+(\deg_h b_j')(\deg_h(Xb_ib_j))}(b_i'b_j')\otimes(Xb_ib_j).\end{aligned}$$

On the other hand, we have

$$\begin{aligned}&\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\mu_2\otimes\mu_2)\circ(\mathrm{id}\otimes\sigma)\circ(\mathrm{id}\otimes\mathrm{id}\otimes\delta_{DD}\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta_{DD})(X)\\
&\qquad=\sum_{i,j}(-1)^{\deg_h b_j\deg_h(b_i')+\deg_h(Xb_ib_j)\deg_h(b_i'b_j')}(b_i'b_j')\otimes(Xb_ib_j).\end{aligned}$$

A direct computation, using the additivity of $\deg_h$ under algebra multiplication, verifies that the signs in these expressions are equal.

Now, we may write term (6-4) as

$$\xi\circ(m_2\otimes\mathrm{id})\circ\big(\mathrm{id}\otimes((\mu_2\otimes\mu_2)\circ\sigma\circ(\mathrm{id}\otimes\delta_{DD}\otimes\mathrm{id})\circ\delta_{DD})\big).$$

Using the type DD bimodule relations for δ, we may replace

$$(\mu_2\otimes\mu_2)\circ\sigma\circ(\mathrm{id}\otimes\delta_{DD}\otimes\mathrm{id})\circ\delta_{DD}$$

with

$$-(\mu_1\otimes|\mathrm{id}|)\circ\delta_{DD}-(\mathrm{id}\otimes\mu_1)\circ\delta_{DD}.$$

Then term (6-4) is equal to

$$-\xi\circ(m_2\otimes\mathrm{id})\circ\big(\mathrm{id}\otimes((\mu_1\otimes|\mathrm{id}|)\circ\delta_{DD})\big)-\xi\circ(m_2\otimes\mathrm{id})\circ\big(\mathrm{id}\otimes((\mathrm{id}\otimes\mu_1)\circ\delta_{DD})\big)$$
$$=-\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\mu_1\otimes|\mathrm{id}|)\circ(\mathrm{id}\otimes\delta_{DD})-\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\mathrm{id}\otimes\mu_1)\circ(\mathrm{id}\otimes\delta_{DD}).$$

The term

$$-\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\mathrm{id}\otimes\mu_1)\circ(\mathrm{id}\otimes\delta_{DD})$$

above cancels the first term

$$(\mu_1\otimes|\mathrm{id}|)\circ\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta_{DD})$$

of the terms in (6-3), whose sum we are trying to show is zero. The remaining terms of (6-3) are, after some simplification,

- $-\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\mu_1\otimes|\mathrm{id}|)\circ(\mathrm{id}\otimes\delta_{DD})$,
- $\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta_{DD})\circ m_1$,
- $(\mathrm{id}\otimes m_1)\circ\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta_{DD})$.

The final of these may be written as

$$\xi\circ(m_1\otimes|\mathrm{id}|)\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta_{DD})=\xi\circ((m_1\circ m_2)\otimes|\mathrm{id}|)\circ(\mathrm{id}\otimes\delta_{DD}).$$

We may use the Leibniz rule on $\widehat{A}$ to replace $m_1\circ m_2$ with $m_2\circ(m_1\otimes|\mathrm{id}|)+m_2\circ(\mathrm{id}\otimes\mu_1)$. Thus, the final of the three remaining terms is equal to

$$\xi\circ(m_2\otimes\mathrm{id})\circ(m_1\otimes|\mathrm{id}|\otimes|\mathrm{id}|)\circ(\mathrm{id}\otimes\delta_{DD})+\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\mu_1\otimes|\mathrm{id}|)\circ(\mathrm{id}\otimes\delta_{DD}).$$

The second of these summands cancels with the first of the other three remaining terms listed above, so it remains to show that

$$\xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta_{DD})\circ m_1+\xi\circ(m_2\otimes\mathrm{id})\circ(m_1\otimes|\mathrm{id}|\otimes|\mathrm{id}|)\circ(\mathrm{id}\otimes\delta_{DD})=0.$$

This follows from the equation

$$(m_1\otimes|\mathrm{id}|\otimes|\mathrm{id}|)\circ(\mathrm{id}\otimes\delta_{DD})=-(\mathrm{id}\otimes\delta_{DD})\circ m_1.$$

Indeed, since all generators of $\widehat{DD}\cong\mathcal{I}$ have bigrading $(0,0)$, the element $\delta(1)$ has homological degree 1. □

Applying this construction to $\widehat{A}=\widehat{A}(r(N))$ with $\widehat{DD}={}^{m(\mathcal{B})^!\odot\mathcal{B}}K^{m(m(\mathcal{B})^!\odot\mathcal{B})^{\mathrm{op}}}$, which is a type DD bimodule over $m(\mathcal{B})^!\odot\mathcal{B}$ and $m(m(\mathcal{B})^!\odot\mathcal{B})$, we get a type D structure $\mathcal{A}(r(N))\boxtimes{}^{m(\mathcal{B})^!\odot\mathcal{B}}K^{m(m(\mathcal{B})^!\odot\mathcal{B})^{\mathrm{op}}}$ over $m(m(\mathcal{B})^!\odot\mathcal{B})$. We can then apply another mirroring operation, analogous to Definition 5.3.1 and in the spirit of Definition 5.2.4, to get a type D structure $\widehat{D}(N)$ over $m(\mathcal{B})^!\odot\mathcal{B}$.

6.3.6 Definition Let $\mathcal{B}$ be a differential bigraded algebra over the idempotent ring $\mathcal{I}_\beta$ and let $(\widehat{D}, \delta)$ be a type D structure over $\mathcal{B}$. The *mirrored type D structure* $(m(\widehat{D}), \delta')$ is defined as follows: as an $\mathcal{I}_\beta$–module, $m(\widehat{D})$ is the mirror of $\widehat{D}$ as defined in Definition 5.2.4. As usual, denote the natural map from $\widehat{D}$ to $m(\widehat{D})$ or $m(\widehat{D})$ to $\widehat{D}$ by mirr. The type D operation on $m(\widehat{D})$ is the following:

$$\delta' = m(\widehat{D}) \xrightarrow{\text{mirr}} \widehat{D} \xrightarrow{\delta} \mathcal{B} \otimes \widehat{D} \xrightarrow{\text{mirr} \otimes \text{mirr}} m(\mathcal{B}) \otimes m(\widehat{D}).$$

6.3.7 Definition Let N be a chain complex of graded projective left H^n–modules satisfying the algebraic condition C_{module} of Definition 6.3.3. The type D structure $\widehat{D}(N)$ over $m(\mathcal{B})^! \odot \mathcal{B}$ is defined to be

$$\widehat{D}(N) := m\big(\widehat{A}(r(N)) \boxtimes {}^{m(\mathcal{B})^! \odot \mathcal{B}} K^{m(m(\mathcal{B})^! \odot \mathcal{B})^{\text{op}}}\big).$$

6.3.8 Definition If N is a chain complex N of graded left projective H^n–modules satisfying C_{module}, the type D structure $\widehat{D}(N)$ over $\mathcal{B}\Gamma_n$ associated to N is induced from the type D structure $\widehat{D}(N)$ over $m(\mathcal{B})^! \odot \mathcal{B}$ defined in Definition 6.3.7, using Proposition 5.3.7.

For convenience, we describe the type D operation δ on $\widehat{D}(N)$ explicitly from the differential d_N on N. Let $\{h_1 \cdot y_i \mid e(y_i) = e_R(h_1)\}$ be the $\mathbb{Z}$–basis for N corresponding to the designated generators $\{y_i\}$ of N. By the condition C_{module}, we may expand $d_N(y_i)$ in this basis as

$$d_N(y_i) = \sum_j c'_{i,j} y_j + \sum_{j, h' \in \beta_{\text{mult}}} \tilde{c}'_{i,j;h'} h' \cdot y_j.$$

Then we have

$$d_N(h_1 \cdot y_i) = \sum_j c'_{i,j} h_1 \cdot y_j + \sum_{j, h' \in \beta_{\text{mult}}, h_2 \in \beta} \tilde{c}'_{i,j;h'} \tilde{\tilde{c}}_{h_1 h'; h_2} h_2 \cdot y_j.$$

We let

$$\{Y_{h_1 \cdot y_i} \mid e(y_i) = e_R(h_1)\}$$

denote the $\mathbb{Z}$–basis of $\widehat{D}(N)$ corresponding to the $\mathbb{Z}$–basis $\{h_1 \cdot y_i\}$ of N.

6.3.9 Proposition *Defining the coefficients $c'_{i,j}$ and $\tilde{c}'_{i,j;h'}$ as above, and $\tilde{\tilde{c}}_{h_1 h'; h_2}$ as in Section 6.1, the type D structure operation δ on $\widehat{D}(N)$ has a basis expansion given by*

$$\delta(Y_{h_1\cdot y_i}) = \sum_j c'_{i,j} Y_{h_1\cdot y_j} + \sum_{j,h'\in\beta_{\mathrm{mult}},h_2\in\beta} \tilde{c}'_{i,j;h'}\tilde{\tilde{c}}_{h_1h';h_2} b_{*;h_1,h_2}\otimes Y_{h_2\cdot y_j} + \sum_{\substack{h_2\in\beta \text{ such that}\\ m(b^*_{*;m(h),m(h_2)}) \text{ exists}}} (-1)^{\deg_h y_i} m(b^*_{*;m(h_1),m(h_2)})\otimes Y_{h_2\cdot y_i},$$

where $b_{*;h_1,h_2}$ *means* $b_{\gamma;h_1,h_2}$ *or* $b_{C;h_1,h_2}$ *as appropriate.*

Proof The operation δ is defined as the mirror of the type D structure operation $\delta^{\boxtimes}$ on $\hat{A}(r(N))\boxtimes{}^{m(\mathcal{B})^!\odot\mathcal{B}}K^{m(m(\mathcal{B})^!\odot\mathcal{B})^{\mathrm{op}}}$, which in turn is defined as

$$\delta^{\boxtimes} := 1\otimes m_1 + \xi\circ(m_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta_{DD}).$$

Here, m_1 is the differential on $\hat{A}(r(N))$ and δ_{DD} is the type DD operation on ${}^{m(\mathcal{B})^!\odot\mathcal{B}}K^{m(m(\mathcal{B})^!\odot\mathcal{B})^{\mathrm{op}}}$. Write an arbitrary generator of $\hat{A}(r(N))$ as $X_{m(y_i)\cdot m(h_1)}$, and we have

$$m_1(X_{m(y_i)\cdot m(h_1)}) = \sum_j c'_{i,j} X_{m(y_j)\cdot m(h_1)},$$

$$m_2(X_{m(y_i)\cdot m(h_1)}, m(b^*_{*;h_1,h_2})) = \sum_{j,h'\in\beta_{\mathrm{mult}}} \tilde{c}'_{i,j;h}\tilde{\tilde{c}}_{h_1h';h_2} X_{m(y_j)\cdot m(h_2)},$$

$$m_2(X_{m(y_i)\cdot m(h_1)}, b_{*;m(h_1),m(h_2)}) = X_{m(y_i)\cdot m(h_2)}.$$

Here, $b_{*;m(h_1),m(h_2)}$ stands for either $b_{\gamma;m(h_1),m(h_2)}$ or $b_{C;m(h_1),m(h_2)}$ as appropriate, and similarly for $m(b^*_{*;h_1,h_2})$. Also note that we have $\tilde{\tilde{c}}_{h_1h';h_2} = \tilde{\tilde{c}}_{m(h')m(h_1);m(h_2)}$.

Thus,

$$\delta^{\boxtimes}(X_{m(y_i)\cdot m(h_1)}) = \sum_j c'_{i,j} X_{m(y_j)\cdot m(h_1)} + \sum_{j,h'\in\beta_{\mathrm{mult}},h_2\in\beta} \tilde{c}'_{i,j;h'}\tilde{\tilde{c}}_{h_1h';h_2} m(b_{*;h_1,h_2})\otimes X_{m(y_j)\cdot m(h_2)} + \sum_{\substack{h_2\in\beta \text{ such that}\\ m(b^*_{*;m(h_1),m(h_2)}) \text{ exists}}} (-1)^{\deg_h y_i} m(m(b^*_{*;m(h_1),m(h_2)}))\otimes X_{m(y_i)\cdot m(h_2)}.$$

Taking the mirror of this formula, we get

$$\delta(Y_{h_1\cdot y_i}) = \sum_j c'_{i,j} Y_{h_1\cdot y_j} + \sum_{j,h'\in\beta_{\mathrm{mult}},h_2\in\beta} \tilde{c}'_{i,j;h'}\tilde{\tilde{c}}_{h_1h';h_2} (b_{*;h_1,h_2})\otimes Y_{h_2\cdot y_j} + \sum_{\substack{h_2\in\beta \text{ such that}\\ m(b^*_{*;m(h_1),m(h_2)}) \text{ exists}}} (-1)^{\deg_h y_i} m(b^*_{*;m(h_1),m(h_2)})\otimes Y_{h_2\cdot y_i}. \qquad \square$$

When N is Khovanov's complex $[T]^{\mathrm{Kh}}$ for a tangle diagram T in $\mathbb{R}_{\geq 0} \times \mathbb{R}$, the induced type D structure $\widehat{D}([T]^{\mathrm{Kh}})$ over $\mathcal{B}\Gamma_n$ is the same as Roberts' type D structure from [12].

6.3.10 Proposition *$\widehat{D}([T]^{\mathrm{Kh}})$, as defined in Definition 6.3.8, agrees with the type D structure over $\mathcal{B}\Gamma_n$ which Roberts associates to T.*

Proof Roberts' type D structure is defined as a bigraded $\mathbb{Z}$–module in Definition 32 of [12]. As such, it agrees with $\widehat{D}([T]^{\mathrm{Kh}})$, and the action of the idempotent ring $\mathcal{I}_\beta$ is the same on both; Roberts defines the action of the idempotent ring at the end of Section 3.2 of [12].

Lastly, the type D operation δ on $\widehat{D}([T]^{\mathrm{Kh}})$ has an explicit form given in Proposition 6.3.9 above. The coefficients $c'_{i,j}$ and $\tilde{c}'_{i,j;h'}$ are either $(-1)^{\#_1(i,j)}$ or zero, just like the coefficients $c_{i,j}$ and $\tilde{c}_{i,j;h'}$ described in Example 6.1.3. Recall that the coefficient $\tilde{\tilde{c}}_{h_1 h';h_2}$ is either one or zero; this was also pointed out in Definition 6.2.1. By comparison, δ agrees with Roberts' type D operation defined at the beginning of [12, Section 5]. □

6.4 Pairing

Let M be a complex of graded projective right H^n–modules and let N be a complex of graded projective left H^n–modules, satisfying the algebraic conditions C_{module} of Definition 6.1.1 and Definition 6.3.3. The natural way to pair M and N and get a chain complex over $\mathbb{Z}$ is to take the tensor product $M \otimes_{H^n} N$. However, we could also use Definition 6.2.4 to construct a type A structure $\widehat{A}(M)$ and use Definition 6.3.7 to construct a type D structure $\widehat{D}(N)$, both over $m(\mathcal{B})^! \odot \mathcal{B}$, and then take their box tensor product. This produces the same chain complex as $M \otimes_{H^n} N$, after a reversal of the intrinsic grading.

6.4.1 Proposition *As differential bigraded $\mathbb{Z}$–modules, $\widehat{A}(M) \boxtimes_{m(\mathcal{B})^! \odot \mathcal{B}} \widehat{D}(N)$ is isomorphic to the module obtained from $M \otimes_{H^n} N$ by multiplying all intrinsic gradings on $M \otimes_{H^n} N$ by -1.*

Proof Let $\{x_i \cdot h_1 \mid e(x_i) = e_L(h_1)\}$ and $\{h_1 \cdot y_i \mid e(y_i) = e_R(h_1)\}$ be the $\mathbb{Z}$–bases for M and N, respectively, corresponding to the sets of designated generators $\{x_i\}$ of M and $\{y_i\}$ of N. Then a $\mathbb{Z}$–basis for $M \otimes_{H^n} N$ is $\{x_i \cdot h_1 \cdot y_j\}$ (we will suppress the idempotent conditions).

Write the differentials on M and N as d_M and d_N. As an element of $M \otimes_{H^n} N$, the differential of $x_i \cdot h_1 \cdot y_j$ is

$$\partial^{\otimes}(x_i \cdot h_1 \cdot y_j) = (-1)^{\deg_h y_j}(d_M(x_i) \cdot h_1 \cdot y_j) + (x_i \cdot h_1 \cdot d_N(y_j)).$$

If we expand out d_M as in Section 6.1 and d_N as in the discussion preceding Proposition 6.3.9, we may write this as

$$(-1)^{\deg_h y_j}\Bigg(\sum_k c_{i,k}(x_k\cdot h_1\cdot y_j)+\sum_{k,h'\in\beta_{\mathrm{mult}},h_2\in\beta}\tilde{c}_{i,k;h'}\tilde{\tilde{c}}_{h'h_1;h_2}(x_k\cdot h_2\cdot y_j)\Bigg) \\ +\sum_l c'_{j,l}(x_i\cdot h_1\cdot y_l)+\sum_{l,h'\in\beta_{\mathrm{mult}},h_2\in\beta}\tilde{c}'_{j,l;h'}\tilde{\tilde{c}}_{h_1h';h_2}(x_i\cdot h_2\cdot y_l). \tag{6-5}$$

Now, as a bigraded $\mathbb{Z}$–module, $\widehat{A}(M)\boxtimes_{m(\mathcal{B})^!\odot\mathcal{B}}\widehat{D}(N)$ is defined as $\widehat{A}(M)\otimes_{\mathcal{I}_\beta}\widehat{D}(N)$. A $\mathbb{Z}$–basis for $\widehat{A}(M)$ (respectively $\widehat{D}(N)$) is given by $\{X_{x_i\cdot h_1}\}$ (respectively $\{Y_{h_1\cdot y_i}\}$). A generator $X_{x_i\cdot h_1}$ of $\widehat{A}(M)$ has the same idempotent in $\mathcal{I}_\beta$ as a generator $Y_{h_2\cdot y_i}$ of $\widehat{D}(N)$ if and only if $h_1=h_2$.

Thus, $\widehat{A}(M)\otimes_{\mathcal{I}_\beta}\widehat{D}(N)$ has a $\mathbb{Z}$–basis consisting of all elements $X_{x_i\cdot h_1}\otimes Y_{h_1\cdot y_j}$, which is in bijection with the basis $\{x_i\cdot h_1\cdot y_j\}$ for $M\otimes_{H^n}N$. The bigradings agree on these two $\mathbb{Z}$–modules after negating the intrinsic gradings on $M\otimes_{H^n}N$: note that for the intrinsic grading on $\widehat{A}(M)\otimes_{\mathcal{I}_\beta}\widehat{D}(N)$, the grading of h_1 in $X_{x_i\cdot h_1}\otimes Y_{h_1\cdot y_j}$ is counted twice with coefficient $-\frac{1}{2}$, while for the intrinsic grading on $M\otimes_{H^n}N$, the grading of h in $x_i\cdot h_1\cdot y_j$ is counted once with coefficient 1. This explains the factor of $\frac{1}{2}$ in Definition 5.1.1.

It remains to show that the differential $\partial^{\otimes}$ on $M\otimes_{H^n}N$ agrees with the differential $\partial^{\boxtimes}$ on $\widehat{A}(M)\boxtimes_{m(\mathcal{B})^!\odot\mathcal{B}}\widehat{D}(N)$. We will use m_1 and m_2 to denote the differential and algebra action on $\widehat{A}(M)$ and δ to denote the type D operation on $\widehat{D}(N)$. Applying $\partial^{\boxtimes}$ to a generator $X_{x_i\cdot h_1}\otimes Y_{h_1\cdot y_j}$, we get

$$(-1)^{\deg_h(Y_{h_1\cdot y_j})}(m_1(X_{x_i\cdot h_1}))\otimes Y_{h_1\cdot y_j}+(m_2\otimes\mathrm{id})\circ(X_{x_i\cdot h_1}\otimes\delta(Y_{h_1\cdot y_j})).$$

Because H^n is concentrated in homological degree zero,

$$\deg_h(Y_{h_1\cdot y_j})=\deg_h(h_1\cdot y_j)=\deg_h(y_j).$$

Thus, the first term of $\partial^{\boxtimes}(X_{x_i\cdot h_1}\otimes Y_{h_1\cdot y_j})$ is

$$(-1)^{\deg_h y_j}\sum_k c_{i,k}X_{x_k\cdot h_1}\otimes Y_{h_1\cdot y_j},$$

which agrees with the first term of expression (6-5) for $\partial^{\otimes}(x_i\cdot h_1\cdot y_j)$ under the bijection between basis elements.

By Proposition 6.3.9, the other term of $\partial^{\boxtimes}(X_{x_i\cdot h_1}\otimes Y_{h_1\cdot y_j})$ can be expanded out as

$$(m_2\otimes\mathrm{id})\circ\Bigg(X_{x_i\cdot h_1}\otimes\Bigg(\sum_l c'_{j,l}Y_{h_1\cdot y_l}+\sum_{l,h'\in\beta_{\mathrm{mult}},h_2\in\beta}\tilde{c}'_{j,l;h'}\tilde{\tilde{c}}_{h_1h';h_2}b_{*;h_1,h_2}\otimes Y_{h_2\cdot y_l}$$
$$+\sum_{\substack{l,h_2\in\beta\text{ such that}\\ m(b^*_{*;m(h_1),m(h_2)})\text{ exists}}}(-1)^{\deg_h y_j}m(b^*_{*;m(h_1),m(h_2)})\otimes Y_{h_2\cdot y_j}\Bigg)\Bigg),$$

where $b_{*;h_1,h_2}$ denotes either $b_{\gamma;h_1,h_2}$ or $b_{C;h_1,h_2}$ and $m(b^*_{*;m(h_1),m(h_2)})$ denotes either $m(b^*_{\gamma;m(h_1),m(h_2)})$ or $m(b^*_{C;m(h_1),m(h_2)})$. This expansion gives us three remaining terms of $\partial^{\boxtimes}(X_{x_i\cdot h_1}\otimes Y_{h_1\cdot y_j})$. The first of these is

$$\sum_l c'_{j,l}X_{x_i\cdot h_1}\otimes Y_{h_1\cdot y_l},$$

which agrees with the third term of expression (6-5) under the bijection between basis elements. The second is

$$\sum_{l,h'\in\beta_{\mathrm{mult}},h_2\in\beta}\tilde{c}'_{j,l;h'}\tilde{\tilde{c}}_{h_1h';h_2}X_{x_i\cdot h_2}\otimes Y_{h_2\cdot y_l},$$

which agrees with the fourth term of expression (6-5). Finally, the remaining term of $\partial^{\boxtimes}(X_{x_i\cdot h_1}\otimes Y_{h_1\cdot y_j})$ is

$$(-1)^{\deg_h y_j}\sum_{k,h'\in\beta_{\mathrm{mult}},h_2\in\beta}\tilde{c}_{i,k;h'}\tilde{\tilde{c}}_{h'h_1;h_2}X_{x_k\cdot h_2}\otimes Y_{h_2\cdot y_j},$$

which agrees with the second term of expression (6-5). Thus, after reversing the intrinsic gradings on $M\otimes_{H^n}N$, we conclude that $M\otimes_{H^n}N$ is isomorphic to $\widehat{A}(M)\otimes_{m(\mathcal{B})^!\odot\mathcal{B}}\widehat{D}(N)$ as differential bigraded $\mathbb{Z}$–modules. □

6.4.2 Remark The negation of the intrinsic gradings on $M\otimes_{H^n}N$ is done for the same reason as in Remark 3.0.3.

6.4.3 Proposition *Let $\mathcal{B}$ be a differential bigraded algebra and let J be a bigrading-homogeneous ideal of $\mathcal{B}$ which is preserved by the differential on $\mathcal{B}$. Let $\widehat{D}$ be a type D structure over $\mathcal{B}$ and let $\widehat{A}$ be a differential bigraded right $\mathcal{B}$–module which descends to a module over $\mathcal{B}/J$. By Proposition 5.3.7, $\widehat{D}$ automatically descends to a type D structure over $\mathcal{B}/J$, and we have*

$$\widehat{A}\boxtimes_{\mathcal{B}}\widehat{D}\cong\widehat{A}\boxtimes_{\mathcal{B}/J}\widehat{D}.$$

Proof This follows immediately from Definition 2.3.5. □

6.4.4 Corollary *Let T_1 and T_2 be oriented tangle diagrams in $\mathbb{R}_{\geq 0}\times\mathbb{R}$ and $\mathbb{R}_{\leq 0}\times\mathbb{R}$ respectively, with orderings chosen of the crossings of T_1 and T_2. Assume that T_1 and T_2 have consistent orientations, so that their horizontal concatenation is an oriented link diagram L in $\mathbb{R}^2$. Order the crossings of L such that those of T_1 come before those of T_2. Then*

$$\mathrm{CKh}(L) \cong \widehat{A}([T_2]^{\mathrm{Kh}}) \boxtimes_{m(\mathcal{B})^! \odot \mathcal{B}} \widehat{D}([T_1]^{\mathrm{Kh}}) \cong \widehat{A}([T_2]^{\mathrm{Kh}}) \boxtimes_{\mathcal{B}\Gamma_n} \widehat{D}([T_1]^{\mathrm{Kh}}).$$

Proof This is a corollary of Proposition 6.4.1, Proposition 6.4.3 and Khovanov's results from [4]. □

Identifying $\widehat{A}([T_2]^{\mathrm{Kh}})$ with Roberts' type A structure over $\mathcal{B}\Gamma_n$ by Proposition 6.2.7, and identifying $\widehat{D}([T_1]^{\mathrm{Kh}})$ with Roberts' type D structure over $\mathcal{B}\Gamma_n$ by Proposition 6.3.10, we obtain an alternate proof of Roberts [11, Proposition 36].

6.5 Equivalences of type A structures

We start by defining $\mathcal{A}_\infty$–morphisms. The following definition is general enough for our purposes, although it is not the most general definition possible. A more general definition is given in Roberts [11, Definition 26]; our sign conventions are the same as Roberts'.

6.5.1 Definition Let $\mathcal{B}$ be a differential bigraded algebra with idempotent ring $\mathcal{I}$. Let $\widehat{A}$ and $\widehat{A}'$ be differential bigraded right modules over $\mathcal{B}$ with differentials m_1, m_1' and algebra actions m_2, m_2' respectively. An $\mathcal{A}_\infty$*–morphism* F from $\widehat{A}$ to $\widehat{A}'$ is a collection

$$F_n\colon \widehat{A} \otimes_{\mathcal{I}} \mathcal{B}^{\otimes(n-1)} \to \widehat{A}'[0, n-1]$$

of bigrading-preserving $\mathcal{I}$–linear maps satisfying the compatibility condition

$$\begin{aligned} m_1' \circ F_n + (-1)^n m_2' \circ (F_{n-1} \otimes |\mathrm{id}|^n) \\ = F_{n-1} \circ (m_2 \otimes \mathrm{id}^{\otimes(n-2)}) + (-1)^{n+1} F_n \circ (m_1 \otimes |\mathrm{id}|^{\otimes(n-1)}) \\ + (-1)^{n+1} \sum_{k=1}^{n-1} F_n \circ (\mathrm{id}^{\otimes k} \otimes \mu_1 \otimes |\mathrm{id}|^{\otimes(n-k-1)}) \\ + \sum_{k=1}^{n-2} (-1)^k F_{n-1} \circ (\mathrm{id}^{\otimes k} \otimes \mu_2 \otimes |\mathrm{id}|^{\otimes(n-k-2)}) \end{aligned}$$

for all $n \geq 1$. Recall that $|\mathrm{id}|^n$ and $|\mathrm{id}|^{\otimes n}$ mean different things; see Section 2.1.

6.5.2 Remark Lipshitz, Ozsváth and Thurston also require all $\mathcal{A}_\infty$–morphisms to satisfy a unitality condition: all algebra inputs must be in the kernel $\mathcal{B}_+$ of the augmentation of $\mathcal{B}$ for the corresponding $\mathcal{A}_\infty$–morphism term to be nonzero. See [7, Remark 2.2.21]. We will not discuss this condition further because it is satisfied for all $\mathcal{A}_\infty$–morphisms and homotopies that we consider.

6.5.3 Example For an $\mathcal{A}_\infty$–morphism F with only F_1 and F_2 nonzero, the condition of Definition 6.5.1 is nontrivial only for $n = 1$, 2 and 3. The $n = 1$ condition is

$$m'_1 \circ F_1 = F_1 \circ m_1,$$

the $n = 2$ condition is

$$m'_1 \circ F_2 + m'_2 \circ (F_1 \otimes \mathrm{id}) = F_1 \circ m_2 - F_2 \circ (m_1 \otimes |\mathrm{id}|) - F_2 \circ (\mathrm{id} \otimes \mu_1),$$

and the $n = 3$ condition is

$$-m'_2 \circ (F_2 \otimes |\mathrm{id}|) = F_2 \circ (m_2 \otimes \mathrm{id}) - F_2 \circ (\mathrm{id} \otimes \mu_2).$$

Let (M, d_M) and $(M', d_{M'})$ be two chain complexes of graded projective right H^n–modules satisfying the algebraic condition C_{module} of Definition 6.1.1 for generating sets $\{x_i\}$ and $\{x'_i\}$ respectively. Let $f\colon M \to M'$ be a bigrading-preserving H^n–linear map such that $d_{M'} f = f d_M$; as shorthand, we will say "let f be a chain map from M to M'". We first show that certain chain maps f induce $\mathcal{A}_\infty$–morphisms of type A structures $\widehat{A}(M) \to \widehat{A}(M')$ over $m(\mathcal{B})^! \odot \mathcal{B}$.

Let $\{x_i \cdot h_1 \mid e(x_i) = e_L(h_1)\}$ be the $\mathbb{Z}$–basis for M corresponding to the generating set $\{x_i\}$ and let $\{x'_i \cdot h_1\}$ be the analogous $\mathbb{Z}$–basis for M' (we will suppress the idempotent conditions). We may expand $f(x_i)$ in the basis for M':

$$f(x_i) = \sum_j f_{i,j} x'_j + \sum_{j,h' \in \beta, \deg h' \neq 0} \tilde{f}_{i,j;h'} x'_j \cdot h'.$$

6.5.4 Definition The chain map f satisfies the algebraic condition C_{morphism} for the generating sets $\{x_i\}$ and $\{x'_j\}$ if $\tilde{f}_{i,j;h'}$ is only nonzero when $h' \in \beta_{\mathrm{mult}}$.

For a chain map f satisfying C_{morphism}, we may write

$$f(x_i) = \sum_j f_{i,j} x'_j + \sum_{j,h' \in \beta_{\mathrm{mult}}} \tilde{f}_{i,j;h'} x'_j \cdot h'.$$

Thus, a basis expansion for $f(x_i \cdot h_1)$ is

$$f(x_i \cdot h_1) = \sum_j f_{i,j} x'_j \cdot h_1 + \sum_{j,h' \in \beta_{\mathrm{mult}}, h_2 \in \beta} \tilde{f}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} x'_j \cdot h_2.$$

Since M and M' satisfy C_{module}, we also have

$$d_M(x_i \cdot h_1) = \sum_j c_{i,j}(x_j \cdot h_1) + \sum_{j,h' \in \beta_{\mathrm{mult}}, h_2 \in \beta} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2}(x_j \cdot h_2),$$

$$d_{M'}(x'_i \cdot h_1) = \sum_j c'_{i,j}(x'_j \cdot h_1) + \sum_{j,h' \in \beta_{\mathrm{mult}}, h_2 \in \beta} \tilde{c}'_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2}(x'_j \cdot h_2).$$

6.5.5 Proposition *Suppose the chain map f satisfies C_{morphism}. The equation $d_{M'} f = f d_M$ gives us the following equations in the coefficients $f_{i,j}$, $\tilde{f}_{i,j;h'}$, $c_{i,j}$, $\tilde{c}_{i,j;h'}$, $c'_{i,j}$ and $\tilde{c}'_{i,j,h'}$:*

(1) *For all generators x_i of M and x'_k of M',*

$$\sum_j f_{i,j} c'_{j,k} = \sum_j c_{i,j} f_{j,k}.$$

(2) *For all generators $x_i \cdot h_1$ of M and $x'_k \cdot h_3$ of M',*

$$\sum_{j,h' \in \beta_\gamma} \tilde{f}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_3} c'_{j,k} + \sum_{j,h' \in \beta_\gamma} f_{i,j} \tilde{c}'_{j,k;h'} \tilde{\tilde{c}}_{h'h_1;h_3} = \sum_{j,h' \in \beta_\gamma} c_{i,j} \tilde{f}_{j,k;h'} \tilde{\tilde{c}}_{h'h_1;h_3} + \sum_{j,h' \in \beta_\gamma} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_3} f_{j,k}.$$

(3) *For all generators $x_i \cdot h_1$ of M and $x'_k \cdot h_3$ of M',*

$$\begin{aligned}\sum_{j,h' \in \beta_\alpha} \tilde{f}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_3} c'_{j,k} &+ \sum_{j,h' \in \beta_\alpha} f_{i,j} \tilde{c}'_{j,k;h'} \tilde{\tilde{c}}_{h'h_1;h_3} \\ &+ \sum_{\substack{j,h' \in \beta_\gamma, \\ h'' \in \beta_\gamma, h_2 \in \beta}} \tilde{f}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} \tilde{c}'_{j,k;h''} \tilde{\tilde{c}}_{h''h_2;h_3} \\ = \sum_{j,h' \in \beta_\alpha} c_{i,j} \tilde{f}_{j,k;h'} \tilde{\tilde{c}}_{h'h_1;h+3} &+ \sum_{j,h' \in \beta_\alpha} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_3} f_{j,k} \\ &+ \sum_{\substack{j,h' \in \beta_\gamma, \\ h'' \in \beta_\gamma, h_2 \in \beta}} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} \tilde{f}_{j,k;h''} \tilde{\tilde{c}}_{h''h_2;h_3}.\end{aligned}$$

(4) *For all generators* $x_i \cdot h_1$ *of* M *and* $x'_k \cdot h_3$ *of* M',

$$\sum_{\substack{j,h' \in \beta_\gamma, \\ h'' \in \beta_\alpha, h_2 \in \beta}} \tilde{f}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} \tilde{c}'_{j,k;h''} \tilde{\tilde{c}}_{h''h_2;h_3} + \sum_{\substack{j,h' \in \beta_\alpha, \\ h'' \in \beta_\gamma, h_2 \in \beta}} \tilde{f}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} \tilde{c}'_{j,k;h''} \tilde{\tilde{c}}_{h''h_2;h_3}$$
$$= \sum_{\substack{j,h' \in \beta_\gamma, \\ h'' \in \beta_\alpha, h_2 \in \beta}} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} \tilde{f}_{j,k;h''} \tilde{\tilde{c}}_{h''h_2;h_3} + \sum_{\substack{j,h' \in \beta_\alpha, \\ h'' \in \beta_\gamma, h_2 \in \beta}} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} \tilde{f}_{j,k;h''} \tilde{\tilde{c}}_{h''h_2;h_3}.$$

(5) *For all generators* $x_i \cdot h_1$ *of* M *and* $x'_k \cdot h_3$ *of* M',

$$\sum_{\substack{j,h' \in \beta_\alpha, \\ h'' \in \beta_\alpha, h_2 \in \beta}} \tilde{f}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} \tilde{c}'_{j,k;h''} \tilde{\tilde{c}}_{h''h_2;h_3} = \sum_{\substack{j,h' \in \beta_\alpha, \\ h'' \in \beta_\alpha, h_2 \in \beta}} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} \tilde{f}_{j,k;h''} \tilde{\tilde{c}}_{h''h_2;h_3}.$$

Proof The proof is very similar to that of Proposition 6.1.2 and will be omitted. Note that explicitly writing the degree conditions in the sums is unnecessary, since the relevant products of coefficients are always zero unless the degree conditions are satisfied. In Proposition 6.1.2, we chose to write out the degree conditions for clarity. □

6.5.6 Definition Suppose (M, d_M) and $(M', d_{M'})$ satisfy C_{module} and $f\colon M \to M'$ is a chain map satisfying C_{morphism}. Define the first component $\hat{A}(f)_1$ of an $\mathcal{A}_\infty$–morphism $\hat{A}(f)\colon \hat{A}(M) \to \hat{A}(M')$ of type A structures over $m(\mathcal{B})^! \odot \mathcal{B}$ by

$$\hat{A}(f)_1(X_{x_i \cdot h_1}) := \sum_j f_{i,j} X_{x'_j \cdot h_1}.$$

The map $\hat{A}(f)_1\colon \hat{A}(M) \to \hat{A}(M')$ respects the right actions of the idempotent ring $\mathcal{I}_\beta$, and it is bigrading-preserving because f is.

If $\hat{A}(f)_1$ were the only nonzero component of $\hat{A}(f)$, then $\hat{A}(f)$ would be an ordinary chain map between differential bigraded $m(\mathcal{B})^! \odot \mathcal{B}$–modules. However, $\hat{A}(f)_2$ will also be nonzero in general; thus, we must deal with genuine higher $\mathcal{A}_\infty$–terms when working with these morphisms. The component

$$\hat{A}(f)_2\colon \hat{A}(M) \otimes_{\mathcal{I}_\beta} m(\mathcal{B})^! \odot \mathcal{B} \to \hat{A}(M')[0,1]$$

of $\hat{A}(f)$ is defined on the generators $X_{x_i \cdot h_1}$ of $\hat{A}(M)$ and $m(b^*_{*;m(h_1),m(h_2)})$ of $m(\mathcal{B})^! \odot \mathcal{B}$ by

$$\hat{A}(f)_2(X_{x_i \cdot h_1}, m(b^*_{*;m(h_1),m(h_2)})) := \sum_{j,h' \in \beta_{\mathrm{mult}}} \tilde{f}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} X_{x'_j \cdot h_2},$$

where $m(b^*_{*;m(h_1),m(h_2)})$ denotes $m(b^*_{\gamma;m(h_1),m(h_2)})$ or $m(b^*_{C;m(h_1),m(h_2)})$ as appropriate. Any action of the form $\hat{A}(f)_2(X_{x_i \cdot h_1}, b_{*;h_1,h_2})$ is defined to be zero.

Suppose that the algebra input $m(b^*_{*;m(h_1),m(h_2)})$ equals $m(b^*_{\gamma;m(h_1),m(h_2)})$. Let $h_1 = (W(a)b,\sigma)$ and $h_2 = (W(a')b,\sigma')$; then $\tilde{\tilde{c}}_{h'h_1;h_2}$ is only nonzero for one value of h', namely $h' = (W(a')a, \text{all plus})$. For this value of h', $\tilde{\tilde{c}}_{h'h_1;h_2}$ is 1. Thus,

$$\widehat{A}(f)_2(X_{x_i\cdot h_1}, m(b^*_{\gamma;m(h_1),m(h_2)})) = \sum_j \tilde{f}_{i,j;h'} X_{x'_j\cdot h_2}. \tag{6-6}$$

Now suppose the algebra input is equal to $m(b^*_{C;m(h_1),m(h_2)})$. As before, write h_1 as $h_1 = (W(a)b,\sigma)$. In this case, $\tilde{\tilde{c}}_{h'h_1;h_2}$ will be nonzero for any h'_α which equals $(W(a)a,\ \text{minus on}\ \alpha)$, where α is any arc of a which is part of the circle C in $W(a)a$. For h' equal to one of the h'_α, we have $\tilde{\tilde{c}}_{h'h_1;h_2} = 1$, and for all other h', $\tilde{\tilde{c}}_{h'h_1;h_2}$ is zero. Thus,

$$\widehat{A}(f)_2(X_{x_i\cdot h_1}, m(b^*_{C;m(h_1),m(h_2)})) = \sum_j \sum_{\text{left arcs } \alpha \text{ of } C} \tilde{f}_{i,j;h'_\alpha} X_{x'_j\cdot h_2}. \tag{6-7}$$

Writing $m(\mathcal{B})^! \odot \mathcal{B}$ as $T(V_{\text{full}})/J_{\text{full}}$ as in Section 5.3, the above formulas define a map

$$\widehat{A}(f)_2\colon \widehat{A}(M) \otimes_{\mathcal{I}_\beta} V_{\text{full}} \to \widehat{A}(M')[0,1].$$

We can extend to a map

$$\widehat{A}(f)_2\colon \widehat{A}(M) \otimes_{\mathcal{I}_\beta} T(V_{\text{full}}) \to \widehat{A}(M')[0,1]$$

which is defined as the sum, over all $n \geq 2$, of the maps

$$\sum_{k=1}^{n-1} m'_2 \circ (m'_2 \otimes \text{id}) \circ \cdots \circ (m'_2 \otimes \text{id}^{\otimes(k-2)}) \\ \circ (\widehat{A}(f)_2 \otimes |\text{id}|^{\otimes(k-1)}) \circ (m_2 \otimes \text{id}^{\otimes k}) \circ \cdots \circ (m_2 \otimes \text{id}^{\otimes(n-2)})$$

from $\widehat{A}(M) \otimes (V_{\text{full}})^{\otimes(n-1)}$ to $\widehat{A}(M')$. In Proposition 6.5.8, we show that $\widehat{A}(f)_2$ descends to a map

$$\widehat{A}(f)_2\colon \widehat{A}(M) \otimes_{\mathcal{I}_\beta} m(\mathcal{B})^! \odot \mathcal{B} \to \widehat{A}(M');$$

in Proposition 6.5.9 we verify that $\widehat{A}(f)_1$ and $\widehat{A}(f)_2$ together satisfy the conditions to form an $\mathcal{A}_\infty$–morphism $\widehat{A}(f)$.

6.5.7 Example The $n=2$ summand of $\widehat{A}(f)_2\colon \widehat{A}(M) \otimes_{\mathcal{I}_\beta} T(V_{\text{full}}) \to \widehat{A}(M')$ is simply $\widehat{A}(f)_2$, the map from $\widehat{A}(M) \otimes_{\mathcal{I}_\beta} V_{\text{full}}$ to $\widehat{A}(M')$ defined above. The $n=3$ summand of $\widehat{A}(f)_2\colon \widehat{A}(M) \otimes_{\mathcal{I}_\beta} T(V_{\text{full}}) \to \widehat{A}(M')$, or in other words the definition of $\widehat{A}(f)_2$ when the algebra input is a quadratic monomial in the generators of V_{full}, is

$$\widehat{A}(f)_2 \circ (m_2 \otimes \text{id}) + m'_2 \circ (\widehat{A}(f)_2 \otimes |\text{id}|),$$

where in this expression $\widehat{A}(f)_2$ again denotes the map from $\widehat{A}(M) \otimes_{\mathcal{I}_\beta} V_{\text{full}}$ to $\widehat{A}(M')$.

6.5.8 Proposition *Write $m(\mathcal{B})^! \odot \mathcal{B}$ as $T(V_{\mathrm{full}})/J_{\mathrm{full}}$ and let r be any element of J_{full}. The map*

$$\widehat{A}(f)_2(-,r)\colon \widehat{A}(M) \to \widehat{A}(M')$$

is identically zero. Thus, we get a well-defined map

$$\widehat{A}(f)_2\colon \widehat{A}(M) \otimes_{\mathcal{I}_\beta} m(\mathcal{B})^! \odot \mathcal{B} \to \widehat{A}(M')$$

which is linear with respect to the right actions of the idempotent ring $\mathcal{I}_\beta$ on $\widehat{A}(M) \otimes_{\mathcal{I}_\beta} m(\mathcal{B})^! \odot \mathcal{B}$ and $\widehat{A}(M')$, and which preserves the intrinsic grading and decreases the homological grading by one.

Proof First, $\widehat{A}(f)_2$ decreases the homological grading by 1 (and thus preserves homological grading when accounting for shifts), since $m(b^*_{*;m(h_1),m(h_2)})$ carries homological grading 1 and $f\colon M \to M'$ preserves homological grading. Also, $\widehat{A}(f)_2$ preserves the intrinsic grading. To see this, note that as elements of M and M', $x_i \cdot h_1$ and $x'_j \cdot h_2$ must have the same intrinsic grading whenever $x'_j \cdot h_2$ appears with nonzero coefficient in the basis expansion of $f(x_i \cdot h_1)$, because f preserves intrinsic grading. As elements of H^n, the intrinsic degree of h_2 is either one or two greater than that of h_1. Since, in $\widehat{A}(M)$ and $\widehat{A}(M')$, the intrinsic degrees of h_1 and h_2 are multiplied by $-\frac{1}{2}$ whereas the intrinsic degrees of x_i and x'_j are multiplied by -1, the element $X_{x'_j \cdot h_2}$ of $\widehat{A}(M')$ should have intrinsic degree which is $\frac{1}{2}$ or 1 greater than the intrinsic degree of $X_{x_i \cdot h_1} \in \widehat{A}(M)$. This extra $\frac{1}{2}$ or 1 is compensated exactly by the intrinsic degree of $m(b^*_{*;m(h_1),m(h_2)})$, which is $\frac{1}{2}$ for $m(b^*_{\gamma;m(h_1),m(h_2)})$ and 1 for $m(b^*_{C;m(h_1),m(h_2)})$.

To show that $\widehat{A}(f)_2(-,r)$ is zero for any $r \in J_{\mathrm{full}}$, note first that Definition 6.5.6 implies that if we have elements r and r' of $T(V_{\mathrm{full}})$ such that $m_2(-,r) = 0$, $m'_2(-,r') = 0$, $\widehat{A}(f)_2(-,r) = 0$ and $\widehat{A}(f)_2(-,r') = 0$, then $\widehat{A}(f)_2(-,r \cdot r') = 0$ as well.

Thus, we only need to show that $\widehat{A}(f)_2(-,r)$ is zero for the multiplicative generators r of J_{full}. These were defined to be the generators of

$$J_{\mathcal{B}} \cap (T^1(V_{\mathcal{B}}) \oplus T^2(V_{\mathcal{B}})), \quad J_{m(\mathcal{B})^!} \cap (T^1(V_{m(\mathcal{B})^!}) \oplus T^2(V_{m(\mathcal{B})^!})) \quad \text{and} \quad J_{\mathrm{extra}}.$$

For generators in $J_{\mathcal{B}} \cap (T^1(V_{\mathcal{B}}) \oplus T^2(V_{\mathcal{B}}))$, there is nothing to show, since $\widehat{A}(f)_2(-,b)$ is zero for any $b \in V_{\mathcal{B}}$.

For the generators in $J_{m(\mathcal{B})^!} \cap (T^1(V_{m(\mathcal{B})^!}) \oplus T^2(V_{m(\mathcal{B})^!}))$, the proof closely follows the proof of Proposition 6.2.2. Write $J_{m(\mathcal{B})^!} \cap (T^1(V_{m(\mathcal{B})^!}) \oplus T^2(V_{m(\mathcal{B})^!}))$ as $m(I^\perp)$ as in that proof.

The generators $m(r^*)$ of $m(I^\perp)$ have intrinsic degree either 1, $\frac{3}{2}$ or 2, as in Section 5.2. For those $m(r^*)$ of intrinsic degree 2, the equations in item (5) of Proposition 6.5.5

above imply that $\widehat{A}(f)_2(X_{x_i\cdot h_1}, m(r^*)) = 0$ for any $X_{x_i\cdot h_1}$. For those $m(r^*)$ of intrinsic degree $\frac{3}{2}$, the equations in item (4) of Proposition 6.5.5 similarly imply that $\widehat{A}(f)_2(-, m(r^*))$ is zero.

The generators $m(r^*)$ of $m(I^\perp)$ which have intrinsic degree 1 are sums of either one, two, three or four terms $m(b^*_\gamma)m(b^*_{\gamma'})$ with all coefficients $+1$. For a fixed $m(r^*)$, let $h_1 \in \mathcal{I}_\beta$ denote its left idempotent and let $h_3 \in \mathcal{I}_\beta$ denote its right idempotent. The element h_3 of β has degree 2 more than h_1, as elements of H^n with its intrinsic grading, and h_3 differs from h_1 by two surgeries on its left crossingless matching. As in Proposition 6.2.2, the left crossingless matchings of h_1 and h_3 are different.

For any generators of $\widehat{A}(M)$ and $\widehat{A}(M')$ of the form $x_i \cdot h_1$ and $x'_k \cdot h_3$, where h_1 and h_3 are as above, the equations from item (3) of Proposition 6.5.5 become

$$\sum_{\substack{j,h'\in\beta_\gamma,\\ h''\in\beta_\gamma,h_2\in\beta}} \tilde{f}_{i,j;h'}\tilde{\tilde{c}}_{h'h_1;h_2}\tilde{c}'_{j,k;h''}\tilde{\tilde{c}}_{h''h_2;h_3} = \sum_{\substack{j,h'\in\beta_\gamma,\\ h''\in\beta_\gamma,h_2\in\beta}} \tilde{c}_{i,j;h'}\tilde{\tilde{c}}_{h'h_1;h_2}\tilde{f}_{j,k;h''}\tilde{\tilde{c}}_{h''h_2;h_3}.$$

The terms involving $h' \in \beta_\alpha$ vanish for these choices of h_1 and h_3. These equations imply that for all generators $m(r^*)$ of $m(I^\perp)$ of intrinsic degree 1, $\widehat{A}(f)_2(-, m(r^*))$ is zero.

Finally, the generators of J_{extra} are listed in items (1)–(5) of Definition 5.3.2. If r is one of these generators, the proof that the map $\widehat{A}(f)_2(-, r)$ is zero is similar to the proof of Proposition 6.2.5.

In more detail, consider a relation

$$r = b_{\gamma;h_1,h_2} m(b^*_{\eta';m(h_2),m(h_3)}) - m(b^*_{\eta;m(h_1),m(\tilde{h}_2)}) b_{\gamma';\tilde{h}_2,h_3}$$

from item (1) of Definition 5.3.2. Write $h_1 = (W(a_1)b_1, \sigma_1)$ and let $X_{x_i\cdot h_1}$ be a generator of $\widehat{A}(M)$. By Example 6.5.7, we have

$$\text{(6-8)}\quad \begin{aligned}\widehat{A}(f)_2(X_{x_i\cdot h_1}, r) = \widehat{A}(f)_2\big(&m_2(X_{x_i\cdot h_1}, b_{\gamma;h_1,h_2}), m(b^*_{\eta';m(h_2),m(h_3)})\big)\\ &- m_2\big(\widehat{A}(f)_2(X_{x_i\cdot h_1}, m(b^*_{\eta;m(h_1),m(\tilde{h}_2)})), b_{\gamma';\tilde{h}_2,h_3}\big)\end{aligned}$$

Write h_2 as $(W(a_1)b_2, \sigma_2)$ and h_3 as $(W(a_2)b_2, \sigma_3)$. Let $h' = (W(a_2)a_1, \text{all plus})$, an element of β_γ. For the first term in (6-8) above, we first multiply $X_{x_i\cdot h_1}$ by $b_{\gamma;h_1,h_2}$ to get $X_{x_i\cdot h_2}$. Applying $\widehat{A}(f)_2(-, m(b^*_{\eta';m(h_2),m(h_3)}))$ to the element $X_{x_i\cdot h_2}$, by (6-6) we get

$$\sum_j \tilde{f}_{i,j;h'} X_{x_j\cdot h_3}.$$

For the second term on the right side of (6-8), we first compute

$$\widehat{A}(f)_2(X_{x_i\cdot h_1}, m(b^*_{\eta;m(h_1),m(\tilde{h}_2)})).$$

This expression equals

$$\sum_j \tilde{f}_{i,j;h'} X_{x_j \cdot \tilde{h}_2}$$

by (6-6) again, where h' is still equal to $(W(a_2)a_1, \text{all plus})$. If we multiply this result by $b_{\gamma';\tilde{h}_2,h_3}$, we get

$$\sum_j \tilde{f}_{i,j;h'} X_{x_j \cdot h_3}.$$

Thus, if a generator r of J_{extra} comes from item (1) of Definition 5.3.2, then the map $\hat{A}(f)_2(-,r)\colon \hat{A}(M) \to \hat{A}(M')$ is zero.

For generators of J_{extra} from items (2)–(5) of Definition 5.3.2, the proof is analogous to that of Proposition 6.2.5 in the same way as above. We will leave the remaining cases to the reader. □

6.5.9 Proposition $\hat{A}(f)$ *satisfies the* $\mathcal{A}_\infty$*–morphism compatibility conditions.*

Proof Since $\hat{A}(f)_n$ is zero for $n > 2$, it suffices to show that the $n = 1$, $n = 2$ and $n = 3$ conditions listed in Example 6.5.3 hold. For the $n = 1$ condition, we want to show that

$$m'_1(\hat{A}(f)_1(X_{x_i \cdot h_1})) = \hat{A}(f)_1(m_1(X_{x_i \cdot h_1}))$$

for each generator $X_{x_i \cdot h_1}$ of $\hat{A}(M)$. The left side is

$$m'_1\Bigl(\sum_j f_{i,j} X_{x_j \cdot h_1}\Bigr) = \sum_{j,k} f_{i,j} c'_{j,k} X_{x'_k \cdot h_1},$$

while the right side is

$$\hat{A}(f)_1\Bigl(\sum_j c_{i,j} X_{x_j \cdot h_1}\Bigr) = \sum_{j,k} c_{i,j} f_{j,k} X_{x'_k \cdot h_1}.$$

These are equal by item (1) of Proposition 6.5.5.

We may write the $n = 3$ condition of Example 6.5.3 as

$$\hat{A}(f)_2 \circ (\mathrm{id} \otimes \mu_2) = \hat{A}(f)_2 \circ (m_2 \otimes \mathrm{id}) + m'_2 \circ (\hat{A}(f)_2 \otimes |\mathrm{id}|). \tag{6-9}$$

In this form, it is clear from the definition of $\hat{A}(f)_2$ in Definition 6.5.6 that this condition holds, generalizing Example 6.5.7.

For the $n = 2$ condition, we want to show that

$$\begin{aligned} m'_1 \circ \hat{A}(f)_2 + m'_2 \circ (\hat{A}(f)_1 \otimes \mathrm{id}) \\ = \hat{A}(f)_1 \circ m_2 - \hat{A}(f)_2 \circ (m_1 \otimes |\mathrm{id}|) - \hat{A}(f)_2 \circ (\mathrm{id} \otimes \mu_1) \end{aligned} \tag{6-10}$$

as maps from $\widehat{A}(M) \otimes_{\mathcal{I}_\beta} m(\mathcal{B})^! \odot \mathcal{B}$ to $\widehat{A}(M')$.

We will first reduce to the case of proving the above equation when applied to elements of the form $X_{x_i \cdot h_1} \otimes b_{*;h_1,h_2}$ or $X_{x_i \cdot h_1} \otimes m(b^*_{*;m(h_1),m(h_2)})$, where $b_{*;h_1,h_2}$ and $m(b^*_{*;m(h_1),m(h_2)})$ are the multiplicative generators of $m(\mathcal{B})^! \odot \mathcal{B}$.

Claim *If b_1 and b_2 are two elements of $m(\mathcal{B})^! \odot \mathcal{B}$ such that the $n=2$ condition is satisfied both when the algebra input is b_1 and when it is b_2, then the $n=2$ condition is also satisfied when the algebra input is $b_1 b_2$.*

Proof of claim We want to show that

$$\begin{aligned} m'_1 \circ \widehat{A}(f)_2 \circ (\mathrm{id} \otimes \mu_2) + m'_2 \circ (\widehat{A}(f)_1 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \mu_2) \\ = \widehat{A}(f)_1 \circ m_2 \circ (\mathrm{id} \otimes \mu_2) - \widehat{A}(f)_2 \circ (m_1 \otimes |\mathrm{id}|) \circ (\mathrm{id} \otimes \mu_2) \\ - \widehat{A}(f)_2 \circ (\mathrm{id} \otimes \mu_1) \circ (\mathrm{id} \otimes \mu_2) \end{aligned} \tag{6-11}$$

when the algebra input to these maps is $b_1 \otimes b_2$, assuming the usual $n=2$ condition (6-10) holds when the algebra input is b_1 or b_2. In the proof of this claim, the algebra input to all maps will be assumed to be $b_1 \otimes b_2$.

The left side of (6-11) can be rewritten as

$$\begin{aligned} m'_1 \circ m'_2 \circ (\widehat{A}(f)_2 \otimes |\mathrm{id}|) + m'_1 \circ \widehat{A}(f)_2 \circ (m_2 \otimes \mathrm{id}) \\ + m'_2 \circ (m'_2 \otimes \mathrm{id}) \circ (\widehat{A}(f)_1 \otimes \mathrm{id} \otimes \mathrm{id}) \end{aligned} \tag{6-12}$$

using the $n=3$ consistency condition (6-9) for $\widehat{A}(f)$ and the associativity of the action of $m(\mathcal{B})^! \odot \mathcal{B}$ on $\widehat{A}(M')$. Call these terms LHS_1, LHS_2 and LHS_3. Now we may use the assumption that the $n=2$ consistency condition (6-10) holds when the algebra input is b_1 to write the third term LHS_3 of expression (6-12) as

$$\begin{aligned} m'_2 \circ (\widehat{A}(f)_1 \otimes \mathrm{id}) \circ (m_2 \otimes \mathrm{id}) - m'_2 \circ (m'_1 \otimes |\mathrm{id}|) \circ (\widehat{A}(f)_2 \otimes |\mathrm{id}|) \\ - m'_2 \circ (\widehat{A}(f)_2 \otimes \mathrm{id}) \circ (m_1 \otimes |\mathrm{id}| \otimes \mathrm{id}) \\ - m'_2 \circ (\widehat{A}(f)_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \mu_1 \otimes \mathrm{id}). \end{aligned}$$

Call these four terms LHS_{3a}, LHS_{3b}, LHS_{3c} and LHS_{3d}.

On the other hand, the right side of (6-11) can be rewritten as

$$\begin{aligned} \widehat{A}(f)_1 \circ m_2 \circ (m_2 \otimes \mathrm{id}) - m'_2 \circ (\widehat{A}(f)_2 \otimes \mathrm{id}) \circ (m_1 \otimes |\mathrm{id}| \otimes \mathrm{id}) \\ - \widehat{A}(f)_2 \circ (m_2 \otimes \mathrm{id}) \circ (m_1 \otimes |\mathrm{id}| \otimes |\mathrm{id}|) \\ - \widehat{A}(f)_2 \circ (\mathrm{id} \otimes \mu_2) \circ (\mathrm{id} \otimes \mu_1 \otimes |\mathrm{id}|) \\ - \widehat{A}(f)_2 \circ (\mathrm{id} \otimes \mu_2) \circ (\mathrm{id} \otimes \mathrm{id} \otimes \mu_1) \end{aligned} \tag{6-13}$$

using the $n=3$ consistency equation (6-9) for $\widehat{A}(f)$, the associativity of the action of $m(\mathcal{B})^! \odot \mathcal{B}$ on $\widehat{A}(M)$ and the Leibniz rule for the derivative μ_1 on $m(\mathcal{B})^! \odot \mathcal{B}$. Call these five terms RHS_1, RHS_2, RHS_3, RHS_4 and RHS_5. Using the $n=3$ consistency equation again, we can rewrite the term RHS_4 as

$$-m_2' \circ (\widehat{A}(f)_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \mu_1 \otimes \mathrm{id}) - \widehat{A}(f)_2 \circ (m_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \mu_1 \otimes |\mathrm{id}|).$$

Call these two terms RHS_{4a} and RHS_{4b}. Similarly, we can rewrite the term RHS_5 as

$$-m_2' \circ (\widehat{A}(f)_2 \otimes |\mathrm{id}|) \circ (\mathrm{id} \otimes \mathrm{id} \otimes \mu_1) - \widehat{A}(f)_2 \circ (m_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \mathrm{id} \otimes \mu_1).$$

Call these terms RHS_{5a} and RHS_{5b}. Using the assumption that the $n=2$ consistency condition (6-10) holds when the algebra input is b_2, we can write the term RHS_1 as

$$\begin{aligned} m_2' \circ (\widehat{A}(f)_1 \otimes \mathrm{id}) \circ (m_2 \otimes \mathrm{id}) + m_1' &\circ \widehat{A}(f)_2 \circ (m_2 \otimes \mathrm{id}) \\ &+ \widehat{A}(f)_2 \circ (m_1 \otimes |\mathrm{id}|) \circ (m_2 \otimes \mathrm{id}) \\ &+ \widehat{A}(f)_2 \circ (m_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \mathrm{id} \otimes \mu_1). \end{aligned}$$

Call these four terms RHS_{1a}, RHS_{1b}, RHS_{1c} and RHS_{1d}.

After rewriting the left and right sides of (6-11) in this way, we want to show that

$$\begin{aligned} \mathrm{LHS}_1 + \mathrm{LHS}_2 + \mathrm{LHS}_{3a} + \mathrm{LHS}_{3b} + \mathrm{LHS}_{3c} + \mathrm{LHS}_{3d} \\ = \mathrm{RHS}_{1a} + \mathrm{RHS}_{1b} + \mathrm{RHS}_{1c} + \mathrm{RHS}_{1d} + \mathrm{RHS}_2 \\ + \mathrm{RHS}_3 + \mathrm{RHS}_{4a} + \mathrm{RHS}_{4b} + \mathrm{RHS}_{5a} + \mathrm{RHS}_{5b}. \end{aligned}$$

Several terms cancel:

- $\mathrm{LHS}_2 = \mathrm{RHS}_{1b}$,
- $\mathrm{LHS}_{3a} = \mathrm{RHS}_{1a}$,
- $\mathrm{LHS}_{3c} = \mathrm{RHS}_2$,
- $\mathrm{LHS}_{3d} = \mathrm{RHS}_{4a}$,
- $\mathrm{RHS}_{1d} + \mathrm{RHS}_{5b} = 0$,
- $\mathrm{RHS}_{1c} + \mathrm{RHS}_3 + \mathrm{RHS}_{4b} = 0$.

The final equality follows from the Leibniz rule for the differential m_1 on $\widehat{A}(M)$. Canceling corresponding terms between the sides, it remains to prove

$$\mathrm{LHS}_1 + \mathrm{LHS}_{3b} = \mathrm{RHS}_{5a}.$$

The Leibniz rule for the differential m_1' on $\widehat{A}(M')$ lets us rewrite $\mathrm{LHS}_1 + \mathrm{LHS}_{3b}$ as

$$m_2' \circ (\mathrm{id} \otimes \mu_1) \circ (\widehat{A}(f)_2 \otimes |\mathrm{id}|).$$

This term equals the remaining right-side term RHS_{5a}, because μ_1 increases homological grading by one. □

Thus, we have reduced to showing that the $n=2$ consistency condition (6-10) holds for $\widehat{A}(f)$ when the algebra input is either $b_{\gamma;h_1,h_2}$, $b_{C;h_1,h_2}$, $m(b^*_{\gamma;m(h_1),m(h_2)})$ or $m(b^*_{C;m(h_1),m(h_3)})$. If the input $b_{*;h_1,h_2}$ is $b_{\gamma;h_1,h_2}$ or $b_{C;h_1,h_2}$, the $\widehat{A}(f)_2$ terms in the $n=2$ equation are zero and we must show that

$$m'_2 \circ (\widehat{A}(f)_1 \otimes \mathrm{id}) = \widehat{A}(f)_1 \circ m_2$$

for these algebra inputs. If $X_{x_i \cdot h_1}$ is a generator of $\widehat{A}(M)$, then

$$m'_2 \circ (\widehat{A}(f)_1 \otimes \mathrm{id})(X_{x_i\cdot h_1}, b_{*;h_1,h_2}) = m'_2\Bigl(\sum_j f_{i,j} X_{x'_j \cdot h_1}, b_{*;h_1,h_2}\Bigr) = \sum_j f_{i,j} X_{x'_j \cdot h_2},$$

while

$$\widehat{A}(f)_1 \circ m_2(X_{x_i\cdot h_1}, b_{*;h_1,h_2}) = \widehat{A}(f)_1(X_{x_i\cdot h_2}) = \sum_j f_{i,j} X_{x'_j\cdot h_2},$$

and these are equal.

Now let the algebra input be a generator $m(b^*_{\gamma;m(h_1),m(h_2)})$. The left side of the $n=2$ condition with this algebra input and module input $X_{x_i\cdot h}$ is

$$\sum_{j,k,h'\in\beta_\gamma} \tilde{f}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} c'_{j,k} X_{x'_k\cdot h_2} + \sum_{j,k,h'\in\beta_\gamma} f_{i,j} \tilde{c}'_{j,k;h'} \tilde{\tilde{c}}_{h'h_1;h_2} X_{x'_k\cdot h_2},$$

and the right side is

$$\sum_{j,k,h'\in\beta_\gamma} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} f_{j,k} X_{x'_k\cdot h_2} + \sum_{j,k,h'\in\beta_\gamma} c_{i,j} \tilde{f}_{j,k;h'} \tilde{\tilde{c}}_{h'h_1;h_2} X_{x'_k\cdot h_2}.$$

These are equal by item (2) of Proposition 6.5.5.

Finally, let the algebra input be a generator $m(b^*_{C;m(h_1),m(h_3)})$. The left side of the $n=2$ condition with this algebra input and module input $X_{x_i\cdot h_1}$ is

$$\sum_{j,k,h'\in\beta_\alpha} \tilde{f}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_3} c'_{j,k} X_{x'_k\cdot h_3} + \sum_{j,k,h'\in\beta_\alpha} f_{i,j} \tilde{c}'_{j,k;h'} \tilde{\tilde{c}}_{h'h_1;h_3} X_{x'_k\cdot h_3}.$$

To compare with the right side of the $n=2$ condition, note that, as in the proof of Proposition 6.2.3, we have

$$\mu_1(m(b^*_{C;m(h_1),m(h_3)})) = -\sum_{h_2\in\beta} m(b^*_{\gamma;m(h_1),m(h_2)}) m(b^*_{\gamma^\dagger;m(h_2),m(h_3)}),$$

where the sum is implicitly over those h_2 such that generators $m(b^*_{\gamma;m(h_1),m(h_2)})$ and $m(b^*_{\gamma^\dagger;m(h_2),m(h_3)})$ exist. Thus,

$$
\begin{aligned}
&\widehat{A}(f)_2 \circ (\mathrm{id} \otimes \mu_1)(X_{x_i \cdot h_1}, m(b^*_{C;m(h_1),m(h_3)})) \\
&\qquad = -\sum_{h_2} \widehat{A}(f)_2(X_{x_i \cdot h_1}, m(b^*_{\gamma;m(h_1),m(h_2)}) m(b^*_{\gamma^\dagger;m(h_2),m(h_3)})) \\
&\qquad = \sum_{h_2 \in \beta} m'_2 \circ (\widehat{A}(f)_2 \otimes \mathrm{id})(X_{x_i \cdot h_1}, m(b^*_{\gamma;m(h_1),m(h_2)}) m(b^*_{\gamma^\dagger;m(h_2),m(h_3)})) \\
&\qquad\quad - \sum_{h_2 \in \beta} \widehat{A}(f)_2 \circ (m_2 \otimes \mathrm{id})(X_{x_i \cdot h_1}, m(b^*_{\gamma;m(h_1),m(h_2)}) m(b^*_{\gamma^\dagger;m(h_2),m(h_3)}))
\end{aligned}
$$

by the $n = 3$ consistency condition (6-9); note that $\deg_h m(b^*_{\gamma^\dagger;m(h_2),m(h_3)}) = 1$. Expanding the above expression out, the top line is

$$\sum_{\substack{j,k,h' \in \beta_\gamma, \\ h'' \in \beta_\gamma, h_2 \in \beta}} \tilde{f}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} \tilde{c}'_{j,k;h''} \tilde{\tilde{c}}_{h''h_2;h_3} X_{x'_k \cdot h_3}$$

and the bottom line is

$$-\sum_{\substack{j,k,h' \in \beta_\gamma, \\ h'' \in \beta_\gamma, h_2 \in \beta}} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} \tilde{f}_{j,k;h''} \tilde{\tilde{c}}_{h''h_2;h_3} X_{x'_k \cdot h_3}.$$

Thus, the right side of the $n = 2$ condition (6-10) with algebra input $m(b^*_{C;m(h_1),m(h_3)})$ and module input $X_{x_i \cdot h_1}$ is

$$
\begin{aligned}
&\sum_{j,k,h' \in \beta_\alpha} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_3} f_{j,k} X_{x'_k \cdot h_3} + \sum_{j,k,h' \in \beta_\alpha} c_{i,j} \tilde{f}_{j,k;h'} \tilde{\tilde{c}}_{h'h_1;h_3} X_{x'_k \cdot h_3} \\
&\qquad - \sum_{\substack{j,k,h' \in \beta_\gamma, \\ h'' \in \beta_\gamma, h_2 \in \beta}} \tilde{f}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} \tilde{c}'_{j,k;h''} \tilde{\tilde{c}}_{h''h_2;h_3} X_{x'_k \cdot h_3} \\
&\qquad\qquad + \sum_{\substack{j,k,h' \in \beta_\gamma, \\ h'' \in \beta_\gamma, h_2 \in \beta}} \tilde{c}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} \tilde{f}_{j,k;h''} \tilde{\tilde{c}}_{h''h_2;h_3} X_{x'_k \cdot h_3}.
\end{aligned}
$$

This is equal to the left side of the $n = 2$ condition with these inputs,

$$\sum_{j,k,h' \in \beta_\alpha} \tilde{f}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_3} c'_{j,k} X_{x'_k \cdot h_3} + \sum_{j,k,h' \in \beta_\alpha} f_{i,j} \tilde{c}'_{j,k;h'} \tilde{\tilde{c}}_{h'h_1;h_3} X_{x'_k \cdot h_3},$$

by item (3) of Proposition 6.5.5. □

We now define the composition of two $\mathcal{A}_\infty$–morphisms as in [11, Definition 27].

6.5.10 Definition Define the *composition* of two $\mathcal{A}_\infty$–morphisms $F\colon \widehat{A} \to \widehat{A}'$ and $G\colon \widehat{A}' \to \widehat{A}''$ to be the $\mathcal{A}_\infty$–morphism $G \circ F$ with

$$(G \circ F)_n := \sum_{i+j=n+1} (-1)^{(i+1)(j+1)} G_i \circ (F_j \otimes |\mathrm{id}|^{(j+1)\otimes(i-1)}).$$

For the purposes of this section, we will only need to compose morphisms $\widehat{A}(f)$ and $\widehat{A}(g)$ such that either f or g satisfies a more restrictive condition than C_{morphism}.

6.5.11 Definition A chain map $f\colon M \to M'$ of complexes of graded projective right H^n–modules satisfies the algebraic condition $\widetilde{C}_{\mathrm{morphism}}$ for the generating sets $\{x_i\}$ of M and $\{x'_j\}$ of M' if it satisfies C_{morphism} of Definition 6.5.4 and furthermore $\tilde{f}_{i,j;h'} = 0$ for all $h' \in \beta_{\mathrm{mult}}$.

6.5.12 Proposition *Let $f\colon M \to M'$ and $g\colon M' \to M''$ be chain maps between complexes of graded projective right H^n modules, such that M, M' and M'' satisfy the algebraic condition C_{module} of Definition 6.1.1, while f and g satisfy the condition C_{morphism} and either f or g satisfies the condition $\widetilde{C}_{\mathrm{morphism}}$. Then $g \circ f$ satisfies C_{morphism} and*

$$\widehat{A}(g \circ f) = \widehat{A}(g) \circ \widehat{A}(f).$$

Proof By the conditions on f and g, the chain map $g \circ f$ satisfies the condition C_{morphism}, so $\widehat{A}(g \circ f)$ is a well-defined $\mathcal{A}_\infty$–morphism. We have

$$(g \circ f)_{i,k} = \sum_j f_{i,j} g_{j,k}.$$

If g satisfies $\widetilde{C}_{\mathrm{morphism}}$, then for $h' \in \beta_{\mathrm{mult}}$ we have

$$\widetilde{(g \circ f)}_{i,k;h'} = \sum_j \tilde{f}_{i,j;h'} g_{j,k},$$

while if f satisfies $\widetilde{C}_{\mathrm{morphism}}$, then

$$\widetilde{(g \circ f)}_{i,k;h'} = \sum_j f_{i,j} \tilde{g}_{j,k;h'}.$$

Let $X_{x_i \cdot h_1}$ be a generator of $\widehat{A}(M)$. We have

$$\widehat{A}(g \circ f)_1(X_{x_i \cdot h_1}) = \sum_{j,k} f_{i,j} g_{j,k} X_{x''_k \cdot h_1},$$

and this sum also equals $(\widehat{A}(g) \circ \widehat{A}(f))_1(X_{x_i \cdot h_1})$.

Let $b_{*;h_1,h_2}$ be a generator of $\mathcal{B} \subset m(\mathcal{B})^! \odot \mathcal{B}$. By the definition of the operation $f \mapsto \widehat{A}(f)$,

$$\widehat{A}(g \circ f)_2(X_{x_i \cdot h_1}, b_{*;h_1,h_2}) = 0 = (\widehat{A}(f) \circ \widehat{A}(g))_2(X_{x_i \cdot h_1}, b_{*;h_1,h_2}).$$

Finally, let $m(b^*_{*;m(h_1),m(h_2)})$ be a generator of $m(\mathcal{B})^! \subset m(\mathcal{B})^! \odot \mathcal{B}$. Suppose g satisfies $\widetilde{C}_{\mathrm{morphism}}$. Then

$$\widehat{A}(g \circ f)_2(X_{x_i \cdot h_1}, m(b^*_{*;m(h_1),m(h_2)})) = \sum_{j,k,h' \in \beta_{\mathrm{mult}}} \tilde{f}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h_2} g_{j,k} X_{x''_k \cdot h_2},$$

while

$$\begin{aligned}(\widehat{A}(g) \circ \widehat{A}(f))_2(X_{x_i \cdot h_1}, m(b^*_{*;m(h_1),m(h_2)}))& \\ &= (\widehat{A}(g)_1 \circ \widehat{A}(f)_2)(X_{x_i \cdot h_1}, m(b^*_{*;m(h_1),m(h_2)})) \\ &= \sum_{j,k,h' \in \beta_{\mathrm{mult}}} \tilde{f}_{i,j;h'} \tilde{\tilde{c}}_{h'h_1;h2} g_{j,k} X_{x''_k \cdot h_2}.\end{aligned}$$

The case when f satisfies $\widetilde{C}_{\mathrm{morphism}}$ is analogous. Thus, $\widehat{A}(g \circ f)_2 = (\widehat{A}(g) \circ \widehat{A}(f))_2$. We have

$$\widehat{A}(g \circ f)_n = (\widehat{A}(g) \circ \widehat{A}(f))_n = 0$$

for all $n > 2$, so $\widehat{A}(g \circ f) = \widehat{A}(g) \circ \widehat{A}(f)$. □

Now we will consider homotopies. The following definition is a special case of [11, Definition 28].

6.5.13 Definition Let $\mathcal{B}$ be a differential bigraded algebra with idempotent ring $\mathcal{I}$. Let $\widehat{A}$ and $\widehat{A}'$ be differential bigraded right modules over $\mathcal{B}$. Let $F = \{F_n\}$ and $G = \{G_n\}$ be $\mathcal{A}_\infty$–morphisms from $\widehat{A}$ to $\widehat{A}'$. An $\mathcal{A}_\infty$–homotopy H between F and G is a collection

$$H_n \colon \widehat{A} \otimes_{\mathcal{I}} \mathcal{B}^{\otimes(n-1)} \to \widehat{A}'[0,n]$$

of bigrading-preserving $\mathcal{I}$–linear maps satisfying the relation

$$\begin{aligned}F_n - G_n = m'_1 \circ H_n &+ (-1)^{n-1} m'_2 \circ (H_{n-1} \otimes |\mathrm{id}|^{n-1}) \\ &+ (-1)^{n+1} H_n \circ (m_1 \otimes |\mathrm{id}|^{\otimes(n-1)}) + H_{n-1} \circ (m_2 \otimes \mathrm{id}^{\otimes(n-2)}) \\ &+ (-1)^{n+1} \sum_{k=1}^{n-1} H_n \circ (\mathrm{id}^{\otimes k} \otimes \mu_1 \otimes |\mathrm{id}|^{\otimes(n-k-1)}) \\ &+ \sum_{k=1}^{n-2} (-1)^k H_{n-1} \circ (\mathrm{id}^{\otimes k} \otimes \mu_2 \otimes \mathrm{id}^{\otimes(n-k-2)})\end{aligned}$$

for all $n \geq 1$.

6.5.14 Example Suppose H is an $\mathcal{A}_\infty$–homotopy between F and G with $H_n = 0$ for $n > 1$. Then the $n = 1$ $\mathcal{A}_\infty$–homotopy condition for H becomes

$$F_1 - G_1 = m'_1 \circ H_1 + H_1 \circ m_1,$$

and the $n = 2$ homotopy condition becomes

$$F_2 - G_2 = -m'_2 \circ (H_1 \otimes |\mathrm{id}|) + H_1 \circ m_2.$$

For $n > 2$, the homotopy condition is $F_n - G_n = 0$.

We will get homotopies $H = \widehat{A}(\psi)$ from certain homotopies ψ between chain maps f and g from a chain complex M of graded projective right H^n modules to another such complex M'; that is, H^n–linear maps $\psi\colon M \to M'$ of bidegree $(0, -1)$ satisfying

$$f - g = d_{M'}\psi + \psi d_M.$$

We will require that M and M' satisfy C_{module} for some generating sets $\{x_i\}$ and $\{x'_j\}$, and that f and g satisfy C_{morphism} for these generating sets. We will only need to consider homotopies ψ which satisfy the analogue of the more restrictive condition $\widetilde{C}_{\text{morphism}}$ on chain maps.

6.5.15 Definition A homotopy ψ as above satisfies the condition $\widetilde{C}_{\text{homotopy}}$ if, for all generators x_i of M,

$$\psi(x_i) = \sum_j \psi_{i,j} x'_j$$

is a basis expansion of $\psi(x_i)$ in the basis $\{x'_j \cdot h_1\}$ for M', for some integer coefficients $\psi_{i,j}$.

If ψ satisfies the condition $\widetilde{C}_{\text{homotopy}}$ (implying, in particular, that M and M' satisfy the condition C_{module} and f and g satisfy the condition C_{morphism}), then the homotopy relation $f - g = d_{M'}\psi + \psi d_M$ becomes the two sets of equations

$$f_{i,k} - g_{i,k} = \sum_j \psi_{i,j} c'_{j,k} + \sum_j c_{i,j} \psi_{j,k} \tag{6-14}$$

for all generators x_i of M and x'_k of M', and

$$\tilde{f}_{i,k;h'} - \tilde{g}_{i,k;h'} = \sum_j \psi_{i,j} \tilde{c}'_{j,k;h'} + \sum_j \tilde{c}_{i,j;h'} \psi_{j,k} \tag{6-15}$$

for all generators $x_i \in M$, $x'_k \in M'$ and $h' \in \beta_{\text{mult}}$.

6.5.16 Definition Suppose M and M' are chain complexes of graded projective right H^n–modules, satisfying the condition C_{module}, and f and g are chain maps from M to M' satisfying the condition C_{morphism}. Let ψ be an H^n–linear chain homotopy between f and g satisfying the condition $\widetilde{C}_{\mathrm{homotopy}}$ defined above.

Define an $\mathcal{A}_\infty$–homotopy $\widehat{A}(\psi)$ between $\widehat{A}(f)$ and $\widehat{A}(g)$ by

$$\widehat{A}(\psi)_1(X_{x_i\cdot h_1}) := \sum_j \psi_{i,j} X_{x'_j\cdot h_1} \quad \text{and} \quad \widehat{A}(\psi)_n = 0$$

for $n > 1$.

6.5.17 Proposition *$\widehat{A}(\psi)$ is a valid $\mathcal{A}_\infty$–homotopy between $\widehat{A}(f)$ and $\widehat{A}(g)$.*

Proof First, $\widehat{A}(\psi)_1$ respects the right action of the idempotent ring $\mathcal{I}_\beta$ on $\widehat{A}(M)$ and $\widehat{A}(M')$, and $\widehat{A}(\psi)_1$ preserves intrinsic grading and decreases homological grading by one because ψ has the same properties.

By Example 6.5.14, the $n=1$ condition is

$$\widehat{A}(f)_1 - \widehat{A}(g)_1 = m'_1 \circ \widehat{A}(\psi)_1 + \widehat{A}(\psi)_1 \circ m_1.$$

If $X_{x_i\cdot h_1}$ is a generator of $\widehat{A}(M)$, then

$$\begin{aligned}(\widehat{A}(f)_1 - \widehat{A}(g)_1)(X_{x_i\cdot h_1}) &= \sum_k (f_{i,k} - g_{i,k}) X_{x'_k\cdot h_1} \\ &= \sum_{j,k} (\psi_{i,j} c'_{j,k}) X_{x'_k\cdot h_1} + \sum_{j,k} (c_{i,j}\psi_{j,k}) X_{x'_k\cdot h_1}\end{aligned}$$

by (6-14), while

$$\begin{aligned} m'_1 \circ \widehat{A}(\psi)_1 X_{x_i\cdot h_1} &= \sum_{j,k} \psi_{i,j} c'_{j,k} X_{x'_k\cdot h_1}, \\ \widehat{A}(\psi)_1 \circ m_1 X_{x_i\cdot h_1} &= \sum_{j,k} c_{i,j}\psi_{j,k} X_{x'_k\cdot h_1}.\end{aligned}$$

Thus, the $n=1$ condition is satisfied.

By Example 6.5.14, the $n=2$ condition is

$$\widehat{A}(f)_2 - \widehat{A}(g)_2 = -m'_2 \circ (\widehat{A}(\psi)_1 \otimes |\mathrm{id}|) + \widehat{A}(\psi)_1 \circ m_2. \tag{6-16}$$

As in Proposition 6.5.9, we first reduce to the case where the algebra input is one of the generators $b_{*;h_1,h_2}$ or $m(b^*_{*;m(h_1),m(h_2)})$ of $m(\mathcal{B})^! \odot \mathcal{B}$.

Claim *If b_1 and b_2 are two elements of $m(\mathcal{B})^! \odot \mathcal{B}$ such that the $n = 2$ homotopy condition* (6-16) *is satisfied both when the algebra input is b_1 and when it is b_2, then the $n = 2$ homotopy condition is also satisfied when the algebra input is $b_1 b_2$.*

Proof of claim When the algebra input is $b_1 \otimes b_2$, we want to show that the maps

$$\hat{A}(f)_2 \circ (\mathrm{id} \otimes \mu_2) - \hat{A}(g)_2 \circ (\mathrm{id} \otimes \mu_2)$$

and

$$-m_2' \circ (\hat{A}(\psi)_1 \otimes |\mathrm{id}|) \circ (\mathrm{id} \otimes \mu_2) + \hat{A}(\psi)_1 \circ m_2 \circ (\mathrm{id} \otimes \mu_2)$$

take the same value. In the proof of this claim, the algebra input to all maps will be assumed to be $b_1 \otimes b_2$.

By Example 6.5.7 and the associativity of the algebra actions m_2 and m_2', we want to show the following equation (6-17), when the algebra input is $b_1 \otimes b_2$:

$$\begin{aligned} m_2' \circ (\hat{A}(f)_2 \otimes |\mathrm{id}|) + \hat{A}(f)_2 \circ (m_2 \otimes \mathrm{id}) \\ - m_2' \circ (\hat{A}(g)_2 \otimes |\mathrm{id}|) - \hat{A}(g)_2 \circ (m_2 \otimes \mathrm{id}) \\ = -m_2' \circ (m_2' \otimes \mathrm{id}) \circ (\hat{A}(\psi)_1 \otimes |\mathrm{id}| \otimes |\mathrm{id}|) + \hat{A}(\psi)_1 \circ m_2 \circ (m_2 \otimes \mathrm{id}). \end{aligned} \tag{6-17}$$

Call the terms on the left side of (6-17) LHS_1, LHS_2, LHS_3 and LHS_4; call the terms on the right side RHS_1 and RHS_2. Using the $n = 2$ homotopy condition for the algebra input b_1, the term RHS_1 can be written as

$$m_2' \circ (\hat{A}(f)_2 \otimes |\mathrm{id}|) - m_2' \circ (\hat{A}(g)_2 \otimes |\mathrm{id}|) - m_2' \circ (\hat{A}(\psi)_1 \otimes |\mathrm{id}|) \circ (m_2 \otimes \mathrm{id}).$$

Call these terms RHS_{1a}, RHS_{1b} and RHS_{1c}. Using the $n = 2$ homotopy condition for the algebra input b_2, the term RHS_2 can be written as

$$\hat{A}(f)_2 \circ (m_2 \otimes \mathrm{id}) - \hat{A}(g)_2 \circ (m_2 \otimes \mathrm{id}) + m_2' \circ (\hat{A}(\psi)_1 \otimes |\mathrm{id}|) \circ (m_2 \otimes \mathrm{id}).$$

Call these terms RHS_{2a}, RHS_{2b} and RHS_{2c}. Then

- $\mathrm{LHS}_1 = \mathrm{RHS}_{1a}$,
- $\mathrm{LHS}_2 = \mathrm{RHS}_{2a}$,
- $\mathrm{LHS}_3 = \mathrm{RHS}_{1b}$,
- $\mathrm{LHS}_4 = \mathrm{RHS}_{2b}$,
- $\mathrm{RHS}_{1c} + \mathrm{RHS}_{2c} = 0$,

proving the claim. □

It remains to show that the $n = 2$ homotopy condition (6-16) is satisfied when the algebra input is one of the multiplicative generators of $m(\mathcal{B})^! \odot \mathcal{B}$. When the input is $b_{*;h_1,h_2}$, the left side of the $n = 2$ condition is zero, so we want to show that the right side is also zero. If $X_{x_i \cdot h_1}$ is a generator of $\widehat{A}(M)$, then the right side of the $n = 2$ condition with algebra input $b_{*;h_1,h_2}$ is

$$-\sum_j \psi_{i,j} X_{x'_j \cdot h_2} + \sum_j \psi_{i,j} X_{x'_j \cdot h_2},$$

which is zero as desired.

Finally, suppose the algebra input is $m(b^*_{*;m(h_1),m(h_2)})$ and let $X_{x_i \cdot h_1}$ be a generator of $\widehat{A}(M)$. The left side of the $n = 2$ condition applied to these inputs is

$$\sum_{k,h' \in \beta} (\tilde{f}_{i,k;h'} - \tilde{g}_{i,k;h'}) \tilde{\tilde{c}}_{h'h_1;h_2} X_{x'_k \cdot h_2},$$

which equals

$$\sum_{j,k,h' \in \beta} \psi_{i,j} \tilde{c}'_{j,k;h'} \tilde{\tilde{c}}_{h'h_1;h_2} X_{x'_k \cdot h_2} + \sum_{j,k,h' \in \beta} \tilde{c}_{i,j;h'} \psi_{j,k} \tilde{\tilde{c}}_{h'h_1;h_2} X_{x'_k \cdot h_2}$$

by (6-15). This expression is also equal to the right side of the $n = 2$ condition applied to these inputs, since $m(b^*_{m(h_1),m(h_2)})$ has homological degree 1. Thus, the $\mathcal{A}_\infty$–homotopy relations are satisfied for $\widehat{A}(\psi)$. □

6.5.18 Corollary *Let M and M' be chain complexes of graded projective right H^n–modules satisfying the algebraic condition C_{module}. Suppose there exist chain maps $f\colon M \to M'$ and $g\colon M' \to M$ satisfying the condition C_{morphism}, with either f or g satisfying the more restrictive condition $\widetilde{C}_{\mathrm{morphism}}$, and chain homotopies ψ between $g \circ f$ and id_M and ψ' between $f \circ g$ and $\mathrm{id}_{M'}$, both satisfying the condition $\widetilde{C}_{\mathrm{homotopy}}$.*

Then $\widehat{A}(M)$ and $\widehat{A}(M')$ are $\mathcal{A}_\infty$–homotopy equivalent type A structures over $m(\mathcal{B})^! \odot \mathcal{B}$.

Proof By Proposition 6.5.12,

$$\widehat{A}(g) \circ \widehat{A}(f) = \widehat{A}(g \circ f) \quad \text{and} \quad \widehat{A}(f) \circ \widehat{A}(g) = \widehat{A}(f \circ g).$$

By Proposition 6.5.17, $\widehat{A}(\psi)$ provides an $\mathcal{A}_\infty$–homotopy between $\widehat{A}(g) \circ \widehat{A}(f)$ and $\mathrm{id}_{\widehat{A}(M)} = \widehat{A}(\mathrm{id}_M)$, and $\widehat{A}(\psi')$ provides an $\mathcal{A}_\infty$–homotopy between $\widehat{A}(f) \circ \widehat{A}(g)$ and $\mathrm{id}_{\widehat{A}(M')} = \widehat{A}(\mathrm{id}_{M'})$. □

The case of interest to us is when $M = [T]^{\mathrm{Kh}}$ and $M' = [T']^{\mathrm{Kh}}$ for two oriented tangle diagrams T and T' in $\mathbb{R}_{\leq 0} \otimes \mathbb{R}$ which are related by a Reidemeister move. In [4], Khovanov shows that $[T]^{\mathrm{Kh}}$ is chain homotopy equivalent to $[T']^{\mathrm{Kh}}$. In the following two propositions, we verify that the maps involved in these homotopy equivalences satisfy the relevant algebraic conditions.

6.5.19 Proposition *Let (M, d_M) be a chain complex of graded projective right H^n–modules. Assume the following conditions hold:*

(1) *M satisfies the algebraic condition C_{module} with respect to a set of generators $\{x_i\}$.*
(2) *$M \cong M_1 \oplus M_2$ as right H^n–modules; furthermore, M_1 is the submodule of M spanned over H^n by some subset of the x_i, while M_2 is the submodule spanned by the rest of the x_i.*
(3) *M_2 is a subcomplex of M. Write d_M in matrix form with respect to the direct sum decomposition as*

$$d_M = \begin{bmatrix} d_1 & 0 \\ d_{1,2} & d_2 \end{bmatrix}.$$

Note that $d_M^2 = 0$ is equivalent to the equations $d_1^2 = 0$, $d_{1,2} \circ d_1 + d_2 \circ d_{1,2} = 0$ and $d_2^2 = 0$.

(4) *There exists an H^n–linear map $\psi' \colon M_2 \to M_2$ of bidegree $(0, -1)$ with $\mathrm{id}_{M_2} = d_2 \psi' + \psi' d_2$, and such that we may write, with integer coefficients $\psi'_{i,j}$,*

$$\psi'(x_i) = \sum_j \psi'_{i,j} x_j.$$

Among the equations implied by $d_M^2 = 0$ is $d_1^2 = 0$; thus (M_1, d_1) is a chain complex of graded projective right modules over H^n. Since (M, d_M) satisfies C_{module} for the generators $\{x_i\}$, (M_1, d_1) satisfies C_{module} for the appropriate subset of $\{x_i\}$. Define $f \colon (M, d_M) \to (M_1, d_1)$, $g \colon (M_1, d_1) \to (M, d_M)$ and $\psi \colon M \to M$ by the following matrix formulas:

$$f := \begin{bmatrix} \mathrm{id}_{M_1} & 0 \end{bmatrix}, \quad g := \begin{bmatrix} \mathrm{id}_{M_1} \\ -\psi' d_{1,2} \end{bmatrix}, \quad \psi := \begin{bmatrix} 0 & 0 \\ 0 & -\psi' \end{bmatrix}.$$

Then f and g are chain maps, $f \circ g = \mathrm{id}_{M_1}$ and $g \circ f - \mathrm{id}_M = d_M \psi + \psi d_M$. Furthermore, f satisfies the condition $\widetilde{C}_{\text{morphism}}$, g satisfies the condition C_{morphism} and ψ satisfies the condition $\widetilde{C}_{\text{homotopy}}$.

Proof Both f and g are bigrading-preserving and H^n–linear; ψ preserves the intrinsic grading and decreases the homological grading by one, because the same holds for ψ'. The map f is a chain map because it is the projection map onto a quotient complex. To show that g is a chain map, we want to show that $g(d_1(x_i)) = d_M(g(x_i))$ for all x_i in M_1. We have

$$g(d_1(x_i)) = d_1(x_i) - \psi' d_{1,2} \circ d_1(x_i)$$

and

$$\begin{aligned} d_M(g(x_i)) &= d_M(x_i) - d_M(\psi' d_{1,2}(x_i)) \\ &= d_M(x_i) + \psi'(d_2 \circ d_{1,2})(x_i) - d_{1,2}(x_i) \\ &= d_1(x_i) - \psi'(d_{1,2} \circ d_1)(x_i). \end{aligned} \tag{6-18}$$

In the second line of (6-18), we use that

$$\begin{aligned} d_{1,2}(x_i) &= \mathrm{id}_{M_2}(d_{1,2}(x_i)) \\ &= d_2(\psi' d_{1,2}(x_i)) + \psi'(d_2 \circ d_{1,2})(x_i) \\ &= d_M(\psi' d_{1,2}(x_i)) + \psi'(d_2 \circ d_{1,2})(x_i) \end{aligned}$$

and in the third line of (6-18) we use the equation $d_2 \circ d_{1,2} + d_{1,2} \circ d_1 = 0$ from item (3) above. Thus, g is a chain map as well, and by definition, $f \circ g = \mathrm{id}_{M_2}$. To verify that ψ is a homotopy between $g \circ f$ and id_M, we can write out the terms of the relevant equation as matrices:

$$g \circ f = \begin{bmatrix} \mathrm{id}_{M_1} & 0 \\ -\psi' d_{1,2} & 0 \end{bmatrix}, \quad d_M \psi = \begin{bmatrix} 0 & 0 \\ 0 & -d_2 \psi' \end{bmatrix}, \quad \psi d_M = \begin{bmatrix} 0 & 0 \\ -\psi' d_{1,2} & -\psi' d_2 \end{bmatrix}.$$

Thus, the equation $g \circ f - \mathrm{id}_M = d_M \psi + \psi d_M$ holds.

By definition, f satisfies the condition $\widetilde{C}_{\mathrm{morphism}}$. By item (4) above and the condition C_{module} for M, g satisfies C_{morphism}. By item (4), ψ satisfies $\widetilde{C}_{\mathrm{homotopy}}$. □

6.5.20 Proposition *Let M be a chain complex of graded projective right H^n–modules. Assume the following conditions hold:*

1. *M satisfies the algebraic condition C_{module} with respect to a set of generators $\{x_i\}$.*
2. *$M \cong M_1 \oplus M_2$ as right H^n–modules, and M_1 is the submodule of M spanned over H^n by some subset of the x_i, say $M_1 = \{x_i \cdot h_1 \mid i \in S\}$. The submodule M_2 has a $\mathbb{Z}$–basis $\{z_i \cdot h_1 \mid i \notin S\}$, where*

$$z_i = x_i + \sum_{j \in S} \tau_{i,j} x_j + \sum_{j \in S, h' \in \beta_{\mathrm{mult}}} \tilde{\tau}_{i,j;h'} x_j \cdot h'$$

for some integer coefficients $\tau_{i,j}$ and $\tilde{\tau}_{i,j;h'}$.

(3) *M_2 is a subcomplex of M. Write d_M in matrix form with respect to the direct sum decomposition:*
$$d_M = \begin{bmatrix} d_1 & 0 \\ d_{1,2} & d_2 \end{bmatrix}.$$

(4) *There exists an H^n–linear map $\psi'\colon M_2 \to M_2$ of bidegree $(0,-1)$ with $\mathrm{id}_{M_2} = d_2\psi' + \psi' d_2$ such that we may write*
$$\psi'(z_i) = \sum_j \psi'_{i,j} z_j$$
for some integer coefficients $\psi'_{i,j}$.

(5) *Write*
$$d_M(x_i) = \sum_j c_{i,j} x_j + \sum_{j,h' \in \beta_{\mathrm{mult}}} \tilde{c}_{i,j;h'} x_j \cdot h'.$$
For all indices $i \in S$, $j \notin S$, $k \notin S$ and elements h' of β_{mult}, we have $\tilde{c}_{i,j;h'}\psi'_{j,k} = 0$.

(6) *For all indices $i \in S$, $j \notin S$, $k \in S$ and elements h', h'' of β_{mult}, we have $\tilde{c}_{i,j;h'}\tau_{j,k} = 0$ and $\tilde{c}_{i,j;h'}\tilde{\tau}_{j,k;h''} = 0$.*

(7) *For all indices $i \notin S$, $j \notin S$, $k \in S$ and elements h' of β_{mult}, $\psi'_{i,j}\tau_{j,k} = 0$ and $\psi'_{i,j}\tilde{\tau}_{j,k;h'} = 0$.*

As in Proposition 6.5.19, we have $d_1^2 = 0$, so (M_1, d_1) is a chain complex. We may write $d_1 = d_M - d_{1,2}$. The right side of this equation has domain restricted to $M_1 \subset M$, and it takes values in M_1.

Define $f\colon (M, d_M) \to (M_1, d_1)$, $g\colon (M_1, d_1) \to (M, d_M)$ and $\psi\colon M \to M$ by the following matrix formulas:
$$f := \begin{bmatrix} \mathrm{id}_{M_1} & 0 \end{bmatrix}, \quad g := \begin{bmatrix} \mathrm{id}_{M_1} \\ -\psi' d_{1,2} \end{bmatrix}, \quad \psi := \begin{bmatrix} 0 & 0 \\ 0 & -\psi' \end{bmatrix}.$$

Then f and g are chain maps, $f \circ g = \mathrm{id}_{M_1}$ and $g \circ f - \mathrm{id}_M = d_M\psi + \psi d_M$. Furthermore, (M_1, d_1) satisfies the condition C_{module} for the generators $\{x_i \mid i \in S\}$, f satisfies C_{morphism}, g satisfies $\widetilde{C}_{\mathrm{morphism}}$ and ψ satisfies $\widetilde{C}_{\mathrm{homotopy}}$. These last three conditions use the generators $\{x_i \mid i \in S\} \cup \{x_j \mid j \notin S\}$ for M and $\{x_i \mid i \in S\}$ for M_1.

Proof The proof that f and g are chain maps, and that ψ is a homotopy between f and g, is the same as in Proposition 6.5.19. Note that here, all the matrices are chosen with respect to the basis $\{x_i \cdot h_1 \mid i \in S\} \cup \{z_j \cdot h_1 \mid j \notin S\}$ of M, since this basis is compatible with the direct sum decomposition $M \cong M_1 \oplus M_2$. For the algebraic conditions, we need to use the basis $\{x_i \cdot h_1 \mid i \in S\} \cup \{x_j \cdot h_1 \mid j \notin S\}$ instead, which is not compatible with the direct sum decomposition. Thus, we want to express d_1, f, g and ψ in this basis and show they satisfy C_{module}, C_{morphism}, $\widetilde{C}_{\text{morphism}}$ and $\widetilde{C}_{\text{homotopy}}$, respectively.

For $i \in S$, we may write

$$\begin{aligned} d_M(x_i) = \sum_{k \in S} c_{i,k} x_k + \sum_{k \in S, h' \in \beta_{\text{mult}}} \tilde{c}_{i,k;h'} x_k \cdot h' + \sum_{j \notin S} c_{i,j} z_j \\ - \sum_{j \notin S, k \in S} c_{i,j} \tau_{j,k} x_k - \sum_{j \notin S, k \in S, h' \in \beta_{\text{mult}}} c_{i,j} \tilde{\tau}_{j,k;h'} x_k \cdot h' \\ + \sum_{j \notin S, h' \in \beta_{\text{mult}}} \tilde{c}_{i,j;h'} z_j \cdot h'. \end{aligned}$$

The final two terms which would appear in this expression are zero by item (6) of the above assumptions. The third and sixth terms of this expression make up $d_{1,2}(x_i)$:

$$d_{1,2}(x_i) = \sum_{j \notin S} c_{i,j} z_j + \sum_{j \notin S, h' \in \beta_{\text{mult}}} \tilde{c}_{i,j;h'} z_j \cdot h'.$$

The rest of the terms make up $d_1(x_i)$:

$$\begin{aligned} d_1(x_i) = \sum_{k \in S} c_{i,k} x_k + \sum_{k \in S, h' \in \beta_{\text{mult}}} \tilde{c}_{i,k;h'} x_k \cdot h' \\ - \sum_{j \notin S, k \in S} c_{i,j} \tau_{j,k} x_k - \sum_{j \notin S, k \in S, h' \in \beta_{\text{mult}}} c_{i,j} \tilde{\tau}_{j,k;h'} x_k \cdot h'. \end{aligned}$$

From this formula we can see that (M_1, d_1) satisfies the condition C_{module} for the generators $\{x_i \mid i \in S\}$.

For x_i with $i \in S$, we have $f(x_i) = x_i$. For z_j with $j \notin S$, we have $f(z_j) = 0$. We may write this equation as

$$f\left(x_j + \sum_{k \in S} \tau_{j,k} x_k + \sum_{k \in S, h' \in \beta_{\text{mult}}} \tilde{\tau}_{j,k;h'} x_k \cdot h'\right) = 0,$$

or equivalently

$$f(x_j) = -\sum_{k \in S} \tau_{j,k} x_k - \sum_{k \in S, h' \in \beta_{\text{mult}}} \tilde{\tau}_{j,k;h'} x_k \cdot h'.$$

By this formula, we see that f satisfies condition C_{morphism}.

For the map g, we can write

$$\psi' \circ d_{1,2}(x_i) = \psi'\Bigg(\sum_{j \notin S} c_{i,j} z_j + \sum_{j \notin S, h' \in \beta_{\text{mult}}} \tilde{c}_{i,j;h'} z_j \cdot h'\Bigg) = \sum_{j \notin S, k \notin S} c_{i,j} \psi'_{j,k} z_k$$

by (5) above. Using (7), we can write this sum as

$$\sum_{j \notin S, k \notin S} c_{i,j} \psi'_{j,k} x_k.$$

Thus,

$$g(x_i) = x_i - \sum_{j \notin S, k \notin S} c_{i,j} \psi'_{j,k} x_k$$

and g satisfies condition $\widetilde{C}_{\text{morphism}}$.

Finally, if $x_i \in S$, then $\psi(x_i) = 0$ by definition. For $x_j \notin S$, we have

$$\begin{aligned}\psi(x_j) &= \psi\Bigg(z_j - \sum_{k \in S} \tau_{j,k} x_k - \sum_{k \in S, h' \in \beta} \tilde{\tau}_{j,k;h'} x_k \cdot h'\Bigg) \\ &= \psi(z_j) = -\sum_{k \notin S} \psi'_{j,k} z_k = -\sum_{k \notin S} \psi'_{j,k} x_k.\end{aligned}$$

The last equality follows from (7). Thus, ψ satisfies the condition $\widetilde{C}_{\text{homotopy}}$. □

6.5.21 Corollary *If T and T' are oriented tangle diagrams in $\mathbb{R}_{\leq 0} \otimes \mathbb{R}$ which are related by a Reidemeister move, then $\mathcal{A}([T]^{\text{Kh}})$ and $\mathcal{A}([T']^{\text{Kh}})$ are $\mathcal{A}_\infty$–homotopy equivalent as type A structures over $m(\mathcal{B})^! \odot \mathcal{B}$.*

Proof When T and T' are related by an R1 move, Khovanov's homotopy equivalence between $[T]^{\text{Kh}}$ and $[T']^{\text{Kh}}$ from [4, Section 4.4] is of the type constructed in Proposition 6.5.19. This is most easily seen by looking at the top diagram of Figure 7. The map ψ' sends a generator x_i of the far-right rectangle to the corresponding generator x_j in the top rectangle, which topologically is x_i isotoped with a plus-labeled free circle added, times $(-1)^{\#_1(i,j)}$. One can check that the four conditions of Proposition 6.5.19 are satisfied.

When T and T' are related by an R2 move, Khovanov's homotopy equivalence from [4, Section 4.5] is a composition of a homotopy equivalence from Proposition 6.5.19 followed by one from Proposition 6.5.20. The relevant diagrams are the middle diagram of Figure 7 and the top diagram of Figure 8. The first homotopy equivalence is very similar to the R1 move, with ψ' defined analogously. For the second homotopy equivalence, ψ' isotopes a generator and deletes a minus-labeled free circle; it still

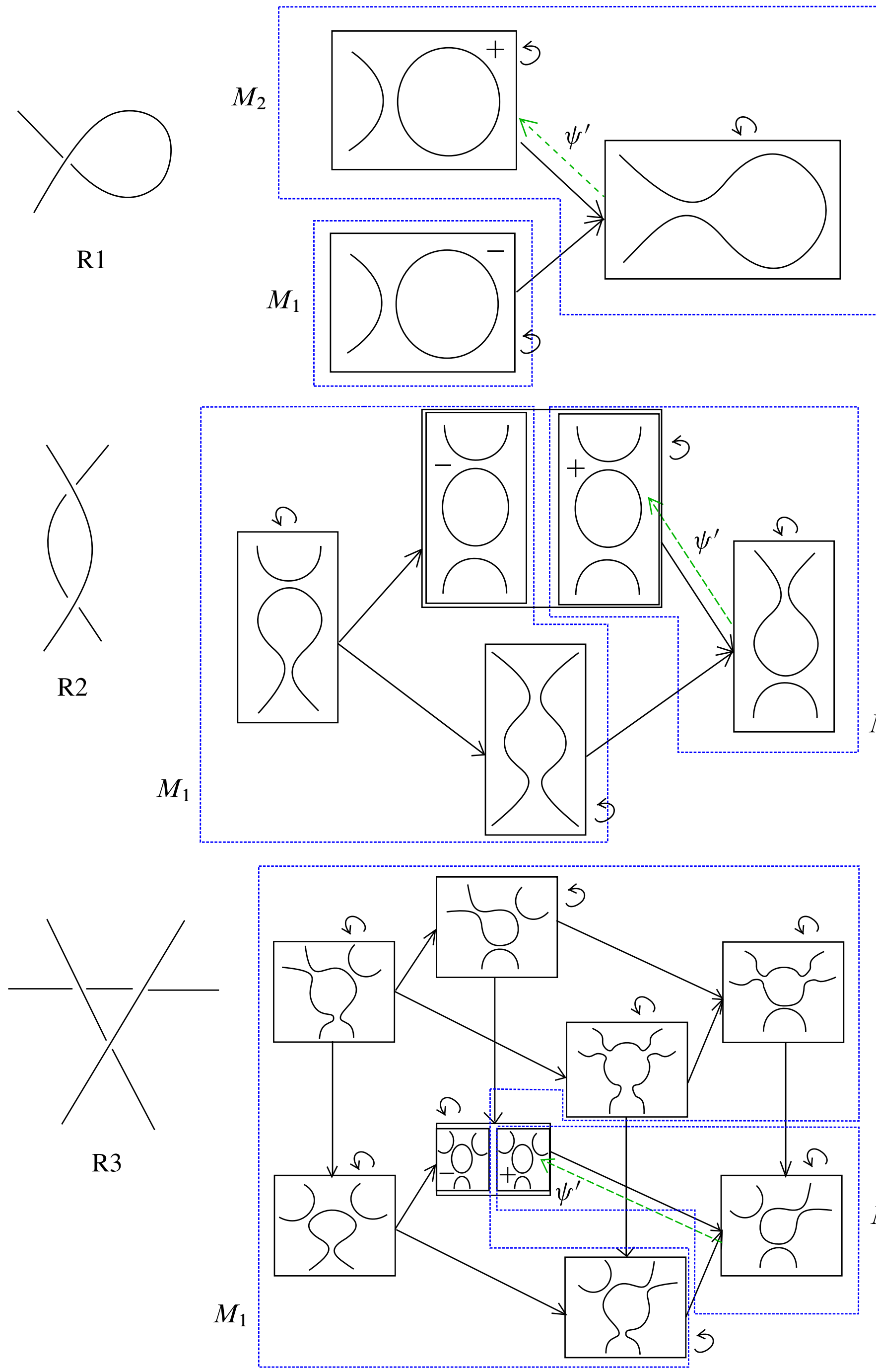

Figure 7: First step of R1, R2 and R3 moves

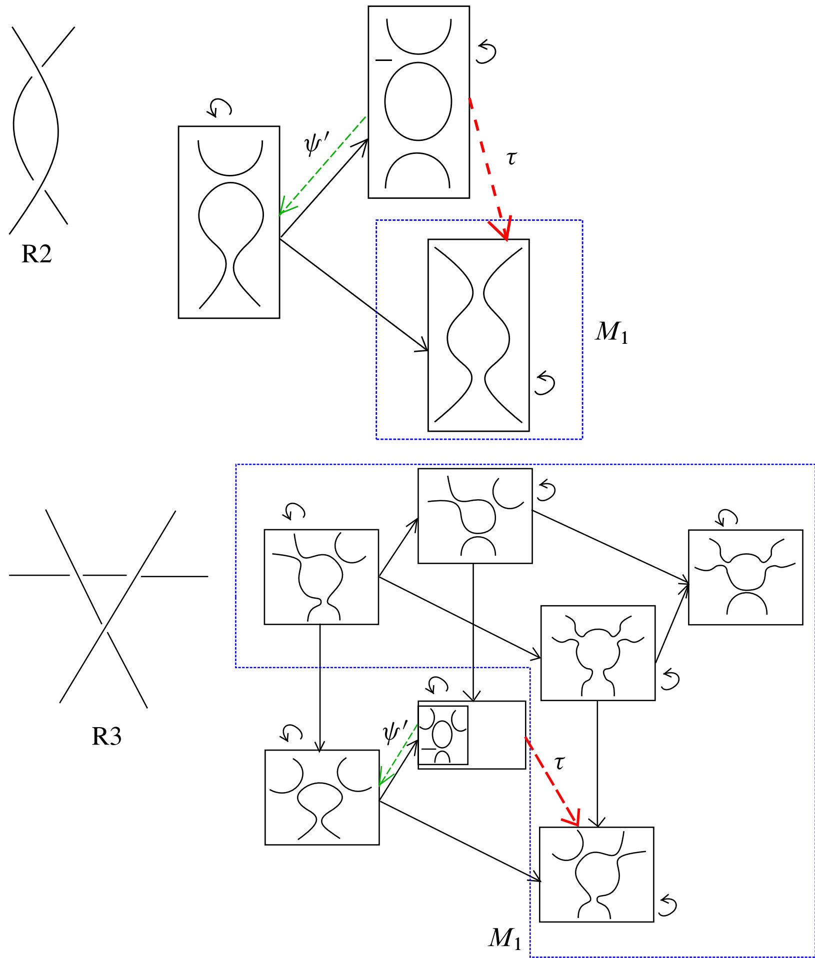

Figure 8: Second step of R2 and R3 moves

carries a coefficient of $(-1)^{\#_1(i,j)}$. The coefficients $\tau_{i,j}$ and $\tilde{\tau}_{i,j;h'}$ may be packaged into a map τ as shown in the diagram. This map is defined as ψ' postcomposed with the differential map from the left rectangle to the bottom rectangle.

For generators x_i in the left rectangle, z_i is x_i. For generators x_i in the top rectangle, z_i is $x_i + \tau(x_i)$; then M_2 is the subcomplex of $[T]^{\mathrm{Kh}}$ generated by the z_i. One can see from the diagram in Figure 8 that the conditions of Proposition 6.5.20 are satisfied. In particular, note that $\psi'(z_i) = \psi'(x_i)$ because ψ' is zero on generators from the bottom rectangle, so condition (4) is satisfied. We have $d_{1,2} = 0$, so conditions (5) and (6) hold automatically. Condition (7) holds because the arrows labeled ψ' and τ are not composable in the diagram.

Finally, when T and T' are related by an R3 move, Khovanov's homotopy equivalence from [4, Section 4.6] comes from doing a homotopy equivalence from Proposition 6.5.19 and then a homotopy equivalence from Proposition 6.5.20, to both $[T]^{\mathrm{Kh}}$ and $[T']^{\mathrm{Kh}}$. After these homotopy equivalences, the complexes are isomorphic. Thus, the full homotopy equivalence from $[T]^{\mathrm{Kh}}$ to $[T']^{\mathrm{Kh}}$ is a composition of four homotopy equivalences from Proposition 6.5.19 and Proposition 6.5.20.

The relevant diagrams are the bottom diagram of Figure 7 and the bottom diagram of Figure 8. The maps ψ' and τ are defined as in the $R2$ move. Again, one can check that the conditions of Proposition 6.5.19 are satisfied using Figure 7 and that the conditions of Proposition 6.5.20 are satisfied using Figure 8. This time, in the second step, $d_{1,2}$ is not zero. However, for one of the two arrows of Figure 8 contributing to $d_{1,2}$, all $\tilde{c}_{i,j;h'}$ are zero. The other arrow is not composable with the arrows labeled ψ' and τ. This suffices to show that conditions (5) and (6) of Proposition 6.5.20 hold. □

6.5.22 Proposition *If T and T' are oriented tangle diagrams in $\mathbb{R}_{\leq 0} \otimes \mathbb{R}$ which are related by a Reidemeister move, then the $\mathcal{A}_\infty$–homotopy equivalence between $\widehat{A}([T]^{\mathrm{Kh}})$ and $\widehat{A}([T']^{\mathrm{Kh}})$ of Corollary 6.5.21 descends to a $\mathcal{A}_\infty$–homotopy equivalence of type A structures over the quotient algebra $\mathcal{B}\Gamma_n$ of $m(\mathcal{B})^! \odot \mathcal{B}$.*

Proof For homotopy equivalences $\{f\colon M \to M_1, g\colon M_1 \to M, \psi\colon M \to M\}$ coming from Proposition 6.5.19, when doing an R1 move, the first step of an R2 move, or the first or fourth step of an R3 move, we only need to show that $\widehat{A}(g)_2$ descends from a map

$$\widehat{A}(M_1) \otimes_{\mathcal{I}_\mathcal{B}} m(\mathcal{B})^! \odot \mathcal{B} \to \widehat{A}(M)$$

to a map

$$\widehat{A}(M_1) \otimes_{\mathcal{I}_\mathcal{B}} \mathcal{B}\Gamma_n \to \widehat{A}(M);$$

since f satisfies $\widetilde{C}_{\text{morphism}}$ and ψ satisfies $\widetilde{C}_{\text{homotopy}}$, we have $\widehat{A}(f)_2 = 0$ and $\widehat{A}(\psi)_2 = 0$.

As in Proposition 5.2.8, let a, b, c and d be vertices of a tetrahedron in the graph G. We want to show that $\widehat{A}(g)_2(-, a+c) = 0$, $\widehat{A}(g)_2(-, a+d) = 0$ and $\widehat{A}(g)_2(-, b+c) = 0$. We will show only that $\widehat{A}(g)_2(-, a+c) = 0$; by symmetry, the proof is the same for the other two extra relations.

Write

$$a = m(b^*_{\gamma;m(h_1),m(h_2)})m(b^*_{\eta';m(h_2),m(h_3)}),$$
$$c = m(b^*_{\eta;m(h_1),m(\tilde{\tilde{h}}_2)})m(b^*_{\gamma';m(\tilde{\tilde{h}}_2),m(h_3)}).$$

Let $X_{x_i \cdot h_1}$ be a generator of $\widehat{A}(M_1)$; we have $i \in S$, in the notation of Proposition 6.5.19. Then

$$\begin{aligned} \text{(6-19)} \quad & \widehat{A}(g)_2(X_{x_i \cdot h_1}, a) \\ &= (\widehat{A}(g)_2 \circ (m_2 \otimes \mathrm{id}))(X_{x_i \cdot h_1}, m(b^*_{\gamma;m(h_1),m(h_2)}) \otimes m(b^*_{\eta';m(h_2),m(h_3)})) \\ &\quad + (m_2' \circ (\widehat{A}(g)_2 \otimes |\mathrm{id}|))(X_{x_i \cdot h_1}, m(b^*_{\gamma;m(h_1),m(h_2)}) \otimes m(b^*_{\eta';m(h_2),m(h_3)})) \end{aligned}$$

using the $n=3$ consistency condition for the $\mathcal{A}_\infty$–morphism $\widehat{A}(g)_2$; see Example 6.5.3. The first term on the right side of (6-19) can be expanded out as

$$\text{(6-20)} \qquad -\sum_{i,j \in S, k,l \notin S} \tilde{c}_{i,j;h'_\gamma} \tilde{c}_{j,k;h''_{\eta'}} \psi'_{k,l} X_{x_l \cdot h_3},$$

where h'_γ and $h''_{\eta'}$ are determined by γ and η', while the second term on the right side of (6-19) can be expanded out as

$$\text{(6-21)} \qquad \sum_{i \in S, j,k,l \notin S} \tilde{c}_{i,j;h'_\gamma} \psi'_{j,k} \tilde{c}_{k,l;h''_{\eta'}} X_{x_l \cdot h_3}.$$

Similarly, we may write $\widehat{A}(g)_2(X_{x_i \cdot h_1}, c)$ as the sum of the expressions

$$\text{(6-22)} \qquad -\sum_{i,j \in S, k,l \notin S} \tilde{c}_{i,j;h'_\eta} \tilde{c}_{j,k;h''_{\gamma'}} \psi'_{k,l} X_{x_l \cdot h_3}$$

and

$$\text{(6-23)} \qquad \sum_{i \in S, j,k,l \notin S} \tilde{c}_{i,j;h'_\eta} \psi'_{j,k} \tilde{c}_{k,l;h''_{\gamma'}} X_{x_l \cdot h_3}.$$

We want to show that the expressions (6-20) and (6-23) sum to zero; the argument that expressions (6-21) and (6-22) sum to zero is very similar.

Indeed, generators of all complexes (M, d_M) and (M_1, d_1) under consideration come from generators of the Khovanov complex $[T]^{\mathrm{Kh}}$ of a tangle T, and by Remark 3.0.4 we may choose any ordering we like for the crossings of T. We will order the crossings of T such that the one, two or three crossings local to the Reidemeister move being performed come first in the ordering.

Now, to each quadruple $(i \in S, j \in S, k \notin S, l \notin S)$ giving rise to a nonzero term of expression (6-20), we may associate a pair of indices $(j' \notin S, k' \notin S)$, such that $\tilde{c}_{i,j';h'_\eta} \tilde{c}_{j',k';h''_{\gamma'}} \psi'_{k',l} \neq 0$. In fact, with the above ordering convention, we will have

$$\tilde{c}_{i,j;h'_\gamma} \tilde{c}_{j,k;h''_{\eta'}} \psi'_{k,l} = \tilde{c}_{i,j';h'_\eta} \psi'_{j',k'} \tilde{c}_{k',l;h''_{\gamma'}}.$$

To construct $(j' \notin S, k' \notin S)$ such that the above equation holds, first note that the only component of g relevant for $\widehat{A}(g)_2$ is $-\psi' \circ d_{1,2}$. Recall that the terms $\tilde{c}$, and thus also the terms ψ', are computed as in Example 6.1.3. A term like $\tilde{c}_{i,j;h'_\gamma}\tilde{c}_{j,k;h''_{\eta'}}\psi'_{k,l}$ corresponds to doing one step of d_1 by changing some crossing away from the local area (and thus higher in the ordering) from 0 to 1, then doing one step of $d_{1,2}$ by changing one of the local crossings from 0 to 1, and then finally doing one step of ψ' by changing a different local crossing from 1 to 0. The indices (j', k') and the corresponding term $\tilde{c}_{i,j';h'_\eta}\psi'_{j',k'}\tilde{c}_{k',l;h''_{\gamma'}}$ come from doing the $d_{1,2}$ and $-\psi'$ steps before changing the nonlocal crossing from 0 to 1. But, when changing the local crossings, the signs are the same for both terms because the local crossings occur at the beginning of the ordering. When changing the nonlocal crossing, the signs are also the same for both terms because doing $d_{1,2}$ and $-\psi$ on the local crossings does not increase or decrease the number of crossings with a 1–resolution ($d_{1,2}$ increases this number by 1 and then $-\psi'$ decreases it by 1).

The correspondence between quadruples $(i \in S, j \in S, k \notin S, l \notin S)$ such that $\tilde{c}_{i,j;h'_\gamma}\tilde{c}_{j,k;h''_{\eta'}}\psi'_{k,l}$ is nonzero and quadruples $(i \in S, j' \notin S, k' \notin S, l \notin S)$ such that $\tilde{c}_{i,j';h'_\eta}\psi'_{j',k'}\tilde{c}_{k',l;h''_{\gamma'}}$ is nonzero is bijective. Thus, expressions (6-20) and (6-23) sum to zero. Analogously, expressions (6-21) and (6-22) sum to zero.

We conclude that $\widehat{A}(g)_2(-, a+c) = 0$. By symmetry, we also have $\widehat{A}(g)_2(-, a+d) = 0$ and $\widehat{A}(g)_2(-, b+c) = 0$, so $\widehat{A}(g)$ descends to an $\mathcal{A}_\infty$–morphism of type A structures over $\mathcal{B}\Gamma_n$.

For homotopy equivalences $\{f\colon M \to M_1, g\colon M_1 \to M, \psi\colon M \to M\}$ coming from Proposition 6.5.20, the argument is similar enough that we will simply outline the differences with the above proof. Homotopy equivalences from Proposition 6.5.20 arise when doing the second step of an R2 move or the second or third step of an R3 move. For these equivalences, we only need to show that $\widehat{A}(f)_2$ descends from a map

$$\widehat{A}(M) \otimes_{\mathcal{I}_\beta} m(\mathcal{B})^! \odot \mathcal{B} \to \widehat{A}(M_1)$$

to a map

$$\widehat{A}(M) \otimes_{\mathcal{I}_\beta} \mathcal{B}\Gamma_n \to \widehat{A}(M_1)$$

since we automatically have $\widehat{A}(g)_2 = 0$ by condition $\widetilde{C}_{\text{morphism}}$ on g and $\widehat{A}(\psi)_2 = 0$ by condition $\widetilde{C}_{\text{homotopy}}$ on ψ.

The only terms of f in the basis expansion $\{x_i \cdot h_1 \mid i \in S\} \cup \{x_j \cdot h_1 \mid j \notin S\}$ of M which are relevant for $\widehat{A}(f)_2$ are the terms with coefficients $-\tilde{\tau}_{j,k;h'}$; see the proof of Proposition 6.5.20. These τ terms play a role analogous to the $-\psi' \circ d_{1,2}$ terms in the proof above for homotopy equivalences from Proposition 6.5.19. Indeed, a τ term corresponds to doing one step of ψ', by changing a local crossing from a 1 to a 0, and then doing one step of d_M, by changing a local crossing from a 0 to a 1. Thus, an argument analogous to the one above shows that $\widehat{A}(f)_2(-, a+c) = 0$, and by symmetry that $\widehat{A}(f)_2(-, a+d) = 0$ and $\widehat{A}(f)_2(-, b+c) = 0$. Hence $\widehat{A}(f)$ descends to an $\mathcal{A}_\infty$–morphism of type A structures over $\mathcal{B}\Gamma_n$. □

Proposition 6.5.22 gives us an alternate proof of Roberts [11, Corollary 33].

6.6 Equivalences of type D structures

We first define morphisms and homotopies of type D structures with sign conventions following Roberts [12, Definition 37].

6.6.1 Definition Let $\mathcal{B}$ be a differential bigraded algebra with idempotent ring $\mathcal{I}$. Let $(\widehat{D}, \delta)$ and $(\widehat{D}', \delta')$ be type D structures over $\mathcal{B}$. A *morphism of type D structures* $F\colon \widehat{D} \to \widehat{D}'$ is a bigrading-preserving $\mathcal{I}$–linear map $F\colon \widehat{D} \to \mathcal{B} \otimes_{\mathcal{I}} \widehat{D}'$ satisfying the type D morphism relation

$$(\mu_1 \otimes |\mathrm{id}|) \circ F = (\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes F) \circ \delta - (\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta') \circ F.$$

The composition of two morphisms of type D structures $F\colon \widehat{D} \to \widehat{D}'$ and $G\colon \widehat{D}' \to \widehat{D}''$ is

$$G \circ F := (\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes G) \otimes F,$$

a bigrading-preserving $\mathcal{I}$–linear map from $\widehat{D}$ to $\mathcal{B} \otimes_{\mathcal{I}} \widehat{D}''$ which also satisfies the type D morphism relation.

6.6.2 Definition Let $F\colon \widehat{D} \to \widehat{D}'$, $G\colon \widehat{D} \to \widehat{D}'$ be morphisms of type D structures over $\mathcal{B}$. A *homotopy of morphisms of type D structures* between F and G is a bigrading-preserving $\mathcal{I}$–linear map $H\colon \widehat{D} \to (\mathcal{B} \otimes_{\mathcal{I}} \widehat{D}')[0, 1]$ satisfying

$$F - G = (\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes H) \circ \delta + (\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta') \circ H + (\mu_1 \otimes |\mathrm{id}|) \circ H.$$

If a homotopy exists between F and G, then F is said to be *homotopic* to G.

Two type D structures $\widehat{D}$ and $\widehat{D}'$ are *homotopy equivalent* if there exist type D structure morphisms $F\colon \widehat{D} \to \widehat{D}'$ and $G\colon \widehat{D}' \to \widehat{D}$, such that $G \circ F$ is homotopic to $\mathrm{id}_{\widehat{D}}$ and $F \circ G$ is homotopic to $\mathrm{id}_{\widehat{D}'}$.

6.6.3 Remark Suppose $\widehat{D}$ and $\widehat{D}'$ are homotopy equivalent type D structures over $\mathcal{B}$ and J is a bigrading-homogeneous ideal of $\mathcal{B}$ which is closed under the differential on $\mathcal{B}$. By Proposition 5.3.7, $\widehat{D}$ and $\widehat{D}'$ induce type D structures over $\mathcal{B}/J$. The induced type D structures are homotopy equivalent. Indeed, one may simply postcompose the algebra outputs of F, G and H with the projection map from $\mathcal{B}$ onto $\mathcal{B}/J$, and all the relevant conditions are still satisfied.

Now let N be a chain complex of graded projective left H^n–modules. In Section 6.3, the type D structure $\widehat{D}(N)$ over $m(\mathcal{B})^! \odot \mathcal{B}$ was defined by applying the mirroring operation of Definition 6.3.6 to $\widehat{A}(r(N)) \boxtimes {}^{m(\mathcal{B})^! \odot \mathcal{B}} K^{m(m(\mathcal{B})^! \odot \mathcal{B})^{\mathrm{op}}}$. One can check that the mirroring operation of Definition 6.3.6 respects homotopy equivalences of type D structures. Thus, to show that $\widehat{D}([T]^{\mathrm{Kh}})$ is a tangle invariant up to homotopy equivalence, it would suffice to prove the following general result: if $\widehat{A}$ and $\widehat{A}'$ are type A structures over a differential bigraded algebra $\mathcal{B}$ which are free as $\mathbb{Z}$–modules, $\widehat{DD}$ is a type DD bimodule over $\mathcal{B}$ and another differential bigraded algebra $\mathcal{B}'$, and $\widehat{A}$ and $\widehat{A}'$ are $\mathcal{A}_\infty$–homotopy equivalent, then $\widehat{A} \boxtimes \widehat{DD}$ and $\widehat{A}' \boxtimes \widehat{DD}$ are homotopy equivalent as type D structures over $\mathcal{B}'$. Over $\mathbb{Z}/2\mathbb{Z}$, this is a standard property of the box tensor product; see [7, Lemma 2.3.13]. Here, we are working over $\mathbb{Z}$, but we will only need a simpler version of this result.

6.6.4 Definition Let $\mathcal{B}$ and $\mathcal{B}'$ be differential bigraded algebras over an idempotent ring $\mathcal{I}$. Let $\widehat{A}$ and $\widehat{A}'$ be differential bigraded right modules over $\mathcal{B}$ and let $(\widehat{DD}, \delta_{DD})$ be a rank-one type DD bimodule over $\mathcal{B}$ and $\mathcal{B}'$. Assume that $\widehat{A}$ and $\widehat{A}'$ are free as $\mathbb{Z}$–modules, with $\mathbb{Z}$–bases consisting of elements which are bigrading-homogeneous and have a unique right idempotent. Let δ and δ' denote the type D structure operations on $\widehat{A} \boxtimes \widehat{DD}$ and $\widehat{A}' \boxtimes \widehat{DD}$ respectively.

Let $F\colon \widehat{A} \to \widehat{A}'$ be an $\mathcal{A}_\infty$–morphism with $F_n = 0$ for $n > 2$. Define a morphism of type D structures $F \boxtimes \mathrm{id}_{DD}$ from $\widehat{A} \boxtimes \widehat{DD}$ to $\widehat{A}' \boxtimes \widehat{DD}$, or in other words a map

$$F \boxtimes \mathrm{id}_{DD}\colon (\widehat{A} \boxtimes \widehat{DD}) \to \mathcal{B}' \otimes_{\mathcal{I}} (\widehat{A}' \boxtimes \widehat{DD}),$$

by the formula

$$F \boxtimes \mathrm{id}_{DD} := 1 \otimes F_1 + \xi \circ (F_2 \otimes |\mathrm{id}|) \circ (\mathrm{id} \otimes \delta_{DD}),$$

where we are identifying $\widehat{A} \boxtimes \widehat{DD}$ with $\widehat{A}$ and $\widehat{A}' \boxtimes \widehat{DD}$ with $\widehat{A}'$. Recall that

$$\xi\colon \widehat{A} \otimes_{\mathcal{I}} (\mathcal{B}')^{\mathrm{op}} \to \mathcal{B}' \otimes_{\mathcal{I}} \widehat{A}$$

was defined in Definition 6.3.4 and

$$\xi\colon \widehat{A}' \otimes_{\mathcal{I}} (\mathcal{B}')^{\mathrm{op}} \to \mathcal{B}' \otimes_{\mathcal{I}} \widehat{A}'$$

is defined analogously. The map $F \boxtimes \mathrm{id}_{DD}$ is bigrading-preserving and respects the actions of $\mathcal{I}$ on $\widehat{A} \boxtimes \widehat{DD}$ and $\widehat{A}' \boxtimes \widehat{DD}$.

6.6.5 Proposition *The map $F \boxtimes \mathrm{id}_{DD}$ defined in Definition 6.6.4 is a morphism of type D structures from $\widehat{A} \boxtimes \widehat{DD}$ to $\widehat{A}' \boxtimes \widehat{DD}$.*

Proof We want to show that

$$(\mu_1 \otimes |\mathrm{id}|) \circ (F \boxtimes \mathrm{id}_{DD}) = (\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes (F \boxtimes \mathrm{id}_{DD})) \circ \delta - (\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta') \circ (F \boxtimes \mathrm{id}_{DD}). \tag{6-24}$$

The left side of (6-24) is

$$(\mu_1 \otimes |\mathrm{id}|) \circ \xi \circ (F_2 \otimes |\mathrm{id}|) \circ (\mathrm{id} \otimes \delta_{DD}) = -\xi \circ (F_2 \otimes |\mathrm{id}|) \circ (\mathrm{id} \otimes \mathrm{id} \otimes \mu_1) \circ (\mathrm{id} \otimes \delta_{DD}).$$

Using the DD bimodule relations for δ_{DD}, we may further rewrite this term as

$$\begin{aligned}&\xi \circ (F_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \mu_1 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD})\\ &\quad + \xi \circ (F_2 \otimes |\mathrm{id}|) \circ (\mathrm{id} \otimes \mu_2 \otimes \mu_2) \circ (\mathrm{id} \otimes \sigma) \circ (\mathrm{id} \otimes \mathrm{id} \otimes \delta_{DD} \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD}).\end{aligned}$$

Using the $n = 2$ and $n = 3$ $\mathcal{A}_\infty$ consistency conditions for F from Example 6.5.3, the sum of these two terms is

$$\begin{aligned}&\xi \circ (F_1 \otimes \mathrm{id}) \circ (m_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD})\\ &\qquad - \xi \circ (F_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes |\mathrm{id}| \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD}) \circ m_1\\ &\qquad - \xi \circ (m_1' \otimes \mathrm{id}) \circ (F_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD})\\ &\qquad - \xi \circ (m_2' \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD}) \circ F_1\\ &\qquad + \xi \circ (F_2 \otimes |\mathrm{id}|) \circ (m_2 \otimes \mathrm{id} \otimes \mathrm{id})\\ &\qquad\qquad \circ (\mathrm{id} \otimes \mathrm{id} \otimes \mathrm{id} \otimes \mu_2) \circ \sigma \circ (\mathrm{id} \otimes \mathrm{id} \otimes \delta_{DD} \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD})\\ &\qquad + (|\mathrm{id}| \otimes \mathrm{id}) \circ \xi \circ (m_2' \otimes \mathrm{id}) \circ (F_2 \otimes |\mathrm{id}| \otimes \mathrm{id})\\ &\qquad\qquad\qquad \circ (\mathrm{id} \otimes \mathrm{id} \otimes \mathrm{id} \otimes \mu_2) \circ \sigma \circ (\mathrm{id} \otimes \mathrm{id} \otimes \delta_{DD} \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD}).\end{aligned}$$

We will refer to these six terms as LHS_1, LHS_2, LHS_3, LHS_4, LHS_5 and LHS_6.

The right side of (6-24) is

$$\begin{aligned}1 \otimes (F_1 \circ m_1) &+ \xi \circ (F_2 \otimes |\mathrm{id}|) \circ (\mathrm{id} \otimes \delta_{DD}) \circ m_1\\ &+ (\mathrm{id} \otimes F_1) \circ \xi \circ (m_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD})\\ &+ (\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \xi) \circ (\mathrm{id} \otimes F_2 \otimes |\mathrm{id}|) \circ (\mathrm{id} \otimes \mathrm{id} \otimes \delta_{DD})\\ &\qquad \circ \xi \circ (m_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD})\\ &- 1 \otimes (m_1' \circ F_1)\\ &- \xi \circ (m_2' \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD}) \circ F_1\\ &- (\mathrm{id} \otimes m_1') \circ \xi \circ (F_2 \otimes |\mathrm{id}|) \circ (\mathrm{id} \otimes \delta_{DD})\\ &- (\mu_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \xi) \circ (\mathrm{id} \otimes m_2' \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \mathrm{id} \otimes \delta_{DD})\\ &\qquad \circ \xi \circ (F_2 \otimes |\mathrm{id}|) \circ (\mathrm{id} \otimes \delta_{DD}).\end{aligned}$$

We will refer to these eight terms as RHS_1 through RHS_8. We have the following:

- $\mathrm{RHS}_1 + \mathrm{RHS}_5 = 0$ by the $n=1$ consistency conditions for F.
- $\mathrm{LHS}_1 = \mathrm{RHS}_3$ because F_1 is bigrading-preserving.
- $\mathrm{LHS}_2 = \mathrm{RHS}_2$ since
$$(\mathrm{id} \otimes |\mathrm{id}|) \circ \delta_{DD} = -(|\mathrm{id}| \otimes \mathrm{id}) \circ \delta_{DD}.$$
- $\mathrm{LHS}_3 = \mathrm{RHS}_7$.
- $\mathrm{LHS}_4 = \mathrm{RHS}_6$.

It remains to show that $\mathrm{LHS}_5 = \mathrm{RHS}_4$ and that $\mathrm{LHS}_6 = \mathrm{RHS}_8$. These claims follow from direct computation: let $\delta_{DD}(1) = \sum_i b_i \otimes (b_i')^{\mathrm{op}}$. The term LHS_5, when applied to a generator X of $\widehat{A}$, gives

$$\begin{aligned}\sum_{i,j} (-1)^{(\deg_h b_i')(\deg_h b_j)} (-1)^{\deg_h b_i' + \deg_h b_j'} (-1)^{(\deg_h X + \deg_h b_i + \deg_h b_j - 1)(\deg_h b_i' + \deg_h b_j')} \\ \cdot b_i' b_j' \otimes F_2(X b_i, b_j) \\ = \sum_{i,j} (-1)^{(\deg_h X)(\deg_h b_i' + \deg_h b_j') + (\deg_h b_i)(\deg_h b_j')} b_i' b_j' \otimes F_2(X b_i, b_j).\end{aligned}$$

To see that the second sum is equal to the first, use the fact that $\deg_h b_i + \deg_h b_i' = 1$ and $\deg_h b_j + \deg_h b_j' = 1$. In particular,

$$(-1)^{(\deg_h b_i)(\deg_h b_i')} = 1 \quad \text{and} \quad (-1)^{(\deg_h b_j)(\deg_h b_j')} = 1.$$

Applying the term RHS_4 to X gives

$$\begin{aligned}\sum_{i,j} (-1)^{(\deg_h X + \deg_h b_i)(\deg_h b_i')} (-1)^{\deg_h b_j'} (-1)^{(\deg_h X + \deg_h b_i + \deg_h b_j - 1)(\deg_h b_j')} \\ \cdot b_i' b_j' \otimes F_2(X b_i, b_j) \\ = \sum_{i,j} (-1)^{(\deg_h X)(\deg_h b_i' + \deg_h b_j') + (\deg_h b_i)(\deg_h b_j')} b_i' b_j' \otimes F_2(X b_i, b_j).\end{aligned}$$

Thus, $\mathrm{LHS}_5 = \mathrm{RHS}_4$.

Similarly, applying the term LHS_6 to X gives

$$\begin{aligned}\sum_{i,j} (-1)^{(\deg_h b_j)(\deg_h b_i')} (-1)^{\deg_h b_j} (-1)^{(\deg_h X + \deg_h b_i + \deg_h b_j - 1)(\deg_h b_i' + \deg_h b_j')} \\ \cdot (-1)^{\deg_h b_i' + \deg_h b_j'} b_i' b_j' \otimes F_2(X b_i, b_j) \\ = \sum_{i,j} -(-1)^{(\deg_h X)(\deg_h b_i' + \deg_h b_j') + (\deg_h b_i')(\deg_h b_j')} b_i' b_j' \otimes F_2(X b_i, b_j).\end{aligned}$$

Applying the term RHS_8 to X gives

$$-\sum_{i,j}(-1)^{\deg_h b_i'}(-1)^{(\deg_h X+\deg_h b_i-1)(\deg_h b_i')}(-1)^{(\deg_h X+\deg_h b_i+\deg_h b_j-1)(\deg_h b_j')} \cdot b_i'b_j' \otimes F_2(Xb_i,b_j)$$
$$=\sum_{i,j}-(-1)^{(\deg_h X)(\deg_h b_i'+\deg_h b_j')+(\deg_h b_i')(\deg_h b_j')}b_i'b_j'\otimes F_2(Xb_i,b_j).$$

Thus, $\mathrm{LHS}_6=\mathrm{RHS}_8$, so $F\boxtimes \mathrm{id}_{DD}$ is a valid morphism of type D structures from $\widehat{A}\boxtimes\widehat{DD}$ to $\widehat{A}'\boxtimes\widehat{DD}$. □

6.6.6 Proposition *If F and G are $\mathcal{A}_\infty$–morphisms from $\widehat{A}$ to $\widehat{A}'$ as described in Definition 6.6.4, with $F_n,G_n=0$ for $n>2$ and either $F_2=0$ or $G_2=0$, then*

$$(G\circ F)\boxtimes\mathrm{id}_{DD}=(G\boxtimes\mathrm{id}_{DD})\circ(F\boxtimes\mathrm{id}_{DD}).$$

Proof First, suppose $G_2=0$. Then $(G\circ F)_1=G_1\circ F_1$ and $(G\circ F)_2=G_1\circ F_2$. We have

$$\begin{aligned}(G\circ F)\boxtimes\mathrm{id}_{DD}&=1\otimes(G\circ F)_1+\xi\circ((G\circ F)_2\otimes|\mathrm{id}|)\circ(\mathrm{id}\otimes\delta_{DD})\\&=1\otimes(G_1\circ F_1)+\xi\circ(G_1\otimes\mathrm{id})\circ(F_2\otimes|\mathrm{id}|)\circ(\mathrm{id}\otimes\delta_{DD}).\end{aligned}$$

On the other hand,

$$\begin{aligned}&(G\boxtimes\mathrm{id}_{DD})\circ(F\boxtimes\mathrm{id}_{DD})\\&\quad=(\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes(G\boxtimes\mathrm{id}_{DD}))\circ(F\boxtimes\mathrm{id}_{DD})\\&\quad=(\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes 1\otimes G_1)\circ(1\otimes F_1+\xi\circ(F_2\otimes|\mathrm{id}|)\circ(\mathrm{id}\otimes\delta_{DD}))\\&\quad=1\otimes(G_1\circ F_1)+(\mathrm{id}\otimes G_1)\circ\xi\circ(F_2\otimes|\mathrm{id}|)\circ(\mathrm{id}\otimes\delta_{DD}).\end{aligned}$$

This expression equals $(G\circ F)\boxtimes\mathrm{id}_{DD}$ because G_1 is bigrading-preserving.

Now suppose instead that $F_2=0$. Then $(G\circ F)_1$ is still $G_1\circ F_1$, and $(G\circ F)_2=G_2\circ(F_1\otimes\mathrm{id})$. We have

$$\begin{aligned}(G\circ F)\boxtimes\mathrm{id}_{DD}&=1\otimes(G\circ F)_1+\xi\circ((G\circ F)_2\otimes|\mathrm{id}|)\circ(\mathrm{id}\otimes\delta_{DD})\\&=1\otimes(G_1\circ F_1)+\xi\circ(G_2\otimes|\mathrm{id}|)\circ(F_1\otimes\mathrm{id}\otimes\mathrm{id})\circ(\mathrm{id}\otimes\delta_{DD}).\end{aligned}$$

On the other hand,

$$\begin{aligned}&(G\boxtimes\mathrm{id}_{DD})\circ(F\boxtimes\mathrm{id}_{DD})\\&\quad=(\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes(G\boxtimes\mathrm{id}_{DD}))\circ(F\boxtimes\mathrm{id}_{DD})\\&\quad=(\mu_2\otimes\mathrm{id})\circ(\mathrm{id}\otimes(1\otimes G_1+\xi\circ(G_2\otimes|\mathrm{id}|)\circ(\mathrm{id}\otimes\delta_{DD})))\circ(1\otimes F_1)\\&\quad=1\otimes(G_1\circ F_1)+\xi\circ(G_2\otimes|\mathrm{id}|)\circ(\mathrm{id}\otimes\delta_{DD})\circ F_1,\end{aligned}$$

which equals $(G \circ F) \boxtimes \mathrm{id}_{DD}$ because

$$(F_1 \otimes \mathrm{id} \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD}) = (\mathrm{id} \otimes \delta_{DD}) \circ F_1. \qquad \square$$

6.6.7 Proposition *As in Proposition 6.6.6, let F and G be $\mathcal{A}_\infty$–morphisms from $\widehat{A}$ to $\widehat{A}'$ with $F_n, G_n = 0$ for $n > 2$ (here we do not require that either $F_2 = 0$ or $G_2 = 0$). Let H be an $\mathcal{A}_\infty$–homotopy between F and G with $H_n = 0$ for $n > 1$. Define*

$$H \boxtimes \mathrm{id}_{DD} := 1 \otimes H_1;$$

then $H \boxtimes \mathrm{id}_{DD}$ is a homotopy of type D morphisms between $F \boxtimes \mathrm{id}_{DD}$ and $G \boxtimes \mathrm{id}_{DD}$.

Proof Let δ and δ' denote the type D operations on $\widehat{A} \boxtimes \widehat{DD}$ and $\widehat{A}' \boxtimes \widehat{DD}$ respectively. We want to show that

$$F \boxtimes \mathrm{id}_{DD} - G \boxtimes \mathrm{id}_{DD} = (\mathrm{id} \otimes H_1) \circ \delta + \delta' \circ H_1;$$

the other term in the type D homotopy relations of Definition 6.6.2 is zero for this special type of H. Expanding out the left side, we want to show that

$$1 \otimes F_1 - 1 \otimes G_1 + \xi \circ ((F_2 - G_2) \otimes |\mathrm{id}|) \circ (\mathrm{id} \otimes \delta_{DD}) = (\mathrm{id} \otimes H_1) \circ \delta + \delta' \circ H_1.$$

By Example 6.5.14, the $\mathcal{A}_\infty$–homotopy relations for H give us the following two equations:

$$\begin{aligned} F_1 - G_1 &= m_1' \circ H_1 + H_1 \circ m_1, \\ F_2 - G_2 &= -m_2' \circ (H_1 \otimes |\mathrm{id}|) + H_1 \circ m_2. \end{aligned}$$

Thus, the left side of the type D homotopy relation is

$$\begin{aligned} &1 \otimes (m_1' \circ H_1) + 1 \otimes (H_1 \circ m_1) \\ &\qquad + \xi \circ ((-m_2' \circ (H_1 \otimes |\mathrm{id}|) + H_1 \circ m_2) \otimes |\mathrm{id}|) \circ (\mathrm{id} \otimes \delta_{DD}) \\ &= 1 \otimes (m_1' \circ H_1) + 1 \otimes (H_1 \circ m_1) - \xi \circ (m_2' \otimes \mathrm{id}) \circ (H_1 \otimes |\mathrm{id}| \otimes |\mathrm{id}|) \circ (\mathrm{id} \otimes \delta_{DD}) \\ &\qquad + \xi \otimes (H_1 \otimes |\mathrm{id}|) \circ (m_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD}) \\ &= 1 \otimes (m_1' \circ H_1) + 1 \otimes (H_1 \circ m_1) + \xi \circ (m_2' \otimes \mathrm{id}) \circ (H_1 \otimes \mathrm{id} \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD}) \\ &\qquad + (\mathrm{id} \otimes H_1) \circ \xi \circ (m_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD}) \\ &= 1 \otimes (m_1' \circ H_1) + 1 \otimes (H_1 \circ m_1) + \xi \circ (m_2' \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD}) \circ H_1 \\ &\qquad + (\mathrm{id} \otimes H_1) \circ \xi \circ (m_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD}). \end{aligned}$$

On the other hand, using the definition of δ and δ' in Definition 6.3.4, the right side of the type D homotopy relation can be expanded out as

$$\begin{aligned} &1 \otimes (H_1 \circ m_1) + (\mathrm{id} \otimes H_1) \circ \xi \circ (m_2 \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD}) \\ &\qquad + 1 \otimes (m_1' \circ H_1) + \xi \circ (m_2' \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_{DD}) \circ H_1 \end{aligned}$$

which is identical to the previous expression after rearranging terms. Thus, $H \boxtimes \mathrm{id}_{DD}$ is a valid type D homotopy between $F \boxtimes \mathrm{id}_{DD}$ and $G \boxtimes \mathrm{id}_{DD}$. $\square$

6.6.8 Corollary *Let $\mathcal{B}$ and $\mathcal{B}'$ be differential bigraded algebras over an idempotent ring $\mathcal{I}$. Let $\widehat{A}$ and $\widehat{A}'$ be differential bigraded right modules over $\mathcal{B}$ and let $(\widehat{DD}, \delta_{DD})$ be a rank-one type DD bimodule over $\mathcal{B}$ and $\mathcal{B}'$. Assume that $\widehat{A}$ and $\widehat{A}'$ are free as $\mathbb{Z}$–modules, with $\mathbb{Z}$–bases consisting of elements which are grading-homogeneous and have a unique right idempotent.*

Suppose there exist $\mathcal{A}_\infty$–morphisms $F\colon \widehat{A} \to \widehat{A}'$ and $G\colon \widehat{A}' \to \widehat{A}$ with $F_n = 0$ and $G_n = 0$ for $n > 2$, and such that either $F_2 = 0$ or $G_2 = 0$. Furthermore, suppose that $G \circ F$ is $\mathcal{A}_\infty$–homotopic to $\mathrm{id}_{\widehat{A}}$ via an $\mathcal{A}_\infty$–homotopy H with $H_n = 0$ for $n > 1$, and $F \circ G$ is $\mathcal{A}_\infty$–homotopic to $\mathrm{id}_{\widehat{A}'}$ via another $\mathcal{A}_\infty$–homotopy H' with $H'_n = 0$ for $n > 1$.

Then the type D structures $\widehat{A} \boxtimes \widehat{DD}$ and $\widehat{A}' \boxtimes \widehat{DD}$ over $\mathcal{B}'$, defined in Definition 6.3.4, are homotopy equivalent.

Proof This follows from Proposition 6.6.5, Proposition 6.6.6 and Proposition 6.6.7, together with the fact that the box tensor product with id_{DD} on morphisms sends identity morphisms to identity morphisms. $\square$

6.6.9 Corollary *If T and T' are oriented tangle diagrams in $\mathbb{R}_{\geq 0} \otimes \mathbb{R}$ which are related by a Reidemeister move, then $\widehat{D}([T]^{\mathrm{Kh}})$ and $\widehat{D}([T']^{\mathrm{Kh}})$ are homotopy equivalent as type D structures over $m(\mathcal{B})^! \odot \mathcal{B}$. Thus, they are also homotopy equivalent as type D structures over the quotient algebra $\mathcal{B}\Gamma_n$.*

Proof The first claim follows from Corollary 6.6.8 and the proof of Corollary 6.5.18, in which the $\mathcal{A}_\infty$–morphisms $F = \widehat{A}(f)$ and $G = \widehat{A}(g)$ and the $\mathcal{A}_\infty$–homotopy $H = \widehat{A}(\psi)$ used to realize the $\mathcal{A}_\infty$–homotopy equivalences satisfy the conditions of Corollary 6.6.8. The second claim follows from Remark 6.6.3 above. $\square$

Corollary 6.6.9 gives us an alternate proof of Roberts [12, Theorem 46].

References

[1] **R M Adin**, **Y Roichman**, *On maximal chains in the non-crossing partition lattice*, J. Combin. Theory Ser. A 125 (2014) 18–46 MR

[2] **D Bessis**, *The dual braid monoid*, Ann. Sci. École Norm. Sup. 36 (2003) 647–683 MR

[3] **T Braden**, *Perverse sheaves on Grassmannians*, Canad. J. Math. 54 (2002) 493–532 MR

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

[5] **R Lipshitz**, **P Ozsvath**, **D Thurston**, *Bordered Heegaard Floer homology: invariance and pairing*, preprint (2008) arXiv

[6] **R Lipshitz**, **P S Ozsváth**, **D P Thurston**, *Heegaard Floer homology as morphism spaces*, Quantum Topol. 2 (2011) 381–449 MR

[7] **R Lipshitz**, **P S Ozsváth**, **D P Thurston**, *Bimodules in bordered Heegaard Floer homology*, Geom. Topol. 19 (2015) 525–724 MR

[8] **A J Manion**, *Constructions and computations in Khovanov homology*, PhD thesis, Princeton Universty (2015) MR Available at `http://search.proquest.com/docview/1707650065`

[9] **D Pálvölgyi**, *For any two noncrossing partitions p,q of n, is the graph of geodesics from p to q in $NC(n)$ connected?* (version: 2015-01-06) Available at `http://mathoverflow.net/q/192273`

[10] **A Polishchuk**, **L Positselski**, *Quadratic algebras*, University Lecture Series 37, Amer. Math. Soc., Providence, RI (2005) MR

[11] **L Roberts**, *A type A structure in Khovanov homology*, Algebr. Geom. Topol. 16 (2016) 3653–3719 MR

[12] **L P Roberts**, *A type D structure in Khovanov homology*, Adv. Math. 293 (2016) 81–145 MR

[13] **C Stroppel**, *Parabolic category $\mathcal{O}$, perverse sheaves on Grassmannians, Springer fibres and Khovanov homology*, Compos. Math. 145 (2009) 954–992 MR

Department of Mathematics, UCLA
520 Portola Plaza, Los Angeles, CA 90095, United States

`manion@math.ucla.edu`

`http://math.ucla.edu/~manion/`

Received: 12 November 2015 Revised: 13 July 2016

Algebraic & Geometric Topology 17 (2017) 1675–1700

The intersection graph of an orientable generic surface

DORON BEN HADAR

The intersection graph $M(i)$ of a generic surface $i\colon F \to S^3$ is the set of values which are either singularities or intersections. It is a multigraph whose edges are transverse intersections of two surfaces and whose vertices are triple intersections and branch values. $M(i)$ has an enhanced graph structure which Gui-Song Li referred to as a "daisy graph". If F is oriented, then the orientation further refines the structure of $M(i)$ into what Li called an "arrowed daisy graph".

Li left open the question "which arrowed daisy graphs can be realized as the intersection graph of an oriented generic surface?" The main theorem of this article will answer this. I will also provide some generalizations and extensions to this theorem in Sections 4 and 5.

57N10, 57N12; 57N35, 57N40, 57N75

1 Introduction: the structure of the intersection graph

A (proper) generic surface in a 3–manifold is a generalization of an immersed surface in a general position. Specifically:

Definition 1.1 A (proper) generic surface in a 3–manifold M is a smooth mapping $i\colon F \to M$, where F is a compact surface (called the underlying surface), and each value of i has a neighborhood U in M such that $U \cap i(F)$ looks like one of the pictures in Figure 1 (The purple part is ∂M).

Figure 1 (top left), (top middle) and (top right) are, respectively, a regular value, a double value and a triple value. Locally they look like the intersection of one, two or three of the coordinate planes in a neighborhood of the origin in $\mathbb{R}^3$. Figure 1 (bottom left) is a branch value — locally, it looks like the renowned "Whitney's Umbrella" — (bottom middle) is a regular boundary (RB) value and (bottom right) is a double boundary (DB) value — they resemble the intersection of one or two of the $[xz]$ and $[yz]$ planes, with the boundary (the $[xy]$ plane) in a neighborhood of the origin inside the upper half space.

Published: 17 July 2017 DOI: 10.2140/agt.2017.17.1675

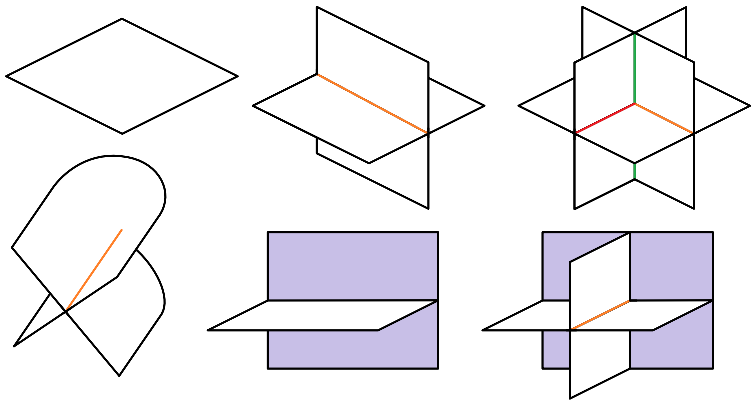

Figure 1: The values of a generic surface

These surfaces are "proper" in the sense that $i(\partial F) = i(F) \cap \partial M$, and they are "generic" in the sense that every proper smooth function from a compact surface to M can be turned into a proper generic surface via an arbitrarily small perturbation. They are also "stable" in the sense that a small enough perturbation can only change i up to an isotopy of F and M. If F is closed, I call i a "closed generic surface".

I am interested in the intersection graph of a generic surface. It is the set $M(i) = \overline{\{p \in i(F) \mid 1 < |i^{-1}(p)|\}}$ of all values which are neither regular nor RB. I will regard the intersection graph in two ways, and each way has its own notations. I will also use specific conventions when I draw it.

Definition 1.2 (1) First, I will regard the intersection graph as a multigraph whose vertices are the triple values (degree 6), branch values and DB values (degree 1) of i, and whose edges are the segments of a "double line" (a line consisting of double values, such as the orange line in the top middle of Figure 1) between two vertices. In addition to this "graph part", $M(i)$ may contain several "double circles" — double lines that close into circles instead of ending at vertices. Having no vertices or edges, double circles do not comply with the traditional definition of a graph, and need to be accounted for separately.

Note that $M(i)$ may have graph-theoretical "loops" — paths whose ends are both at the same vertex (which must be a triple value, as its degree is not 1). Due to this, it will be important later on to distinguish between the two "ends of an edge". I use the term "half-edge" to describe such an end.

In Figure 1 (top right), I show that three "segments" of double line (marked in orange, red and green) intersect at each triple value. The intersection cuts each segment into a

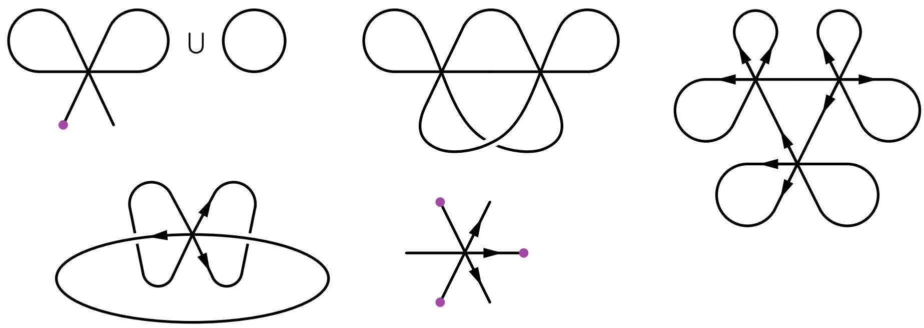

Figure 2: Some examples of daisy graphs and arrowed daisy graphs

pair of half-edges that seem to be in continuation to one another. I say that such a pair of half edges is "consecutive". Each triple value has three disjoint pairs of consecutive half edges. Notice that these two half-edges may originate in the same edge, in which case said edge is a loop.

(2) Secondly, I will regard a pair of consecutive half-edges to be parts of a long path that crosses the triple value. $M(i)$ is then the union of several of these long lines, which I call "double arcs". Three double arcs intersect at each triple value, but this number includes multiplicity; it may be that a double arc crosses itself at a triple value. Double arcs are thus immersed 1–manifolds in M, not embedded ones.

A double arc may, as in Figure 1 (bottom left) and (bottom right), end in a branch value or a DB value on each side of it. Otherwise, it may close up into a circle. I refer to arcs of the latter kind as "closed" and to the former type as "open".

According to the "double arc" notation, a double circle is just a closed double arc that does not pass any triple values. A closed double arc that passes a triple value only once is simply a loop whose two half-edges are consecutive.

(3) When drawing a diagram of the intersection graph, make sure that each triple value looks like the intersection of three lines, as in Figure 2 (top left) and (top middle). In this way, one can see which half-edges are consecutive and what the double arcs are. A "purple point" at an end of an edge symbolizes that this edge ends in a DB value, as opposed to a branch value. Double circles are, of course, drawn as circles disjoint from the graph part.

Many authors have studied the intersection graph from various angels. For instance, in [6], Izumiya and Marar showed a connection between the Euler characteristic of the (image of the) closed generic surface, and the Euler characteristic of its underlying surface. Namely, $\chi(i(F)) = \chi(F) + T(i) + \frac{1}{2}B(i)$ where $T(i)$ is the number of triple values and $B(i)$ is the number of branch values of the surface (they called the latter "branch points", hence B). This generalized an earlier result by Carter and Ko [7], which in turn generalized a result by Banchoff [1].

I am interested in the intersection graph because it encodes several important properties of the surface. For instance, Giller [5] showed that examining the intersection set can tell us if a generic surface in $\mathbb{R}^3$ can be lifted into an embedded surface in $\mathbb{R}^4$. Several other authors, including Carter and Saito [3], and Satoh [9], have looked into the connection between liftings and the intersection graph.

One can, to some extent, classify generic surfaces according to their intersection graph. An early example of this can be found in [4], where Cromwell and Marar classified the kind of surface in $\mathbb{R}^3$ that can have an intersection graph of a certain form (connected and with only one vertex which is a triple value).

In this article, I address the case in which both the 3–manifold M and the underlying surface F are oriented. In this case, one can add more data to a diagram of the intersection graph. To see this, one must first recall the notion of a co-orientation:

Definition 1.3 A co-orientation on a generic surface $i\colon F \to M$ is a continuous choice, for every non-branch-value x, of a normal vector in $T_{i(x)}M$ that is orthogonal to $D_i(T_x F)$. If M is oriented, then there is a one-to-one correspondence between orientations on F and co-orientations on i. It matches each orientation on F with the normal $\vec{n}$ for which $(D_i(\vec{v}_1), D_i(\vec{v}_2), \vec{n})$ upholds the orientation of M whenever $(\vec{v}_1, \vec{v}_2)$ upholds that of F.

I use co-orientations to indicate the orientation of a generic surface in illustrations. In particular, Figure 3 depicts the neighborhood of a triple value. Notice that each of the three "double arc segments" that pass through a triple value (a) consists of the intersection of two of the three planes that intersect in the value, and (b) intersects the remaining plane transversally. The normal arrows on this last plane point toward one "preferred direction" on this arc or, equivalently, toward one of the two consecutive half-edges. I refer to the half-edge the arrow points toward as the "preferred" one of the two.

I can encode this information in a diagram of $M(i)$. I use a small arrow to mark the preferred direction on each intersecting arc segment at each triple value as in Figure 2 (top right), (bottom left) and (bottom middle). I can formally define a type of combinatorial structure that encodes the relevant information about $M(i)$. This is a generalization of definitions made by Li in [8, pages 3721, 3723].

Definition 1.4 (1) A daisy graph (DG) (V, E, n, B, C) is a 5–tuple where (V, E) is multigraph whose vertices are all of degree 1 or 6, n is a nonnegative integer, B

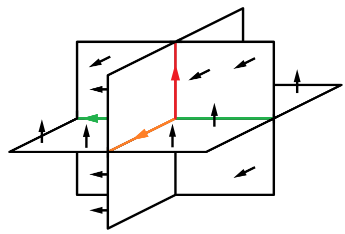

Figure 3: The preferred directions at a triple value

is a subset of the set of degree-1 vertices, and for each degree-6 vertex v, we have a division $C(v)$ of the set of the half-edges of v into three pairs.

For the DG of a generic surface, n will indicate the number of double circles the surface has, B is the set of DB values (the other degree-1 vertices are branch values), and for each triple value v, the three pairs in $C(v)$ are the three pairs of consecutive half-edges. In the interest of convenience, I call the vertices of any DG triple values, branch values, and DB values, in accordance with their degrees and belonging to B. I call a pair of half-edges in $C(v)$ consecutive. I draw a DG according to the conventions of Definition 1.2(3).

(2) An arrowed daisy graph (ADG) (V,E,n,B,C,A) is a 6–tuple where (V,E,n,B,C) is a DG, and for each triple value v and each pair $p \in c(v)$, we let $A(v,p)$ be one of the half-edges in p, which I call the "preferred half-edge". In diagrams I mark each preferred half-edge with an arrow as in Figure 2 (top right), (bottom left) and (bottom middle). In the ADG of an oriented generic surface I choose the preferred half-edges as per the co-orientation, as explained above.

Remark 1.5 (1) Li assumed the surface is an immersion of a closed surface, so he did not have branch values or DB values.

(2) Despite the similarity to graphs on surfaces, daisy graphs are not planar. One arc can go "above" another. I mark it as a crossing in a knot (see Figure 2 (top middle), (bottom left) and (bottom middle)) to avoid confusion, but it is not a real crossing; it does not matter which arc is higher and which is lower.

In [8], Li found out which DGs can be realized as the intersection graph of an orientable generic surface in $\mathbb{R}^3$. He defined the ADG in order to answer this, but this led to a new question: which ADGs can be realized as the intersection graph of an oriented generic surface in a given oriented 3–manifold M? (Li posted this as an open question; see [8, page 3725]). The main purpose of this article is to answer this question. It turns out that there are two inherently different cases: the case where $H(M;\mathbb{Z})$ is periodic (all its elements have a finite order), and the case where it is not. I solve the first case in Sections 2 and 3 and the second case in Section 4. In Section 5, I will discuss a refinement of the notion of ADGs.

Remark 1.6 The solution I give in this paper is not fully constructive. Specifically, when I prove that an ADG is realizable, I only construct a part of the realizing surface. I then use arguments of homology and surgery to prove that this part can be extended to a whole surface. I have, by now, found several ways to construct an entire surface, but they are unneeded here, and will only lengthen the proof.

I explained one of these constructions in my Ph D Thesis [2], where I show that the problem of determining if a generic surface is liftable into an embedded surface in 4–space (a knotted surface) is NP complete. The purpose of [2] only requires me to construct a realizing surface for a very specific type of realizable ADGs, but the construction given therein can be easily generalized to fit all constructible ADGs.

Acknowledgement This article contains some results from my doctoral thesis, conducted under the advisement of Dr. Tahl Nowik, at the Department of Mathematics at Bar Ilan University. Approved 15 November 2016.

2 Gradings and winding numbers

Definition 2.1 (1) Let G be an ADG. I say that an edge e is "preferred" (resp. "nonpreferred") at a vertex v if one of the ends of e is a preferred (resp. nonpreferred) half-edge at the vertex v.

(2) A grading of an ADG G is a choice of a number $g(e) \in \mathbb{Z}$ (called the grade of e) for every edge e of G, such that, at each triple value v, all the nonpreferred edges at v have the same grade $a(v)$ and all preferred edges have the grade $a(v)+1$. An ADG that has a grading is called "gradable".

The grading concerns only the "graph part" of the ADG and ignores the double circles. Since double circles pose no obstruction to gradability, I consider an ADG that consists solely of double circles to be gradable.

Figure 4 (far left) depicts a graded ADG. The ADG in Figure 4 (far right) is not gradable. The reason for this is that the red and green edges are both nonpreferred at the upper triple value, implying that they ought to have the same grade, but at the bottom triple value one of them is preferred and the other is not, implying that they ought to have different grades. I will discuss the connection between the gradability and the realizability of an ADG shortly. However, I will begin by explaining the factors that make an ADG gradable.

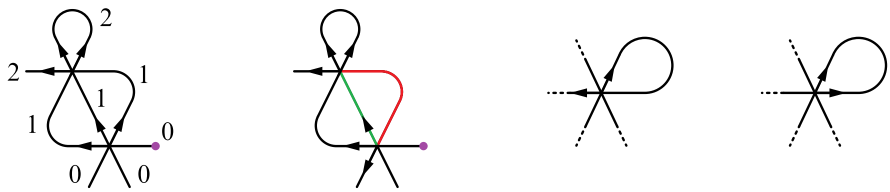

Figure 4: The obstructions to gradability of an ADG

Definition 2.2 A "grade obstructing" loop of an ADG is a loop (a path whose ends both lay on the same vertex v) with one preferred end and one nonpreferred end at v. For example, the loop in Figure 4 (middle right) is grade obstructing, while the loop in Figure 4 (far right) is not.

Remark 2.3 (1) A gradable ADG cannot have grade obstructing loops since the grade of such a loop would have to be $a(v) = g(e) = a(v) + 1$.

(2) If an ADG has no grade obstructing loops, then the sets of preferred edges at v and nonpreferred edges at v are mutually exclusive. This simplifies the following definition.

Definition 2.4 (1) Given a DADG G with no grade obstructing loops, and two edges e and f that share a vertex v (which must be a triple value since its degree cannot be 1), define the "grading difference" $\Delta g(e, v, f)$ to be 1 if f is preferred at v and e is not, -1 if it is the other way around, and 0 if either both f and g are preferred or if they are both nonpreferred.

(2) The grading difference of a path $e_0, v_0, e_1, v_1, \ldots, v_{r-1}, e_r$ in G is the sum

$$\sum_{k=1}^{r} \Delta g(e_{k-1}, v_{k-1}, e_k).$$

Lemma 2.5 (1) *If an ADG has a grading g, then the grading difference of a path $e_0, v_0, e_1, v_1, \ldots, v_{r-1}, e_r$ is equal to $g(e_r) - g(e_0)$.*

(2) *An ADG is gradable if and only if it has no grade obstructing loop and, for every pair of edges e and f, every path between e and f has the same grading difference.*

(3) *One can check if an ADG G is gradable, and therefore construct a grading, in linear $O(|E|)$ time, where E is the set of the edges of G.*

Proof (1) For a short path e, v, f, this follows directly from Definitions 2.1 and 2.4(1). Induction implies the general case.

(2) ($\Leftarrow$) The first part is Remark 2.3, and the second follows from (1).

($\Rightarrow$) For every connected component G' of G (that is not a double circle), do the following: Choose one edge e in G' and give it the grade 0. Next, for every other edge f in G' choose a path $e = e_0, v_0, e_1, \dots, e_r = f$ and set the grade $g(f)$ of f to be the relative grade of this path. By assumption, this is independent of the path. If f shares a vertex v with another edge h, then $e = e_0, v_0, e_1, \dots, e_r = f, v, h$ is a path from e to h, and so

$$g(h) = \sum_{k=1}^{r} (\Delta g(e_{k-1}, v_{k-1}, e_k)) + \Delta g(f, v, h) = g(f) + \Delta g(f, v, h).$$

This holds for every adjacent pair of edges. In particular, if v is a vertex and f is nonpreferred at v, then for the number $a(v) = g(f)$, every nonpreferred edge h at v upholds $g(h) = g(f) + \Delta g(f, v, h) = a(v)$, and every preferred edge h at v upholds $g(h) = g(f) + \Delta g(f, v, h) = a(v) + 1$, and so g is a grading.

(3) It takes $O(|E|)$ time to go over the edges of G and check if any of them is a grade-obstructing loop. If no such loop exists, I will assign each edge f of G a number $g(f)$ which, if the graph is gradable, will be a grading. I say that the algorithm "reached" (resp. "exhausted") a vertex if it assigned a grading to at least one (resp. all) of the edges of this vertex. I begin by choosing one edge e and grading it $g(e) = 0$. For each vertex of e, I set $a(v) = g(e) - 1 = -1$ or $a(v) = g(e) = 0$ if e is respectively preferred or nonpreferred at v.

Next, I choose a vertex v that the algorithm has reached but has not exhausted (currently, this means that v is one of the vertices of e) and go over the edges of v. If a preferred or nonpreferred edge f has yet to be graded, then grade it $g(f) = a(v) + 1$ or $g(f) = a(v)$, respectively, then look at the other vertex w of f. If this is the first time the algorithm reaches w, set $a(w) = g(f) - 1$ or $a(w) = g(f)$ if f is respectively preferred or nonpreferred at w. If the algorithm reached w before, then $a(w)$ has already been set previously. In order for g to be a grading, w must uphold $a(w) = g(f) - 1$ or $a(w) = g(f)$, depending on if f is preferred at v or not. Check if this equality holds.

If the equality holds, move on to the other edges of v and do the same. Since v has no more than six edges, this takes $O(1)$ time. When you have exhausted v, move on to another vertex G that the algorithm has reached but has yet to exhaust. Continue like this until either (a) you grade an edge f whose "other vertex" w has already been reached and for which the appropriate equality, $a(w) = g(f) - 1$ or $a(w) = g(f)$, fails, or (b) if you have not reached such an edge but there are no more vertices that the algorithm reached but has yet to exhaust.

If you stop because of (a), then G is not gradable. In order to see this, notice that if you reached a vertex v via an edge e_v, and then you grade another edge f at v, then

$g(f) = \Delta(e_v, v, f) + g(e_v)$. This can be proven on a case per case basis. For instance, if both f and e_v are preferred at v, then $\Delta(e_v, v, f) = 0$, and according to the above,

$$a(v) = g(e_v) - 1 \quad \text{and} \quad g(f) = a(v) + 1 = \Delta(e_v, v, f) + g(e_v),$$

as required. Similar considerations show that if the other vertex w of f has already been reached, and the appropriate equality, $a(w) = g(f) - 1$ or $a(w) = g(f)$, fails, then $g(f) \neq \Delta(e_w, w, f) + g(e_w)$.

Induction implies that every edge f that the algorithm has already graded has a path $e = e_0, v_0, e_1, \dots, e_r = f$ such that $g(f)$ is equal to the grading difference of this path. Indeed, it holds for e itself, and if you assume that it holds for every edge you graded before, in particular for e_v, then $g(e_v)$ is equal to the grading difference of the path $e = e_0, v_0, e_1, \dots, e_{r-1} = e_v$, and

$$g(f) = g(e_v) + (g(f) - g(e_v)) = \sum_{k=1}^{r-1} (\Delta g(e_{k-1}, v_{k-1}, e_k)) + \Delta(e_v, v, f),$$

which is the grading difference of the path $e = e_0, v_0, e_1, \dots, e_{r-1} = e_v, v, f$.

Now, assuming as before that the algorithm already reached w and that the appropriate equality, $a(w) = g(f) - 1$ or $a(w) = g(f)$, fails, one can realize $g(e_w)$ as the grading difference of a path $e = h_0, w_0, h_1, \dots, w_{r-1}, h_r = e_w$. There is thus a second path between e and f, namely $e = h_0, w_0, h_1, \dots, w_{r-1}, h_r = e_w, w, f$, whose grading difference is

$$\sum_{k=1}^{r} (\Delta g(h_{k-1}, w_{k-1}, h_k)) + \Delta(e_w, w, f) = g(e_w) + \Delta(e_w, w, f) \neq g(f).$$

As these two paths have different grading differences, (2) implies that G is not gradable.

If the algorithm stopped because of (b), then it provided a grading $g(f)$ for every edge f in the connected component of G that contains e. Since the equality never failed, every vertex v and every preferred or nonpreferred edge f at v upholds $a(v) = g(f) - 1$ or $a(v) = g(f)$, respectively. This means that g is indeed a grading of this connected component. If there are any vertices left that the algorithm hasn't reached yet, then they belong to a different connected component. Choose a new ungraded edge e and grade it $g(e) = 0$, and then proceed to grade its connected component. Eventually, either you will reach stop condition (a), meaning that G is not gradable, or you will exhaust all the vertices of G, in which case you finished grading all of G.

In total, the algorithm went over every edge f of G, determined $g(f)$, and either determined $a(w)$ for one or both of its vertices, or checked if it upheld the equality $a(w) = g(f) - 1$ or $a(w) = g(f)$. This takes $O(|E|)$ time. □

Remark 2.6 If the graph part of an ADG G is a forest, then the algorithm will never reach the stop condition (a), and so G is gradable.

I can now formulate the main theorem:

Theorem 2.7 *Let M be an oriented 3–manifold for which $H(M;\mathbb{Z})$ is periodic.*

(1) *If M has no boundary, then an ADG G can be realized as the intersection graph of an oriented generic surface $i\colon F \to M$ if and only if G is gradable and has no DB values.*

(2) *If M has a boundary, then an ADG G can be realized as the intersection graph of an oriented generic surface $i\colon F \to M$ if and only if G is gradable.*

Result 2.8 *In* [8], *Li showed that a DG with no DB values or branch values is realizable if and only if any arc in it is composed of an even number of edges. Theorem 2.7 implies a generalization of this: a general DG is realizable via an orientable generic surface if and only if every* closed *arc is composed of an even number of edges.*

Proof of Result 2.8 If a DG is realizable via an orientable generic surface, then any orientation of the surface gives the DG an ADG structure (arrows), and this ADG is realizable and therefore gradable. The grading of each subsequent edge on an arc will have a different parity than the grading of the previous edge and, in particular, closed arcs must have an even number of edges on them.

On the other hand, given a DG that upholds this condition (every closed arc must have an even number of edges), it is possible to give the DG a "short grading": number the edges with only 0 and 1 in such a way that consecutive edges have different numbers. Clearly, the only obstruction to this is the existence of closed arcs with an odd number of edges. Now, one half-edge in every consecutive pair will belong to an edge whose grade is 1, and the other will belong to an edge whose grade is 0. You can give the DG an ADG structure that matches this grading by choosing the former half-edges to be preferred. This graded ADG is realizable, and in particular, the underlying DG is realizable via an orientable surface. □

In the remainder of this section, I will prove the "only if" direction of the articles of Theorem 2.7. The "if" direction will be proven in the next section. One part of the "only if direction" is trivial: a generic surface in a boundaryless 3–manifold cannot have DB values. In order to prove the other part, that the intersection graph of a generic surface is a gradable ADG, I use 3–dimensional winding numbers:

Definition 2.9 Let $i\colon F \to M$ be a proper generic surface in a 3–manifold M.

(1) A face (resp. body) of i is a connected component of $i(F) \setminus M(i)$ (resp. $M \setminus i(F)$).

(2) Each face V is an embedded surface in M, and there is a body on each side of it. I say these two bodies are adjacent (via V). A priori, it is possible that these two bodies are in fact two parts of the same body, and even that V is a one-sided surface. In these cases, this body will be self adjacent, but this does not happen in any of the cases I am interested in.

(3) If i has a co-orientation, then each face V is two sided, and the arrows on the face point towards one of its two sides. I say that the body on the side that the arrows point toward is "greater" (via V) than the body on the other side of V.

(4) A choice of "winding numbers" for i is a choice of an integer $w(U) \in \mathbb{Z}$ for every body U of i such that if U_1 and U_2 are adjacent, and U_1 the greater of the two, then $g(U_1) = g(U_2) + 1$.

Lemma 2.10 *If M is a connected and orientable 3–manifold, $H_1(M;\mathbb{Z})$ is periodic, and $i\colon F \to M$ is a co-oriented generic surface, then i has a choice of winding numbers.*

Proof Pick one body U_0 to be "the exterior" of the surface, and set $w(U_0) = 0$. Next, define the winding numbers for every other body U like so:

Take a smooth path from U_0 to U that is in general position to i (it intersects $i(F)$ only at faces, and does so transversally), and set $w(U)$ to be the signed number of times it crosses $i(F)$, that is, the number of times it intersects $i(F)$ in the direction of the co-orientation minus the number of times it crosses against the co-orientation. This is well defined, since any two such paths α and β must give the same number. Otherwise, the composition $\beta^{-1} * \alpha$ is a 1–cycle whose intersection number with the 2–cycle represented by i is nonzero. This implies that this 1–cycle is of infinite order in $H_1(M;\mathbb{Z})$, contradicting the fact that this group is periodic.

It is also clear that if U_1 and U_2 are adjacent, and U_1 is the greater of the pair, then $g(U_1) = g(U_2) + 1$. □

Remarks 2.11 (1) It is clear that two different choices of "winding numbers" for i will differ by a constant, and that the one I created is unique in satisfying $w(U_0) = 0$.

(2) I can do a similar process on a loop γ in $\mathbb{R}^2$ instead of a surface in a 3–manifold. If I choose the component U_0 of $\mathbb{R}^2 \setminus \mathrm{Im}(\gamma)$ to be the actual exterior, then this will produce the usual winding numbers: $w(U)$ will be the number of times γ winds around a point in U.

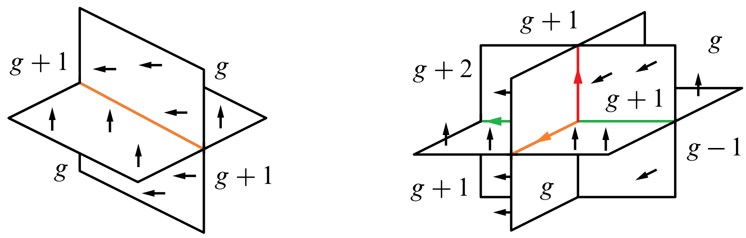

Figure 5: The winding number of bodies around an edge of $M(i)$ and a triple value

I will use the winding numbers to induce a grading in the following manner: the neighborhood of a double value includes four bodies, with the possibility that some of them are, in fact, different parts of the same body. If the surface has a co-orientation and winding numbers, then there is a number g such that two of these bodies have the winding number g, one has $g+1$ and one has $g-1$. Figure 5 (left) depicts this.

Due to continuity, this will be the same value g for all the double values on the same edge (or double circle). I call this number the grading of the edge, and denote the grading of an edge e by $g(e)$. This is indeed a grading in the sense of Definition 2.1. In order to prove this, I need to show that at every triple value of the surface, all the preferred half-edges have the same grading, which is greater by 1 than the grading of all the nonpreferred ones. This can be seen in Figure 5 (right), which depicts the winding numbers of the bodies around an arbitrary triple value. Indeed, you can see that the preferred half-edges, the ones going up, left and outwards (toward the reader), have the grading $g+1$, while the other edges have the grading g. This proves the "only if" direction of Theorem 2.7.

3 Realizing graded arrowed daisy graphs

In order to prove the "if" direction of Theorem 2.7, I will first prove a partial result. I will limit the discussion to connected ADGs with no DB values.

Lemma 3.1 *Every* connected, gradable *ADG G* without DB values *has a closed generic surface $i\colon F \to S^3$ such that the intersection graph of i is equal, as an ADG, to G.*

Remark 3.2 It may be assumed that $i(F)$ is connected. Otherwise, one of its components will contain the connected intersection graph, and the rest will be embedded connected surfaces in S^3. They can be removed by deleting their preimages from F.

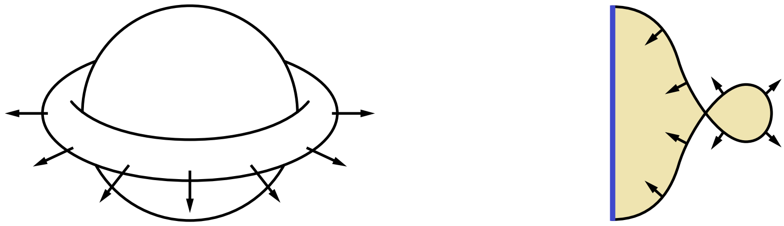

Figure 6: A surface whose intersection graph is a double circle

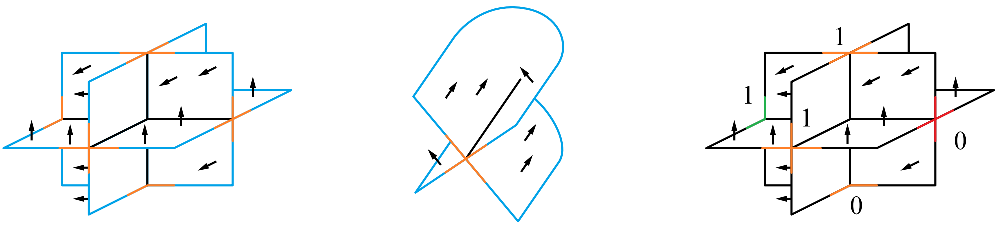

Figure 7: The vertices' neighborhoods and their gluing zones

I begin with the unique case where the ADG is a double circle. The generic surface from Figure 6 (left) has a single double circle as its intersection graph. It is the surface of revolution of the curve from Figure 6 (right) around the blue axis. Both figures have indication for the co-orientation. The intersection graph will be the revolution of the orange dot where the curve intersects itself, and will thus be a circle. The underlying surface is clearly a sphere.

Any other connected ADG is a "graph ADG"—it will have no double circles. In this case, I begin by constructing a part of the matching generic surface, the regular neighborhood of the intersection graph. Li defined something similar in [8, Figure 2, page 3723], which he called a "cross-surface", and I will use the same notation.

Definition 3.3 Given an ADG G that has no DB values and no double circle, a "cross-surface" X_G of G is a shape in S^3 that is built via the following two steps:

(1) For every triple value v of G, embed a copy of Figure 7 (left) in S^3. This shape is called the "vertex neighborhood" of v. Similarly, for every cross cap v of G, embed a vertex neighborhood that looks like Figure 7 (middle) in S^3. Make sure that the different vertex neighborhoods will be pairwise disjoint. These will be the neighborhoods of the actual triple values and branch values of the surface we are constructing. The vertex neighborhoods have little arrows on them which indicate the co-orientation on this part of the surface. It is important to remember which vertex neighborhood corresponds to which vertex of G.

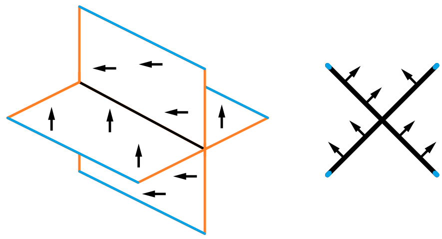

Figure 8: The X-bundle of an edge and a cross-section of it

Each vertex v is supposed to have either one or six half-edges enter it. One can see the ends of these half-edges in Figure 7. Note that each half-edge will cross the boundary of the vertex neighborhood at one point. Given such an intersection point, I refer to its regular neighborhood *inside the boundary of the vertex neighborhood* as its "gluing zone". In Figure 7 (left) and (middle), I colored the gluing zones in orange and the rest of the boundary of the vertex neighborhoods in blue.

The above implies that each such "gluing zone" on the vertex neighborhood of v should correspond to a unique half-edge of G that ends in v. The reader has some freedom in choosing which gluing zone corresponds to which half-edge, but in accordance with Definition 1.4, the following must happen for every triple value, in order for this association to reflect the ADG structure of G:

(1a) Two gluing zones on opposite sides of the value's neighborhood, like those marked red and green in Figure 7 (right), must correspond to a pair of consecutive half-edges.

(1b) In compliance with the co-orientation (the little arrows) on the vertex neighborhoods, the zones that the arrows point toward — those marked with the number 1 in Figure 7 (right) — must correspond to the preferred half-edges. The other zones, marked with 0, will correspond to the nonpreferred half-edges.

Again, it is important to remember which gluing zone corresponds to which half-edge.

(2) Step 2 will realize the edges of G. The "ends" of each edge have already been realized inside the corresponding vertex neighborhoods. Each end is realized by the double line between the vertex and the corresponding gluing zone. I want to add the "middle of the edge" to our construction. This should be a double line, the intersection of two surfaces, as in Figure 8 (left). For each edge e of G, embed a matching copy of this shape in S^3. A closer look reveals that this shape is a bundle over a closed interval, whose fiber looks like the "X" in Figure 8 (right). I therefore call this shape "the X-bundle of e". The embedding of the X-bundles must follow the following rule:

Figure 9: The gluing must preserve the orientation.

(2a) The boundary of each X-bundle is composed of two parts: the fibers at the ends of the interval, colored orange, and the (union of the) ends of all the fibers, colored blue. It is natural to identify the two orange fibers with the two ends (half-edges) of the matching edge e. Until now, I identified each half-edge of the ADG with both (i) a "gluing zone" on the boundary of the neighborhood of some vertex, and (ii) a fiber at the end of some X-bundle. Make sure to embed the X-bundles so that each end fiber coincides with the matching gluing zone. Additionally, make sure that the rest of the X-bundle (the X-bundle sans the end fiber) is disjoint from the vertex neighborhoods, and that X-bundles of different edges do not touch one another.

The resulting shape is the cross-surface. It is similar to a generic surface, but it has a boundary: the union of all the "blue parts" of the boundaries of the vertex neighborhoods and the X-bundles. The intersection graph of this "generic surface with a boundary" is clearly isomorphic (as a multigraph) to G: I already identified each vertex (resp. edge) of it with a unique vertex (resp. edge) of G and made sure that each edge ends in the vertices in which it should properly end. Rule (1a) implies that this identification will preserve the consecutive pairs of half-edges. This means that the intersection graph is isomorphic to G as a DG, not just as a multigraph. In order for this to be an ADG isomorphism as well, the embedding of the X-bundles must follow another rule:

(2b) To have an ADG structure, the cross-surface must have a co-orientation. Note that both the vertex neighborhoods and the X-bundles have arrows on them, which represent co-orientations. When you embed the X-bundles, these co-orientations on them must match, as in Figure 9 (left), and unlike Figure 9 (right). This way they will merge into a continuous co-orientation on the entire cross-surface.

Rule (1b) implies that the preferred half-edges of the intersection graph will correspond to the preferred half-edges of G. This means that the intersection graph will be isomorphic to G as an ADG as well.

The boundary of the cross-surface is the union of many embedded intervals in S^3 — the "blue parts" of the boundaries of the vertex neighborhoods and the X-bundles. Since each end of every interval coincides with an end of one other interval, and the intervals do not otherwise intersect, their union is an embedded compact 1–manifold in S^3. The

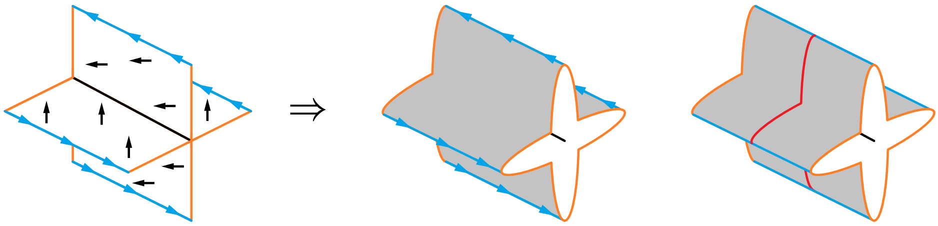

Figure 10: Thickening the cross-surface into a handle body, and the handle body's meridians

cross-surface induces an orientation on this 1–manifold, the usual orientation that an oriented manifold induces on its boundary. It is depicted in the far left of Figure 10.

I will show that the boundary of the cross-surface of a connected ADG G is also the oriented boundary of an embedded surface which is disjoint from the cross-surface. It follows that the union of the cross-surface and the embedded surface, with the orientation on the embedded surface reversed, will be a closed and oriented generic surface whose intersection graph will be isomorphic to G. This will prove Lemma 3.1.

In order to prove that such an embedded surface exists, I begin by "thickening" the cross-surface as in Figure 10 (middle); this only shows how to do this to an X-bundle, but you can similarly do this for all the vertex neighborhoods. This results in a handle body H in S^3, and our 1–cycle is on its boundary. It will suffice to prove that the 1–cycle is the boundary of some embedded surface in the complement of H. This happens if and only if is the cycle is a "boundary" in the homological sense — it is equal to 0 in $H_1(\overline{S^3\setminus H};\mathbb{Z})$.

For any loop γ in the intersection graph, I define a functional $f_\gamma\colon H_1(\overline{S^3\setminus H};\mathbb{Z})\to\mathbb{Z}$ such that $f_\gamma(c)$ is the linking number of γ and a representative of c. It is well-defined, since cycles in $\overline{S^3\setminus H}$ are disjoint from γ, and since the linking number of γ with any boundary in $H_1(\overline{S^3\setminus H})$ is 0, as the boundary bounds a surface in $\overline{S^3\setminus H}$ which is disjoint from γ.

In case the first Betti number of G (and therefore of the intersection graph) is n, then the genus of H is n, and the intersection graph has n simple cycles $C_1,\dots,C_n$, such that each cycle C_i contains an edge e_i that is not contained in any of the other cycles. For every cycle C_i, I take a small meridian m_i around the edge C_i (as depicted in red in Figure 10 (right)). It follows that $f_{c_i}([m_j])=\delta_{ij}$, where δ is the Kronecker delta function. Additionally, since $\overline{S^3\setminus H}$ is the complement of an n–handle body, $H_1(\overline{S^3\setminus H})\equiv\mathbb{Z}^n$. I will prove that:

Lemma 3.4 *These meridians form a base of* $H_1(\overline{S^3\setminus H})$.

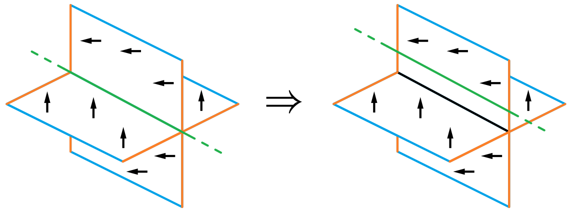

Figure 11: Moving the intersection graph away from the cross-surface

Proof First, I show that the meridians are independent. This is because a boundary in $\overline{S^3 \setminus H}$ would have 0 as the linking number with every c_i, but the linking number of a nontrivial combination $x = \sum a_i[m_i]$ with any c_j will be a_j for some j and $a_j \neq 0$. Second, notice that this implies that $N = \mathrm{Span}_{\mathbb{Z}}\{[m_1], \dots, [m_n]\}$ is a maximal lattice in $H_1(\overline{S^3 \setminus H}) \equiv \mathbb{Z}^n$, and therefore has a finite index.

Third, had N been a strict subgroup of $H_1(\overline{S^3 \setminus H}; \mathbb{Z})$, then there would be an element $y \in H_1(\overline{S^3 \setminus H}; \mathbb{Z}) \setminus N$. Define $b_i = \mathrm{lk}(y, c_i)$ and $y' = y - \sum_{i=1}^n b_i[m_i]$. Then y' will have 0 as the linking number with every c_i, but it will not belong to N. The finite index of N implies that $ky' \in N$ for some k, but $\mathrm{lk}(ky', c_i) = k\dot{0} = 0$ for all i, and thus $ky' = 0$. This means that y' is a nonzero element of finite order in $H_1(\overline{S^3 \setminus H}; \mathbb{Z}) \equiv \mathbb{Z}^n$, but no such element exists. □

Lemma 3.5 *Let G be a connected ADG that has no DB values, which is not a double circle, and is gradable. Then the linking number of the boundary of its cross-surface with any simple cycle in the intersection graph of this cross-surface is 0.*

Proof Let C be a simple cycle in the intersection graph. It is composed of distinct vertices and edges $e_0, v_1, e_1, v_2, \dots, v_n, e_n = e_0$. Each v_i is a triple value, since it is not a degree-1 vertex. I will perturb C until it's in general position to the cross-surface and calculate the intersection number of the "moved C" with the cross-surface. This will be equal to the linking number of C and the boundary of the cross-surface.

I begin by pushing each edge e_i away from its matching X-bundle in a direction that agrees with the co-orientation on both of the surfaces that intersect in this X-bundle, as in Figure 11.

I need to continue this "pushing" at the vertex neighborhood of each v_i. Figures 12, 13 and 14 demonstrate how to push away the half-edges from their original position. The half-edges I push are colored green, and the arrows on them indicate the direction of the cycle: the half-edge whose arrow points toward (resp. away from) the triple value is a part of e_{i-1} (resp. e_i). Continuity dictates that I must always push in the direction

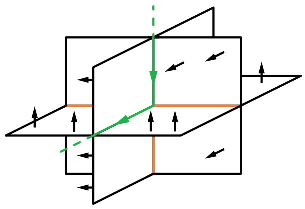

Figure 12: Moving the intersection graph away from a triple value when both sides are preferred

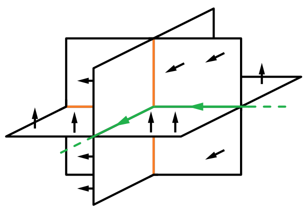

Figure 13: Moving the intersection graph away from a triple value when only one side is preferred

indicated by the orientations on the surface as we did in Figure 11, and Figures 12, 13 and 14 indeed comply with this.

Each of the three figures depict a different situation with regards to which of the two half-edges, if any, is preferred at v_i. Figure 12 depicts the case where both the half-edges are preferred. In this case, after being pushed away from the cross-surface, C will not intersect the cross-surface at the neighborhood of v_i.

Figure 13 depicts the case where the half-edge that is a part of e_{i-1}, the one entering the triple value, is not preferred, and the half-edge that is a part of e_i, the one exiting the triple value, is preferred. In this case, after being pushed away from the cross-surface, C will intersect the cross-surface once, and it will do so agreeing with the direction of the co-orientation on the surface (that's why there is a little $+1$ next to the intersection.)

Figure 13 depicts the case where the two half-edges are not consecutive, but even if they were, the same thing would happen; C would intersect the cross-surface once, in agreement with the co-orientation. The only difference would be that the half-edge that was exiting v_i would have continued leftwards instead of turning outwards towards the reader. Furthermore, had the half-edge coming from e_{i-1} been preferred and the one coming from e_i hadn't, then the pushing would still occur as in Figure 13, except that the arrows on the green line would point the other way. In this case, C would still intersect the cross-surface once after the pushing, but it would be against the direction on the co-orientation.

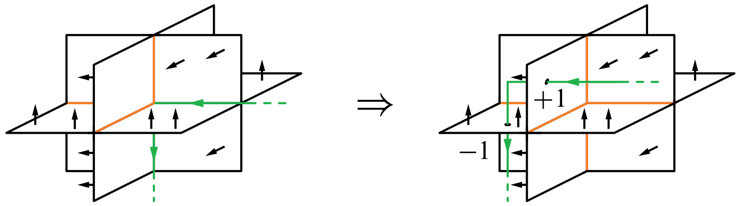

Figure 14: Moving the intersection graph away from a triple value when both sides are nonpreferred

Lastly, Figure 14 depicts the case in which both half-edges are not preferred. In this case, after being pushed away from the cross-surface, C will intersect the cross-surface twice in the neighborhood of v_i. One intersection, marked $+1$, is in the direction of the co-orientation, and the other intersection, marked -1, is against it.

Let G be a grading of the intersection graph. Since e_{i-1} and e_i share a vertex, the difference between their grading is at most 1. If $g(e_i)-g(e_{i-1})=1$ (resp. -1), then e_i (resp. e_{i-1}) is preferred and e_{i-1} (resp. e_i) is not. I just showed that in this case the signed number of intersections between the "pushed away" C and the cross-surface is 1 (resp. -1). If $g(e_i)-g(e_{i-1})=0$, then either both e_i and e_{i-1} are preferred, in which case C does not intersect the cross-surface around v_i, or they are both nonpreferred, in which case they intersect once with and once against the co-orientation.

In all cases the signed number of intersections between the pushed C and the cross-surface around v_i is equal to $g(e_i)-g(e_{i-1})$. The pushed C does not intersect the cross-surface anywhere else, and so their intersection number is $\sum_{i=1}^{n}(g(e_i)-g(e_{i-1})) = g(e_n)-g(e_0)=0$. Since C did not cross the boundary of the cross-surface during the pushing, this is equal to the linking number of C and the boundary. □

Having proven Lemmas 3.5 and 3.1, I can now prove the "if" direction of the articles of Theorem 2.7:

Proof (1) Each connected component G_k of G is gradable and lacks DB values, and thus has a realizing surface $i_k\colon F_k \to S^3$. Simply remove a point from $S^3 \setminus i_k(F_k)$ to regard i_k as a surface in $\mathbb{R}^3$, and embed these copies of $\mathbb{R}^3$ as disjoint balls in the interior of M.

(2) If G has no DB values, the proof of (1) holds. Otherwise, define a new ADG G' in which each DB value of G is replaced with a branch value. Realize G', via (1), with a closed generic surface $i\colon F \to S^3$ for which F is connected.

Take a small ball around each of the branch values that replaces a DB value of G, as in Figure 15 (left). Figure 15 (middle) depicts the intersection of the surface with the

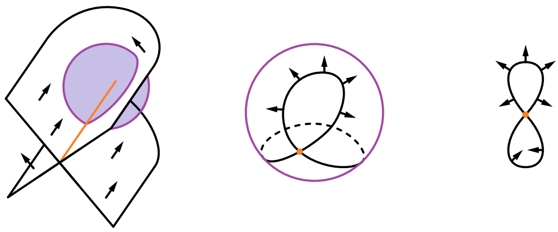

Figure 15: Turning a branch value into a DB value

boundary of the ball. It is an "8–figure" as in Figure 15 (right), and the orange dot (the intersection in the 8–figure) is the intersection of the boundary with the intersection graph. If you remove this ball from S^3, then instead of ending at the cross cap, the edge will end at the orange dot in the 8–figure, which will become a DB value. It follows that after removing all these balls, the intersection graph will be an ADG isomorphic to G.

The generic surface now lays in S^3 minus some number of balls. Choose one spherical boundary component and connect it via a path to each of the other ones. Make sure that the path is in general position to the generic surface: it may intersect the generic surface only at faces and will do so transversally. Thicken these paths into narrow 1–handles and remove them from the 3–manifold. This may remove some disc from the surface, but will not effect its intersection graph. You now have a generic surface that realizes G in D^3. Remove a point from the boundary of D^3, making it diffeomorphic to the closed half space $\{(x,y,z) \in \mathbb{R}^3 \mid z \geq 0\}$, which can be properly embedded in any 3–manifold with a boundary. This finishes the proof. □

Remark 3.6 If needed, you can make sure that the underlying surface F is connected. This involves modifying the surface in two ways.

(a) You can modify the proof of article (1) to produce a surface $i\colon F \to M$ with a connected image. Begin by assuming that the image of each i_k is connected via Remark 3.2. Pick a face v_k in each i_k. The co-orientation on v_k points towards a body U_k. When you remove a point from S^3, make sure you remove it from U_k. This way, U_k (minus a point) becomes the exterior body of $i_k\colon F_k \to \mathbb{R}^3$. When you embed the copies of $\mathbb{R}^3$ in M, the co-orientation on all the v_k will point towards the same connected component of $M \setminus \bigcup i_k(F_k)$. You may connect each V_k to V_{k+1} with a handle going through this component as in Figure 18 (ignore the letters "A" and "B" in the drawing). This connects the images of all the i_k without sacrificing the orientation or changing the intersection graph.

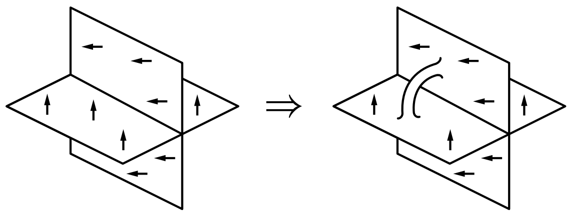

Figure 16: Turning a disconnected surface into a connected one

In article (2) you take a surface from (1) and modify it. It is clear that none of these modifications can disconnect the image of the surface, and so (2) may also produce surfaces with a connected image.

(b) In case $i(F)$ is connected but F has more than one connected component, the images of some pair of connected components must intersect generically at a double line. This is depicted in the left part of Figure 16, where the vertical surface comes from one connected component of F and the horizontal comes from another. Connect them via a handle in an orientation preserving way, as in the right part of Figure 16, thereby decreasing the number of connected components of F. Continue in this manner until F is connected.

4 Infinite homology

In this section, I deal with a 3–manifold whose first homology group contains an element of infinite order.

Theorem 4.1 *If M is an oriented, compact and boundaryless 3–manifold with an infinite first homology group, then any ADG G with no DB values can be realized as the intersection graph of an oriented generic surface in M. If M has a boundary, then any ADG G can be realized in M.*

The proof relies on two lemmas:

Lemma 4.2 *M has a connected, compact, oriented and properly embedded surface $S \subseteq M$ that is nondividing ($M \setminus S$ is connected).*

Proof $H_2(M;\mathbb{Z})$ is generated by 2–cycles of the form $[S]$ where $S \subseteq M$ is a connected, compact, oriented and properly embedded surface. If the statement of the lemma is false, then each such surface divides M into two connected components and will therefore be a boundary in $H_2(M;\mathbb{Z})$. This implies that $H_2(M;\mathbb{Z}) \equiv \{0\}$. According to

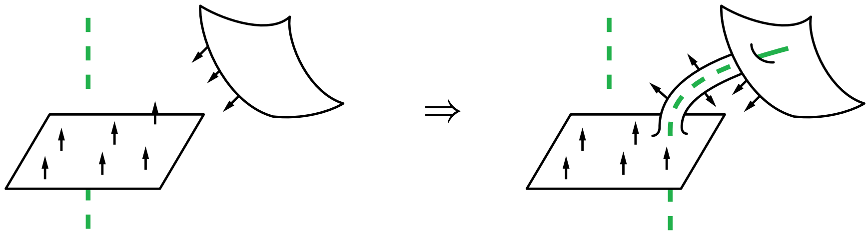

Figure 17: Giving the surface a face that has the same body on both sides

Poincaré's duality, $\{0\} \equiv H_2(M;\mathbb{Z})/\operatorname{Tor}(H_2(M;\mathbb{Z})) \equiv H_1(M;\mathbb{Z})/\operatorname{Tor}(H_1(M;\mathbb{Z}))$. This implies that every element of $H_1(M;\mathbb{Z})$ is of finite order, contradicting the assumption. □

Lemma 4.3 *If G is gradable, then there is a generic surface $i\colon F \to M$ which realizes G, and for which $M \setminus i(F)$ is connected (equivalently, i has only one body).*

Proof Take the generic surface $S \subseteq M$ from Lemma 4.2 and a subset $M' \subseteq M$ that is disjoint from S and is homomorphic to a half-space (if M has a boundary) or to $\mathbb{R}^3$ (if it does not). According to Theorem 2.7, there is a generic surface $i\colon F \to M'$ which realizes G. Connect some face V of the generic surface to S with a handle, as in Figure 17 (the handle does not intersect $i(F)$ or S). If needed, reverse the co-orientation of S so that the resulting surface will be continuously co-oriented.

You now have a new generic surface $i'\colon F \# S \to M$ whose intersection graph is still isomorphic to G. Since S was nondividing, the connected sum of V and S is a face of this surface that has the same body A on both sides (as indicated by the green path which does not intersect the surface in Figure 17). If this is the surface's only body, then you are done. If not, you can decrease the number of bodies as follows:

Let B be another body of the surface that is adjacent to A. Connect the face W which separates A and B to the face $V \# S$ with a path that goes through A, and does not intersect our generic surface except at the ends of the path. Since $V \# S$ has A on both sides, you can approach it from either side. If the arrows on W points toward A (resp. B), make sure the path enters $V \# S$ from the direction the arrows point towards (resp. point away from). Next, attach the faces V and W with a handle that runs along this path. Figure 18 depicts the case where the arrows on W point towards A. Reverse the direction of all arrows to get the other case.

The resulting generic surface has one body less since A and B have merged. It still realizes G and has a face with the same body on both sides. Repeat this process until you get a surface with only one body. □

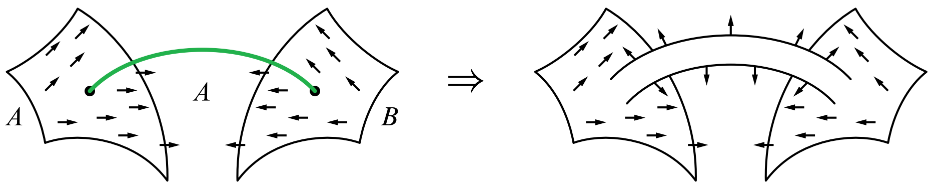

Figure 18: Reducing the number of bodies

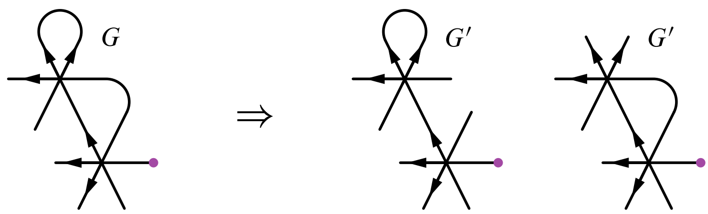

Figure 19: Cutting an edge and adding two branch values to an arrowed daisy graph

I will now prove Theorem 4.1:

Proof Let H be the graph part of G, that is, G without the double circles. I use induction on the genus of H. If the genus is 0, then G is the union of a forest with some double circles; Remark 2.6 implies that it is gradable, and the theorem follows from Lemma 4.3. If the genus of H is positive, pick an edge $e \in H$ such that $H \setminus \{e\}$ has a smaller genus. This means that removing e does not divide the connected component of H that contains e. Note that both ends of e are on triple values, since branch values and DB values are of degree 1, and removing their single edge divides the graph.

Define a new ADG G' in the following manner: Start with a copy of G and cut the edge e in the middle. Instead of e, you will get two "new edges" e_1 and e_2. Each e_i has one end on a new branch value while the other end "replaces" one of the ends of e; it enters the triple value that the said end of e was on and retains the ADG data, it is preferred if and only if the half-edge of e was preferred, and it has the same consecutive half-edge. Figure 19 depicts two possible ways to construct G' from a given graph G.

The graph structure of G', which we denote by H', has a lower genus then H. I assume, by induction, that there is a generic surface in M that realizes G' and *has only one body*. I will modify this surface so that it realizes G. Observe the new branch values at the ends of e_1 and e_2. Change the surface in a small neighborhood of each branch value as per Figure 20 (left), deleting the branch value and leaving instead a "figure 8 boundary" of the surface.

This figure 8 boundary is depicted in Figure 15 (right). Take a bundle over an interval whose fibers are "8–figures", as in Figure 20 (right), and embed it in M in such a way

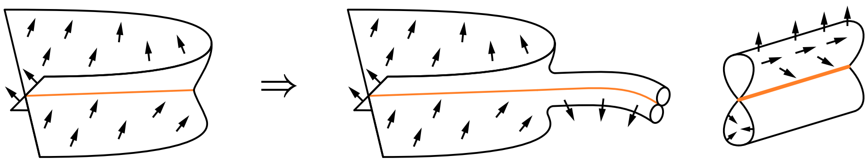

Figure 20: Removing two branch values from a generic surface and restoring the previously cut edge

that its end-fibers coincide with the said "figure 8 boundaries" (in a way that preserves the arrows of the co-orientation). Since the complement of the original surface was connected, you can make sure that the bundle does not intersect the surface anywhere except its ends. This closes e_1 and e_2 into one edge, reversing the procedure that created G' from G, and so this new surface realizes G while still having only one body. The proof follows by induction. □

Remark 4.4 It is possible once more to make sure that the underlying surface F is connected. First, you may connect the different connected components of $i(F)$ via handles, similarly to the way you connected faces in the proof of Lemma 4.3. You may then proceed as in Remark 3.6(b).

5 Ordered daisy graphs

In the last section, I will refine the enhanced graph structure of the intersection graph of a generic surface from an ADG to a structure I call an ordered daisy graph, or ODG, which encodes more information regarding the topology of the surface.

For motivation, attempt to construct a cross-surface for an ADG G as per Definition 3.3. Figure 21 (left) depicts a vertex neighborhood of a triple value v. Its preferred half-edges are indexed as $+1$, $+2$ and $+3$ (and the corresponding nonpreferred ones as -1, -2 and -3). While constructing the cross-surface, you glue the end of some X-bundle to each of these half-edges. Let $\sigma \in S_3$ be the even permutation $(1,2,3)$. If, for each $k = 1,2,3$, you take the "end of an X-bundle" that is supposed to be glued to the half-edge $\pm k$ and instead glue it to the half-edge $\pm\sigma(k)$, then you would end up with essentially the same cross-surface.

By "essentially the same cross-surface", I mean that you could pick neighborhoods $H_1 \subseteq S^3$ of one cross-surface and $H_2 \subseteq S^3$ of the other so that there is an orientation preserving homeomorphism $f : H_1 \to H_2$ that sends the first cross-surface to the second one in a manner preserving the co-orientation on them. Here f acts as a rotation

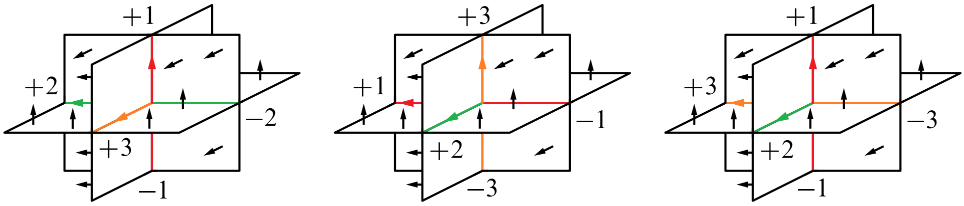

Figure 21: Even and odd permutations of a triple value neighborhood

in the neighborhood of v (the rotation that sends Figure 21 (left) to Figure 21 (middle)), and as the identity on the rest of the cross-surface.

Using an odd permutation, such as $\tau = (2,3)$, instead of σ, might produce a fundamentally different cross-surface. While I can define a similar f via a reflection on the neighborhood of v (such as the reflection that sends Figure 21 (left) to Figure 21 (right)), this f does not preserve the orientation on H_1. Furthermore, the two cross-surfaces may not even be homeomorphic subsets of S^3. You may complete each of the fundamentally different cross-surfaces into a generic surface. The intersection graphs of these two surfaces will be isometric as ADGs, but their neighborhoods, the cross-surfaces, will be topologically distinct. I refine the enhanced graph structure of the intersection graph so that it reflects the difference between them.

Definition 5.1 (1) An ordered daisy graph (ODG) (V, E, n, B, C, A, O) is a 7–tuple where (V, E, n, B, C, A) is an ADG (see Definition 1.4), and for each triple value v, we let $O(v)$ be an ordering (v_1, v_2, v_3) of the three preferred half-edges at v. This ordering is unique up to an even permutation. In the ODG of an oriented generic surface, I choose the ordering so that the triple value looks like Figure 21 (left) (and unlike Figure 21 (right)).

(2) A cross-surface of an ODG G is defined similarly to a cross-surface of an ADG (Definition 3.3) with the additional requirements that the vertex neighborhood of each triple value comes with indexed half-edges as per Figure 21 (left) and that, when you attach the X-bundles to such a vertex neighborhood, you comply with the indexing of the ODG structure.

The benefit of using ODGs is that all the cross-surfaces one might produce for the same ODG are essentially the same. All of the results in the previous sections, and in particular Theorems 2.7 and 4.1, can be rephrased to use ODGs instead of ADGs, and the same proofs will still apply.

Acknowledgments I would like to thank Dr Tahl Nowik, my Ph D adviser, for introducing me to this subject and to the wonderful and intriguing field of geometric topology, and my wife, Hila, for editing my work.

References

[1] **T F Banchoff**, *Triple points and surgery of immersed surfaces*, Proc. Amer. Math. Soc. 46 (1974) 407–413 MR

[2] **D Ben-Hadar**, *The liftings problem of generic surfaces is NP complete*, PhD thesis, Bar-Ilan University (2016) Available at `https://arxiv.org/abs/1605.08460`

[3] **J S Carter**, **M Saito**, *Surfaces in* 3*–space that do not lift to embeddings in* 4*–space*, from "Knot theory" (V F R Jones, J Kania-Bartoszyńska, J H Przytycki, V G Traczyk, Pawełand Turaev, editors), Banach Center Publ. 42, Polish Acad. Sci., Warsaw (1998) 29–47 MR

[4] **P R Cromwell**, **W L Marar**, *Semiregular surfaces with a single triple-point*, Geom. Dedicata 52 (1994) 143–153 MR

[5] **C A Giller**, *Towards a classical knot theory for surfaces in* $\mathbf{R}^4$, Illinois J. Math. 26 (1982) 591–631 MR

[6] **S Izumiya**, **W L Marar**, *The Euler characteristic of a generic wavefront in a* 3*–manifold*, Proc. Amer. Math. Soc. 118 (1993) 1347–1350 MR

[7] **K H Ko**, **J S Carter**, *Triple points of immersed surfaces in three-dimensional manifolds*, from "Proceedings of the 1987 Georgia Topology Conference" (N Habegger, C McCrory, editors), volume 32 (1989) 149–159 MR

[8] **G-S Li**, *On self-intersections of immersed surfaces*, Proc. Amer. Math. Soc. 126 (1998) 3721–3726 MR

[9] **S Satoh**, *Lifting a generic surface in* 3*–space to an embedded surface in* 4*–space*, Topology Appl. 106 (2000) 103–113 MR

Department of Mathematics, Bar-Ilan University
5290002 Ramat Gan, Israel

`doron.ben@live.biu.ac.il`

`http://math.biu.ac.il/node/641`

Received: 20 January 2016

Algebraic & Geometric Topology 17 (2017) 1701–1742

Embedding calculus knot invariants are of finite type

RYAN BUDNEY
JAMES CONANT
ROBIN KOYTCHEFF
DEV SINHA

We show that the map on components from the space of classical long knots to the n^{th} stage of its Goodwillie–Weiss embedding calculus tower is a map of monoids whose target is an abelian group and which is invariant under clasper surgery. We deduce that this map on components is a finite type-$(n-1)$ knot invariant. We compute the E^2–page in total degree zero for the spectral sequence converging to the components of this tower: it consists of $\mathbb{Z}$–modules of primitive chord diagrams, providing evidence for the conjecture that the tower is a universal finite-type invariant over the integers. Key to these results is the development of a group structure on the tower compatible with connected sum of knots, which in contrast with the corresponding results for the (weaker) homology tower requires novel techniques involving operad actions, evaluation maps and cosimplicial and subcubical diagrams.

55P65, 57M25

1 Introduction

We connect three current threads in studying knots and the moduli space of all knots: the Goodwillie–Weiss embedding calculus, Budney's operad actions and Vassiliev's theory of finite-type invariants. This work also connects two fundamental results on commutativity which are over fifty years old. In 1949, H Schubert [25] established and applied the fact that connected sum of knots is commutative. In 1947, Steenrod [30] gave explicit formulae exhibiting commutativity of cup product, in the course of defining the cohomology operations which bear his name. We establish and use compatibility of these classical results as we develop a group structure on the components of the Goodwillie–Weiss tower.

One of our main results is that the Goodwillie–Weiss tower for knots yields additive invariants of finite type. Such a result is modest compared to the conjecture made by Budney, Conant, Scannell and Sinha [4] that the tower gives all such invariants. We provide evidence for this conjecture through calculation of the E^2–term of the spectral

Published: 17 July 2017 DOI: 10.2140/agt.2017.17.1701

sequence for the components of this tower. The current results are meaningful first steps, and establishing them requires new combinations of tools in the areas of operad actions, clasper surgery, cosimplicial and cubical diagrams, and evaluation maps.

A large part of the work is "getting off the ground", showing that the Goodwillie–Weiss tower yields abelian group-valued invariants compatible with connected sum of knots. Because the knot invariants are defined as an induced map on π_0, such invariants are a priori only set-valued. A number of authors have already studied multiplicative structures on the closely related Goodwillie–Weiss tower for the homotopy fiber of the map from embeddings to immersions, namely Sinha [27], Turchin [31], Dwyer and Hess [8] and Boavida and Weiss [1], the latter three giving deloopings. The techniques of Turchin, and Boavida and Weiss, can be adapted to imply that the components of stages of the tower for classical framed knots is an abelian group (while the result of Dwyer and Hess applies to the limit and Sinha's result only implies an abelian monoid structure in the limit). Thus, including the present results, there are now five different monoid structures on the components of the inverse limit of the tower for framed classical knots, four of which are group structures, which are all conjecturally equivalent. For our present application, we need compatibility with connected sum of knots, which our approach here provides, as does the approach of Boavida and Weiss, which appeared after our work. Our techniques involve modeling the totalization tower of a cosimplicial space using cubical diagrams, which is of independent interest. Once we establish the group structure and compatibility with connected sum, we use the Habiro surgery criterion for finite type to establish the main result that the n^{th} stage in the tower defines a type-$(n-1)$ invariant.

Our work has many predecessors. The first is [4], in which we establish that the third stage of the tower is a universal type-2 invariant. Here the group structure and the finite-type result were straightforward from ad hoc arguments. The more interesting aspect of this work was the development of geometry to explicitly distinguish components in the tower, yielding a new interpretation of the type-2 knot invariant by counting collinearities of points. This geometric approach was continued by Flowers [9].

The second predecessor is Volić's thesis, which establishes that the Goodwillie–Weiss tower for the space level rational homology of knots is a universal rational finite-type invariant. The map to the homology tower factors through the tower we consider (the "homotopy tower"), which thus encodes such a universal rational invariant as well. The question of whether the tower gives a universal invariant over the integers is of considerable interest, since Vassiliev invariants with values in arbitrary $\mathbb{Z}$–modules are not well understood. A universal invariant is known to exist over the integers (see for example Habiro's ψ_n map, used in Theorem 6.1 of this paper), but it is not computable nor does it directly relate to the combinatorics of weight systems. Identification of the tower as a universal invariant would likely resolve questions about the combinatorics and geometry of finite-type invariants and, in particular, the open question of whether weight systems "integrate".

Our techniques are distinct from those of Volić in three ways. In our result the type-n invariants are potentially realized at the $n+1^{\text{st}}$ stage, while in the rational homology tower they are realized precisely at stage $2n$. Volić uses the Vassiliev derivative approach to establish his finite-type result, while we use Habiro's clasper surgery. Finally, Volić's result proceeds by extending the theory of Bott–Taubes integration to the homology tower, while our approach through the homotopy tower invites new techniques from geometric and algebraic topology.

We extend the abelian group structures to the spectral sequence level, which is crucial for analysis since studying components of an arbitrary tower of spaces can potentially lead to unending ad hoc calculations. To do so we employ C_1–operad actions, which necessitates their development throughout our work despite the fact that our main theorems are at the level of components. With group structure in hand, we immediately establish convergence at each finite stage and then use results of the second author [5] to identify the E^2 term in degree zero as the $\mathbb{Z}$–modules of primitive chord diagrams.

Based on this, we conjecture that the map from the knot space to the tower sends linear combinations of knots given by resolving a singular knot to the corresponding elements of E^2. We also conjecture that the spectral sequence collapses and, together, these two conjectures would imply that weight systems over the integers all "integrate" to finite-type invariants. Unlike Vassiliev's original spectral sequence, this spectral sequence does not involve a subtle limit process (see Giusti [10]) but instead is simply the spectral sequence of a tower of fibrations. It thus is more amenable to tools from algebraic topology such as generalized Hopf invariants.

The paper is organized as follows. Section 2 gives needed general background on compactified configuration spaces and on cubical and cosimplicial diagrams. Section 3 recalls the resulting mapping space and cosimplicial models we prefer to use, interchangeably, for the n^{th} stage of the Taylor tower for $\mathrm{Emb}^{\mathrm{fr}}(\mathbb{R},\mathbb{R}^3)$. Some readers may prefer to only look back at these sections as needed, in particular for the "infinitesimal transformation" map between the spatial and operadic mapping space models.

In Section 4, we construct homotopy-commutative multiplications on the n^{th} stage of the Taylor tower for $\mathrm{Emb}^{\mathrm{fr}}(\mathbb{R},\mathbb{R}^3)$. This shows that π_0 of each stage of the Taylor tower is an abelian monoid. Section 5 contains two main results, the first being that the projections in the tower are surjective on π_0. This is then used to show that π_0 of each stage is a group.

In Section 6, we show that π_0 of the map from the space of knots to its $n+1^{\text{st}}$ Goodwillie–Weiss approximation is invariant under clasper surgery and thus of type n. In Section 7, we show that the homotopy spectral sequence for the tower is a spectral sequence of abelian groups (in particular in total degrees zero and one), and we identify the E^2 term.

2 Compactification of configuration spaces

2.1 Basic definition

We briefly review the simplicial compactification of configuration space, defined in [26; 28]. For any manifold M — not necessarily compact, and possibly with boundary — let $C_n(M)$ denote the configuration space of n points in M. It is the space of distinct ordered n–tuples of points in M.

Definition 2.1 For an M embedded in some Euclidean space $\mathbb{R}^d$, let $C_n\langle M\rangle$ denote the closure of the image of

$$C_n(M) \hookrightarrow M^n \times (S^{d-1})^{\binom{n}{2}}, \quad (x_1, \dots, x_n) \mapsto \left((x_1, \dots, x_n), \left(\frac{x_i - x_j}{|x_i - x_j|}\right)_{i<j}\right).$$

As shown in [26], this is independent of embedding and is a quotient of $C_n[M]$, the Fulton–MacPherson compactification of $C_n(M)$. If M is noncompact, then $C_n\langle M\rangle$ as defined above as well as the Fulton–MacPherson compactification are not compact but are more accurately described as completions. Informally, we refer to the projection to M^n as the "spatial" information, while information which distinguishes points with the same projection to M is "infinitesimal". If M is one-dimensional and connected, then $C_n\langle M\rangle$ is in general not connected. For such M, by abuse we use $C_n\langle M\rangle$ to denote the connected component where the n points are in (cyclic) order.

2.2 Framings and tangent data

We need framed configurations. For any manifold M, define $C_n^{\mathrm{fr}}\langle M\rangle$ as the pullback

$$\begin{array}{ccc} C_n^{\mathrm{fr}}\langle M\rangle & \longrightarrow & (\mathrm{Fr}\, M)^n \\ \big\downarrow & & \big\downarrow \\ C_n\langle M\rangle & \longrightarrow & M^n \end{array}$$

where $\mathrm{Fr}\, M \to M$ is the unit frame bundle of the tangent bundle of M. If M is parallelizable and d–dimensional, then $C_n^{\mathrm{fr}}\langle M\rangle$ is homeomorphic to $C_n\langle M\rangle \times O(d)^n$.

Let $C'_n\langle M\rangle$ be defined similarly, with the unit tangent bundle STM in place of the unit frame bundle.

2.3 Distinguished boundary points

Suppose M has two distinguished points y_0 and y_1 in its boundary. Then, as in [4; 28], let $C_n\langle M,\partial\rangle$ denote the subspace of $C_{n+2}\langle M\rangle$ where the first and last points are located at y_0 and y_1, by abuse omitting dependence on these points from notation.

Further, if there are distinguished tangent vectors $v_0 \in TM|y_0$ and $v_1 \in TM|y_1$, let $C'_n\langle M,\partial\rangle$ be the subspace of $C'_{n+2}\langle M\rangle$ where (p_1, v_1) and (p_{n+2}, v_{n+2}) are (y_0, v_0) and (y_1, v_1), respectively. By fixing framings at y_0 and y_1, define $C^{\mathrm{fr}}_n\langle M,\partial\rangle$ similarly.

Define $C_n\langle \mathbb{I},\partial\rangle$ by taking the two endpoints to be the distinguished points. The fact that $C_n\langle \mathbb{I},\partial\rangle$ is the n–simplex is the main rationale for calling this compactification "simplicial". Define $C'_n\langle \mathbb{I}^d\rangle$ by taking $\{y_0, y_1\}$ to be $\partial\mathbb{I}\times\{(0,\dots,0)\}$ and $v_0 = v_1 = (1,0,\dots,0)$ (so our knots will "proceed from left to right") and similarly define $C^{\mathrm{fr}}_n\langle \mathbb{I}^d,\partial\rangle$ by using the identity element in $O(d)$ for framings at those boundary points (using the standard parallelization of $\mathbb{I}^d\subset\mathbb{R}^d$).

2.4 Quotients by translation and scaling, and insertion maps

There are maps between products of $C_n\langle\mathbb{R}^d\rangle$ and $C_n\langle \mathbb{I}^d,\partial\rangle$ defined by "inserting an infinitesimal configuration into a point of another configuration". Let $\widetilde{C}_n(\mathbb{R}^d) := C_n(\mathbb{R}^d)/(\mathbb{R}^d\rtimes\mathbb{R}_+)$ be the quotient of configuration space by translations and positive scalings of all n points.

Definition 2.2 Define $\widetilde{C}_n\langle\mathbb{R}^d\rangle$ as the closure of the image of the map

$$\widetilde{C}_n(\mathbb{R}^d) \xrightarrow{e} (S^{d-1})^{\binom{n}{2}}, \quad (x_1,\dots,x_n)\mapsto \left(\frac{x_i-x_j}{|x_i-x_j|}\right)_{i<j},$$

which is injective except on collinear configurations.

We let v_{ij} denote the projection of $\widetilde{C}_n\langle\mathbb{R}^d\rangle$ to the $(i,j)^{\mathrm{th}}$ factor of S^{d-1}.

The similarly defined $\widetilde{C}_n\langle\mathbb{I}^d\rangle$ is homeomorphic to $\widetilde{C}_n\langle\mathbb{R}^d\rangle$, so both $C_n\langle\mathbb{R}^d\rangle$ and $C_n\langle\mathbb{I}^d\rangle$ naturally surject onto $\widetilde{C}_n\langle\mathbb{R}^d\rangle$.

We proceed directly to the framed setting in defining insertion maps. Let $\widetilde{C}^{\mathrm{fr}}_n\langle\mathbb{R}^d\rangle := \widetilde{C}_n\langle\mathbb{R}^d\rangle\times O(d)^n$, the framed version of the "infinitesimal configuration space".

For every m, n, and $i\in\{1,\dots,n\}$, we define a map $\circ_i$ which, informally, inserts a configuration of m points (with framings) into the i^{th} point of a configuration of n

points (with framings). In the resulting configuration of $m+n-1$ points, the m points form an "infinitesimal configuration". Precisely, in coordinates we define

$$\circ_i\colon C_n^{\mathrm{fr}}\langle \mathbb{I}^d, \partial\rangle \times C_m^{\mathrm{fr}}\langle \mathbb{I}^d\rangle \to C_{m+n-1}^{\mathrm{fr}}\langle \mathbb{I}^d, \partial\rangle$$

as follows. First, suppressing the dependence on i, we let

$$\hat{j} = \begin{cases} j & \text{if } j \le i, \\ i & \text{if } i \le j \le i+m-1, \\ j+m-1 & \text{if } j > i+m-1. \end{cases}$$

Now define $\circ_i$ as sending

$$((x_j)_{j=1}^n, (u_{jk})_{j<k}, (\alpha_j)_{j=1}^n) \times ((y_j)_{j=1}^m, (v_{jk})_{j<k}, (\beta_j)_{j=1}^m)$$
$$\mapsto ((z_j)_{j=1}^{m+n-1}, (w_{jk})_{j<k}, (\gamma_j)_{j=1}^{m+n-1}),$$

where $z_j = x_{\hat{j}}$,

$$w_{jk} = \begin{cases} \alpha_i v_{(j-i+1)(k-i+1)} & \text{if } i \le j,k \le i+m-1, \\ u_{\hat{j}\hat{k}} & \text{otherwise,} \end{cases}$$

and

$$\gamma_j = \begin{cases} \alpha_i \beta_{j-i+1} & \text{if } i \le j \le i+m-1, \\ \alpha_{\hat{j}} & \text{otherwise.} \end{cases}$$

One must check that such maps send subspaces of $(\mathbb{R}^d)^n \times (S^{d-1})^{\binom{n}{2}}$ to each other appropriately. Here we only cite a similar check, namely [26, Proposition 6.6], using the description of the $C_n\langle \mathbb{R}^d\rangle$ as a subspace of $(\mathbb{R}^d)^n \times (S^{d-1})^{\binom{n}{2}}$ given in [26, Theorem 5.14]. The results are not dependent on the y_j coordinates of $C_m^{\mathrm{fr}}\langle \mathbb{I}^d\rangle$, which means that these insertion maps factor through the projection to $\widetilde{C}_m^{\mathrm{fr}}\langle \mathbb{R}^d\rangle$ on that factor.

3 The models

We study the space of framed knots $\mathrm{Emb}^{\mathrm{fr}}(\mathbb{R}, \mathbb{R}^3)$. A framed knot is a smooth embedding of $\mathbb{R}$ into $\mathbb{R}^3$ together with a smooth map $\mathbb{R} \to O(3)$ whose first (column) vector is the unit derivative map. Embeddings take $\mathbb{I} = [-1, 1] \subset \mathbb{R}$ into $\mathbb{I}^3 \subset \mathbb{R}^3$, and on $\mathbb{R} \setminus \mathbb{I}$ are standard, given by $t \mapsto (t, 0, 0)$. The framing is required to be constant at the identity on $\mathbb{R} \setminus \mathbb{I}$.

We primarily use a mapping space model for the n^{th} stage of the Goodwillie–Weiss tower for the space of framed knots $\mathrm{Emb}^{\mathrm{fr}}(\mathbb{R}, \mathbb{R}^d)$. But, for both spectral sequence calculations and for clarity through comparison, we use a cosimplicial model as well. For each of these, there are a few variants depending on the choice of compactifications of configuration spaces $C_n(M)$. The Fulton–MacPherson (Axelrod–Singer) compactification $C_n[M]$ is a smooth manifold with corners. In this paper, we use instead the simplicial compactification $C_n\langle M\rangle$, developed in the previous section. It is not a manifold with corners, but it has the advantage that one component of $C_n\langle \mathbb{I}\rangle$ is the n–simplex. This is needed to define a cosimplicial model, and the corresponding mapping space model is defined in terms of the face poset of the simplex rather than that of the associahedron.

3.1 Cosimplicial models

A cosimplicial space is a functor to $\mathcal{T}op$ from the category Δ with one object for each ordered set $[n]=(0,\dots,n)$ and morphisms given by order-preserving maps. We rely on the standard set of generating morphisms, denoting coface maps by d^i and codegeneracy maps by s^i. We let Δ_n be the full subcategory of Δ containing the first $n+1$ objects.

Goodwillie–Weiss embedding calculus leads to forming a cosimplicial space from the spaces $C^{\mathrm{fr}}_\bullet\langle \mathbb{I}^d,\partial\rangle$, by using the first vector in the framing as the "doubling direction". This was first done in the unframed setting in [28, Corollary 4.22]. In more detail, we have the following:

Definition 3.1 The spatial cosimplicial model $C^{\mathrm{fr}}_\bullet\langle \mathbb{I}^d,\partial\rangle$ has n^{th} entry $C^{\mathrm{fr}}_n\langle \mathbb{I}^d,\partial\rangle$.

The codegeneracy $s_i\colon C^{\mathrm{fr}}_n\langle \mathbb{I}^d,\partial\rangle \to C^{\mathrm{fr}}_{n-1}\langle \mathbb{I}^d,\partial\rangle$ is the extension to compactifications of the projection which forgets the i^{th} point.

The coface d^i is given by $d^i(x)=x\circ_i\mu$, which "doubles" the i^{th} point by inserting into its position the infinitesimal two-point configuration μ rotated by the i^{th} framing.

So here d^0 and d^{n+1} "double" the "extra" points which are located in the middle of the left and right faces of $\mathbb{I}^d$.

Definition 3.2 The operadic cosimplicial model $\widetilde{C}^{\mathrm{fr}}_\bullet\langle \mathbb{R}^d\rangle$ has n^{th} entry $\widetilde{C}^{\mathrm{fr}}_n\langle \mathbb{R}^d\rangle$.

The codegeneracy $s_i\colon \widetilde{C}^{\mathrm{fr}}_n\langle \mathbb{R}^d\rangle \to \widetilde{C}^{\mathrm{fr}}_{n-1}\langle \mathbb{R}^d\rangle$ is the extension to compactifications modulo translation and scaling of the projection which forgets the i^{th} point.

For $1\le i\le n$, the coface d^i is given by $d^i(x)=x\circ_i\mu$, which "doubles" the i^{th} point by inserting into its position the infinitesimal two-point configuration μ rotated by the i^{th} framing. The coface $d^0(x)$ equals $\mu\circ_2 x$, while the coface $d^{n+1}(x)$ equals $\mu\circ_1 x$.

Thus, in the operadic model, the first and last coface maps insert configurations into two-point configuration, which has the effect of "adding a point at infinity". Along with the obvious passing to quotients by translation and scaling, this is the difference between the spatial and operadic models. We discuss the relative advantages of these models as well as why we need both when we discuss their associated mapping space models below. The operadic model is so-called because it fits into a framework by which operads with multiplication produce cosimplicial objects, following Gerstenhaber and Voronov, and McClure and Smith [18].

The fourth author showed [27; 28] that the homotopy-invariant totalizations of similar cosimplicial spaces for unframed knots, which we will use and denote by $C'_\bullet\langle \mathbb{I}^d, \partial\rangle$ and $\widetilde{C}'_\bullet\langle \mathbb{R}^d\rangle$, give models for the Goodwillie–Weiss tower. The framed version of this construction was studied by Salvatore [23, Section 3].

By needing to use a homotopy-invariant totalization to model the towers, some control of geometry and combinatorics is lost. A standard approach to cosimplicial spaces through (sub)cubical diagrams, reviewed in the next section, retains both combinatorics and geometry by sacrificing some symmetry. In particular the resulting spatial model is compatible with the evaluation map, also known as a Gauss map, from the knot space.

3.2 Mapping space models

Our mapping space models are defined as homotopy limits of subcubical diagrams of compactified configuration spaces. A subcubical diagram is a functor from $\mathcal{P}_\nu[n]$, the poset of nonempty subsets of $[n]$, which is the face poset of the n–simplex. Because the cosimplicial and subcubical diagram categories both involve ordered sets, there is an immediate relationship between them. In general Δ admits a canonical functor from any category defined through finite ordered sets, as follows:

Definition 3.3 Let C be a category whose objects are given by ordered finite sets and whose morphisms are subsets of the order-preserving maps between those sets. Define $\mathcal{G}_C\colon C \to \Delta$ to be the functor which sends an ordered finite set S to $[\#S-1]$ and which sends an order-preserving map $S \to T$ to the composite

$$[\#S-1] \cong S \to T \cong [\#T-1],$$

where the isomorphisms are order-preserving.

For $C = \mathcal{P}_\nu[n]$, we abbreviate $\mathcal{G}_{\mathcal{P}_\nu[n]}$ to $\mathcal{G}_n\colon \mathcal{P}_\nu[n] \to \Delta_n$.

The functor $\mathcal{G}_n$ was used [28; 22] to use cubical diagrams to model cosimplicial spaces. The resulting diagrams use all of the coface maps "multiple times", while the codegeneracy maps are ignored.

Definition 3.4 The spatial mapping space model $\mathrm{AM}_n^{\mathrm{fr}}$ is the homotopy limit of the composite functor $C_\bullet^{\mathrm{fr}}\langle \mathbb{I}^3, \partial\rangle \circ \mathcal{G}_n \colon \mathcal{P}_\nu[n] \to \mathcal{T}op$.

The operadic mapping space model $\widetilde{\mathrm{AM}}_n^{\mathrm{fr}}$ is the homotopy limit of the composite functor $\widetilde{C}_\bullet^{\mathrm{fr}}\langle \mathbb{R}^3\rangle \circ \mathcal{G}_n \colon \mathcal{P}_\nu[n] \to \mathcal{T}op$.

We are henceforth focusing on classical knots, in three dimensions, so we are suppressing the ambient dimension from notation. The mapping space model AM_n of [4] is defined similarly to $\mathrm{AM}_n^{\mathrm{fr}}$, but pulled back from $C'_\bullet\langle \mathbb{I}^3, \partial\rangle$ and so with tangent vectors instead of frames.

Using the definition of homotopy limit through nerves of under-categories, the homotopy limit of a subcubical diagram is given by a collection of maps from simplices. Since the structure maps d^i are injective, an element φ of $\mathrm{AM}_n^{\mathrm{fr}}$ is determined by a map $\Delta^n \to C_n^{\mathrm{fr}}\langle \mathbb{I}^3, \partial\rangle$, so $\mathrm{AM}_n^{\mathrm{fr}}$ is a subspace of $\mathrm{Map}(\Delta^n, C_n^{\mathrm{fr}}\langle \mathbb{I}^3, \partial\rangle)$. Because the faces of the simplex map to configurations which are degenerate in an "aligned" manner we sometimes refer to this as the subspace of aligned maps. Explicitly, if some (consecutive) collection of t_i in $\vec{t} = (t_1, \dots, t_n) \in \Delta^n$ are equal, then the corresponding points in the configuration $\varphi(\vec{t})$ have "collided" in $\mathbb{I}^3$, their framings ($\alpha_i \in O(3)$) are all equal, and the first column of α_i is the direction of collision of these points.

To interrelate cosimplicial and mapping space models, a main technical result is [28, Theorem 6.7], which establishes that $\mathcal{G}_n$ is left cofinal. So if $X^\bullet$ is a cosimplicial space then the homotopy limit of the subcubical diagram $X^\bullet \circ \mathcal{G}_n$ is equivalent to the homotopy limit of the restriction of $X^\bullet$ to Δ_n. By work of Bousfield and Kan [2], this is homotopy equivalent to $\widetilde{\mathrm{Tot}}^n X^\bullet$, the n^{th} stage in the homotopy invariant totalization tower. In [28] this is used to establish the validity of the cosimplicial models, building from that of the mapping space models.

3.2.1 Evaluation maps Our main results make use of evaluation maps, which naturally connect with the spatial mapping space model. By functoriality for embeddings of compactified configuration spaces [26, Corollary 4.8], we have that an embedding $f \colon \mathbb{I} \to \mathbb{I}^3$ will extend to a map from $\Delta^n = C_n\langle \mathbb{I}, \partial\rangle$ to $C_n\langle \mathbb{I}^3, \partial\rangle$. For a framed embedding with framing $\alpha \in \mathrm{Map}(\mathbb{I}, O(3))$ to go along with the embedding f, this map is given as follows:

Definition 3.5 Define $\mathrm{ev}_n \colon \mathrm{Emb}^{\mathrm{fr}}(\mathbb{R}, \mathbb{R}^3) \to \mathrm{AM}_n^{\mathrm{fr}} \subset \mathrm{Map}(\Delta^n, C_n\langle \mathbb{I}^3, \partial\rangle \times O(3)^n)$ as sending an embedding f and framing α to the map which sends

$$(-1 \le t_1 \le \dots \le t_n \le 1) \mapsto \big((f(t_1), \dots, f(t_n)), \alpha(t_1), \dots, \alpha(t_n)\big).$$

Intuitively, ev_n samples the knot and its framing at n points in the domain.

One of the main results of [28], namely Theorem 5.4, immediately extends to the framed setting to establish that ev_n agrees in the homotopy category with the canonical map from the embedding space $\mathrm{Emb}^{\mathrm{fr}}(\mathbb{R}, \mathbb{R}^d)$ to the n^{th} stage of the Taylor tower, $T_n\,\mathrm{Emb}$.

In dimensions $d > 3$ it is known that the connectivity of ev_n increases with n, and hence the tower converges to the embedding space. Thus, the homotopy theory of evaluation maps captures the nature of knots in these dimensions. For $d = 3$, it is not known if ev_n is even 0–connected (that is, surjective) much less injective in the limit. We show below that on components ev_n defines abelian group-valued invariants of finite type $n-1$.

3.2.2 Translation between mapping space models In addition to the quotienting by translation and scaling, AM_n and $\widetilde{\mathrm{AM}}_n$ differ by the first and last coface maps, which in the former case "add a point to the configuration on the left or right face of $\mathbb{I}^3$" while in the latter case such points are added "at infinity to the left or right". This means that the obvious quotient maps do not define a map between these models. We require a map between them, as the spatial model is the target of the evaluation map as we just saw and the operadic model is more convenient for defining multiplications below.

Observe that the evaluation of a long knot can be extended to include times greater than 1 or less than -1, in which case the corresponding configuration points are standard along the x–axis. We achieve something similar for arbitrary elements of AM_n to obtain maps $\widetilde{\varphi}\colon C_n\langle\mathbb{R}\rangle \to C_n\langle\mathbb{R}^3, \partial\rangle$.

Definition 3.6 For $t_1 \le t_2 \le \cdots \le t_n$ with all $t_i \in \mathbb{R}$, we let

$$\widehat{t_i} = \begin{cases} -1 & \text{if } t_i \le -1, \\ t_i & \text{if } -1 \le t_i \le 1, \\ 1 & \text{if } t_i \ge 1. \end{cases}$$

Then we define $\widetilde{\varphi}(t_1 \le t_2 \le \cdots \le t_n)$ to be the configuration which is the union of the following:

(1) One point at $(t_i, 0, 0)$ for each $t_i \le -1$ or $t_i \ge 1$.

(2) The configuration obtained by taking $\varphi(\widehat{t_1} \le \cdots \le \widehat{t_n})$ and applying the projection which forgets points x_i for which $t_i \le -1$ or $t_i \ge 1$.

Moreover, after we compose with the quotient $C_n\langle\mathbb{R}^3, \partial\rangle \to \widetilde{C}_n\langle\mathbb{R}^3, \partial\rangle$, the map $\widetilde{\varphi}$ extends to a map from $C_n\langle[-\infty, \infty]\rangle$ to $\widetilde{C}_n\langle\mathbb{R}^3, \partial\rangle$ since the limit as some t_i goes to $-\infty$ or ∞ will have all vectors v_{ij} approaching $(1, 0, 0)$ or $(-1, 0, 0)$, as all points in the resulting configurations not in $\mathbb{I}^3$ lie along the x–axis.

Let h be an order-preserving homeomorphism between $[-1,1]$ and $[-\infty,\infty]$. By our usual abuse of notation, let h also denote the induced map on collections of points.

Definition 3.7 Let $\iota\colon \mathrm{AM}_n \to \widetilde{\mathrm{AM}}_n$ be the map which sends φ to the composite

$$\Delta^n \xrightarrow{h} C_n\langle[-\infty,\infty]\rangle \xrightarrow{\widetilde{\varphi}} \widetilde{C}_n\langle\mathbb{R}^3,\partial\rangle.$$

Proposition 3.8 *The map ι is a homotopy equivalence.*

Proof To compare AM_n and $\widetilde{\mathrm{AM}}_n$ it is simplest to map them both to a third space which is easily seen to be equivalent. Let $\widetilde{\mathrm{AM}}_n^{\mathrm{hs}}$, where "hs" stands for hemispherical, be the subspace of $\mathrm{Map}(\Delta^n, \widetilde{C}_n\langle\mathbb{I}^3\rangle)$ defined by the same conditions as for $\widetilde{\mathrm{AM}}_n$ for all faces except the $t_1=-1$ and $t_n=1$ faces. On those faces, instead of the vectors v_{1j} and v_{kn} being $(1,0,0)$ as in the definition of $\widetilde{\mathrm{AM}}_n$, for $\widetilde{\mathrm{AM}}_n^{\mathrm{hs}}$ those vectors are simply required to lie in the hemisphere with nonnegative x–coordinate. (We could also define this as a homotopy limit.)

Consider the diagram

$$\begin{tikzcd} \mathrm{AM}_n \arrow[d, "q"', "\wr"] \arrow[dr, "\iota"] & \\ \widetilde{\mathrm{AM}}_n^{\mathrm{hs}} & \widetilde{\mathrm{AM}}_n \arrow[l, "\simeq"', "i"] \end{tikzcd}$$

where q is the map induced by composition of $\Delta^n \to C_n\langle\mathbb{I}^3,\partial\rangle$ with the quotient map to $\widetilde{C}_n\langle\mathbb{I}^3\rangle$ and i is the inclusion. Both of these are homotopy equivalences. The diagram commutes up to homotopy, interpolating between q and $i\circ\iota$ by first extending $\widetilde{\varphi}$ to $[-x,x]$. Then use a continuous family of homeomorphisms h_x of $\mathbb{I}$ with $[-x,x]$ to define a map ι_x which serves as a homotopy. It is then elementary that the composite of i with the homotopy inverse of q serves as a homotopy inverse to ι. □

The main idea in the definition of ι, and in particular the construction of $\widetilde{\varphi}$, is "focusing" on the interval $\mathbb{I}=[-1,1]$ by modifying configurations in $\mathbb{R}$ with points outside the interval to have points at the endpoints of the interval instead. Corresponding points in the configuration are replace by standard points along the x–axis. We will use similar ideas in the development of multiplicative structures on AM_n and $\widetilde{\mathrm{AM}}_n$.

4 Multiplicative structures

In this section we first define an action of the little intervals operad $\mathcal{C}_1$ on the framed knot space and $\mathrm{AM}_n^{\mathrm{fr}}$, which are compatible via the evaluation map. Then we construct a multiplication which is homotopy commutative on $\widetilde{\mathrm{AM}}_n^{\mathrm{fr}}$. We show that these actions are all compatible up to homotopy.

Definition 4.1 The little intervals operad $\mathcal{C}_1$ has as its n^{th} entry $\mathcal{C}_1(n)$ the space of n disjoint subintervals of $\mathbb{I}$, which we topologize as a subspace of $\mathbb{I}^{2n}$ through the endpoints of the intervals.

Given an subinterval $L \subseteq \mathbb{I}$ we by abuse let L also denote the orientation-preserving affine-linear transformation which sends $\mathbb{I}$ to L.

For use with operad actions on knot spaces, we let $\widehat{L}$ denote the map $\mathbb{I}^3 \to \mathbb{I}^3$ which applies L to the first coordinate and then shrinks the second and third coordinates by the same scaling factor (but doesn't translate them).

4.1 The little intervals action on the knot space

If $\mathcal{L} = \bigcup L_i$ is a union of k disjoint little intervals, its action on a k–tuple of embeddings $f_i\colon \mathbb{R} \to \mathbb{R}^3$ yields the embedding which at time t has value

$$\mathcal{L} \cdot (f_1, \dots, f_k)(t) = \begin{cases} \widehat{L}_i \circ f_i \circ {L_i}^{-1}(t) & \text{if } t \in L_i \text{ for some } i, \\ (t, 0, 0) & \text{otherwise.} \end{cases}$$

That is, the embeddings are "shrunk and placed in succession" according to $\mathcal{L}$. The action on $\operatorname{Emb}^{\text{fr}}(\mathbb{R}, \mathbb{R}^3)$ is similar, with the framings unchanged by the shrinking. We will view this action as a case of "insertion into the trivial embedding, with standard framing".

4.2 The spatial little intervals action on aligned maps

Defining a $\mathcal{C}_1$ action on AM_n is more involved, guided by wanting the evaluation map ev_n to be compatible with the action. We want to take a configuration in the interval and evaluate points in the various L_i on different elements of AM_n. But typically fewer than n points will be in each L_i, so we adjust accordingly. We first define restriction of an aligned map to some interval L.

Definition 4.2 For $L \subset \mathbb{I}$, define L^{-1} on some $\vec{t}$ in the interval by applying to each t_i the piecewise-linear map which is L^{-1} on the image of L, sends points to the left of L to -1, and sends points to the right of L to 1.

Define the restriction of $\varphi \in \mathrm{AM}_n$ to L, denoted by $\varphi|_L$, by applying φ to $L^{-1}(\vec{t})$ and then applying projection maps to forget all of the points in the resulting configurations whose indices j do not correspond to a t_j in the interior of L.

This is not continuous, as the different projection maps depending on the number of t_i in L produce points in different configuration spaces, but it is an essential auxiliary construction.

Definition 4.3 Define the action of a union of little intervals $\mathcal{L} = \bigcup L_i$ on a collection of maps $\varphi_i \in \mathrm{AM}_n$ as the map in AM_n which takes $\vec{t} \in \Delta^n$ to the union of all of the $\widehat{L}_i \circ \varphi_i|_{L_i}$ applied to $\vec{t}$ along with a point in the configuration at $(t_j, 0, 0)$ for each t_j which is not in the interior of any L_i. If for such a t_j we have, say, $t_j = t_{j+1}$, then we set the "direction of collision from $(t_j, 0, 0)$ to $(t_{j+1}, 0, 0)$", denoted by $v_{i,i+1}$ when we have used coordinates, to be the positive x–axis direction.

The $\mathcal{C}_1$–action on $\mathrm{AM}_n^{\mathrm{fr}}$ is defined in the same manner, where the framings are unchanged because each interval shrinks an aligned map equally in all directions.

This action varies continuously as t_j approaches an endpoint e of some L_i because the limit from either side of this endpoint is the configuration with the corresponding point x_j at $(e, 0, 0)$. Checking continuity elsewhere is immediate, as is checking the usual conditions required for an action of the operad $\mathcal{C}_1$. The definition was arranged so that this action is compatible with the $\mathcal{C}_1$ action on the knot space via the evaluation map.

Proposition 4.4 *The map* $\mathrm{ev}_n\colon \mathrm{Emb}^{\mathrm{fr}}(\mathbb{R}, \mathbb{R}^3) \to \mathrm{AM}_n^{\mathrm{fr}}$ *is a map of* $\mathcal{C}_1$*–spaces.*

For this paper we only need that this is a map of H–spaces. We will also need that the action is compatible with the structure maps in the Goodwillie–Weiss tower, which will be a main focus of Section 5.

4.3 A spatial-infinitesimal single little interval action

In his work on operad actions on knot spaces [3], the first author extensively uses the fact that embeddings of $\mathbb{R} \times D^2$ into $\mathbb{R} \times D^2$ can be composed. Our main idea in establishing homotopy commutativity of the multiplication(s) on $\mathrm{AM}_n^{\mathrm{fr}}$ is to set up composition. In order to do so, we produce products where the spatial points are along the evaluation map of the unknot, and use infinitesimal composition for the essential part of the multiplication. Given an interval L and $\varphi \in \widetilde{\mathrm{AM}}_n^{\mathrm{fr}}$, we produce an aligned map with infinitesimal configuration at the point $(L(0), 0, 0)$ together with some points along $\mathbb{I} \times (0, 0)$. For continuity, we need points near the boundary and, say, inside of L to be pulled towards $L(0)$.

Let $L^\circ = L\big(\big[-\frac{1}{2}, \frac{1}{2}\big]\big)$, which is the "core" of L. Let $e_L\colon \mathbb{I} \to \mathbb{I}$ be a monotone continuous map which sends L° to the point $L(0)$ and which is the identity outside of L. By abuse use the same notation e_L for the induced map on collections of points in the interval.

As in previous constructions, we define the map $L \cdot \varphi$ piecewise on Δ^n according to the partition defined by which indices of t_i occur in $\vec{t} \cap L^\circ$. Let i and j be the indices of the leftmost and rightmost t_k in L°, which we will consider as functions of L° and $\vec{t}$. For an element $\varphi \in \widetilde{\mathrm{AM}}_n^{\mathrm{fr}}$, let $\varphi|_L$ be defined similarly as in Definition 4.2 for the AM_n setting, though this now produces equivalence classes of configurations modulo translation and scaling, with framings.

Definition 4.5 With notation as above, for $\varphi \in \widetilde{\mathrm{AM}}_n$ define $L \cdot \varphi \in \mathrm{AM}_n$ by

$$(L \cdot \varphi)(\vec{t}) = \big(e_L(t_1, \dots, t_{i-1}, L(0), t_{j+1}, \dots, t_n) \times (0,0)\big) \circ_i \varphi|_{L^\circ}(\vec{t}).$$

For a fixed L, it is clear that the output varies continuously with φ and with all $\vec{t}$ which have the same indices occurring in $\vec{t} \cap L^\circ$. To check that the various pieces fit together to a continuous function, suppose that some t_i is equal to $L\big(-\frac{1}{2}\big)$, the left endpoint of L°. The formulae on the two pieces of Δ^n that meet at $t_i = L\big(-\frac{1}{2}\big)$ are

$$\big(e_L(t_1, t_{i-1}, L(0), t_{j+1}, \dots, t_n) \times (0,0)\big) \circ_i \varphi((L^\circ)^{-1}(t_i, t_{i+1}, \dots, t_j))$$

and, using that $e_L(t_i) = e_L\big(L\big(-\frac{1}{2}\big)\big) = e_L(L(0)) = L(0)$,

$$\big(e_L(t_1, \dots, t_{i-1}, L(0), L(0), t_{j+1}, \dots, t_n) \times (0,0)\big) \circ_{i+1} \varphi((L^\circ)^{-1}(t_{i+1}, \dots, t_j)).$$

The key point is that, by definition of the doubling maps in $\widetilde{C}^{\mathrm{fr}}_\bullet\langle \mathbb{I}^3, \partial\rangle$, the map φ sends a point in $\partial\mathbb{I}$ to a point "that looks infinitely far away from the images of the interior points in $\mathbb{I}$". Here it is essential that we are starting with elements of the operadic mapping space model rather than the spatial model.

To elaborate, using our standard coordinates on these compactifications, $v_{ik} = (1,0,0)$ for any $k \in \{i+1, \dots, j\}$ in either of the two configurations above, and every other $v_{k\ell}$ for $1 \le k \le \ell \le n$ is the same in these two configurations as well. Furthermore, the projections to $(\mathbb{I}^3)^n$ of the two expressions agree, so they are the same configuration in $C_n^{\mathrm{fr}}\langle \mathbb{I}^3, \partial\rangle$.

Checking continuity between any other two pieces of Δ^n is similar, as is checking continuity if we vary both $\vec{t}$ and L.

If we compose $L \cdot \varphi$ with the projection to $\widetilde{C}_n\langle \mathbb{R}^3\rangle$, the resulting map satisfies the conditions of being in $\widetilde{\mathrm{AM}}_n$. We need the spatial information in the next section.

4.4 A commutative multiplication on infinitesimal aligned maps

Now we define a multiplication on $\widetilde{\mathrm{AM}}_n^{\mathrm{fr}}$ determined by a choice of two intervals L_1, $L_2 \in \mathcal{C}_1(1)$, one entry in the operad of "overlapping intervals". Here, when two intervals overlap, they are also ordered. Informally, this order says which interval is "on top". Since $\mathcal{C}_1(1)$ is connected, this product will be homotopy-commutative.

To see that homotopy-commutativity of connected sum holds for knots coherently as part of an operad action, it is technically necessary to "thicken" the knots, by studying embeddings of $\mathbb{R} \times D^2$ in $\mathbb{R} \times D^2$. In [3], the first author shows that such embedding spaces have an action of the overlapping intervals operad and, moreover, uses this action to determine the homotopy type of the space of classical knots. As we work with evaluation maps between compactified configuration spaces, our substitute for $\mathbb{R} \times D^2$ is the following:

Definition 4.6 Let $C_n^{\mathrm{it}}\langle \mathbb{I} \times (0,0), \partial\rangle$, the space of infinitesimally thickened configurations in an interval, be the subspace of $C_n^{\mathrm{fr}}\langle \mathbb{I}^3, \partial\rangle$ whose image under the projection to $(\mathbb{I}^3)^n$ lies in $(\mathbb{I} \times (0,0))^n$.

So far, $L \cdot \varphi$ is a map $C_n\langle \mathbb{I}, \partial\rangle \to C_n^{\mathrm{fr}}\langle \mathbb{I}^3, \partial\rangle$ whose image lies in the subspace $C_n^{\mathrm{it}}\langle \mathbb{I} \times (0,0), \partial\rangle$. We can view $C_n\langle \mathbb{I}, \partial\rangle$ as a subspace of $C_n^{\mathrm{it}}\langle \mathbb{I} \times (0,0), \partial\rangle$ in an obvious way with identity framings at every point. We will now extend the domain of $L \cdot \varphi$ from $C_n\langle \mathbb{I}, \partial\rangle$ to all of $C_n^{\mathrm{it}}\langle \mathbb{I} \times (0,0), \partial\rangle$.

For any $c \in C_n^{\mathrm{it}}\langle \mathbb{I} \times (0,0), \partial\rangle$, let $\vec{t}(c) = (t_1 < \cdots < t_m)$ be the set of *distinct* points in $p(c)$ (so $m \le n$). Then c can be written as

$$c = \big(\cdots\big((\vec{t}(c) \times (0,0)) \circ_{m_1} c_1\big) \circ_{m_2} \cdots\big) \circ_{m_k} c_k \tag{1}$$

for some $c_i \in \widetilde{C}_{n_i}\langle \mathbb{I}^3\rangle$. This expression is unique once we require that $m_i \ge m_{i-1} + n_{i-1}$, so that the underlying points of insertion are distinct.

Definition 4.7 Define the extension of $L \cdot \varphi$ to $C_n^{\mathrm{it}}\langle \mathbb{I} \times (0,0), \partial\rangle$ as

$$(L \cdot \varphi)(c) = \big(\cdots\big((L \cdot \varphi)(\vec{t}(c)) \circ_{m_1} c_1\big) \circ_{m_2} \cdots\big) \circ_{m_k} c_k.$$

Here the $L \cdot \varphi$ on the right-hand side is as in Definition 4.5, using the identification of $C_n\langle \mathbb{I}, \partial\rangle$ with Δ^n.

The key point now is checking continuity when points enter or exit the infinitesimal configurations c_i. The argument is just as for continuity in Definition 4.5 but with $(1,0,0)$ replaced by the tangent vector, which is the first vector in the framing, at $\varphi(t_{m_i})$.

Now both the input and output of the map $L \cdot \varphi$ can be regarded as elements in $C_n^{\mathrm{it}}\langle \mathbb{I} \times (0,0), \partial\rangle$. Thus two such maps can be composed, and we denote the composition by $\circ$.

Definition 4.8 Given two little intervals $L_1,\ L_2 \in \mathcal{C}_1(1)$, define the product of elements $\varphi,\ \psi \in \widetilde{\mathrm{AM}}_n^{\mathrm{fr}}$ as

$$\mu_{L_1,L_2}(\varphi, \psi)\colon \vec{t} \mapsto ((L_2 \cdot \psi) \circ (L_1 \cdot \varphi))(\vec{t}),$$

where we take the equivalence class in $\widetilde{C}_n^{\mathrm{fr}}\langle \mathbb{I}^3\rangle$ of the right-hand side.

Let $L_- = [-1,0]$ and $L_+ = [0,1]$. Abbreviate μ_{L_-,L_+} as μ, and abbreviate $\mu(\varphi,\psi)$ as $\varphi\cdot\psi$.

Theorem 4.9 *$\varphi\cdot\psi$ is homotopic to $\psi\cdot\varphi$.*

Proof The map $(L,\varphi)\mapsto L\cdot\varphi$ varies continuously with L. Because $\mathcal{C}_1(1)$ is connected, any two multiplications induced by choices of (L_1,L_2) — in particular (L_-,L_+) and (L_+,L_-) — are homotopic. □

4.5 A compatible little intervals action on infinitesimal aligned maps

We now compare the multiplications on the spatial model $\mathrm{AM}_n^{\mathrm{fr}}$ and the infinitesimal model $\widetilde{\mathrm{AM}}_n^{\mathrm{fr}}$. We first present a simpler multiplication μ' on $\widetilde{\mathrm{AM}}_n^{\mathrm{fr}}$, which is homotopic to the one defined above. The multiplication μ' avoids the use of the maps e_L which pull points towards L, which are needed homotopy-commutativity. The multiplication μ' also has the advantage of extending to an action of the little intervals operad.

Definition 4.10 Let μ' be the map

$$\mu'\colon \mathcal{C}_1(2)\times(\widetilde{\mathrm{AM}}_n^{\mathrm{fr}})^2 \to \widetilde{\mathrm{AM}}_n^{\mathrm{fr}}$$

defined by

$$\begin{aligned}\mu'_{L_1,L_2}(\varphi,\psi)(\vec{t}) \\ := \big(\big((t_1,\dots,t_{i-1},L_1(0),t_{j+1},\dots,t_{k-1},L_2(0),t_{\ell+1},\dots,t_n)\times(0,0)\big)\circ_i\varphi|_{L_1}(\vec{t})\big) \\ \circ_k \psi|_{L_2}(\vec{t}),\end{aligned}$$

where $t_i,\dots,t_j$ are the points in L_1 and $t_k,\dots,t_\ell$ are the points in L_2. Similarly, for any $k\ge 1$, define a map

$$\mathcal{C}_1(k)\times(\widetilde{\mathrm{AM}}_n^{\mathrm{fr}})^k \to \widetilde{\mathrm{AM}}_n^{\mathrm{fr}} \tag{2}$$

which sends $((L_1,\dots,L_k),(\varphi_1,\dots,\varphi_k))$ to the result of inserting each $\varphi_i|_{L_i}$ into the point $(L_i(0),0,0)$.

It is straightforward to verify that the maps (2) define a $\mathcal{C}_1$–action on $\widetilde{\mathrm{AM}}_n^{\mathrm{fr}}$, using the fact that $\widetilde{C}_n\langle\mathbb{I}^3\rangle$ records only directions of collision, as well as again using the fact that in the operadic models configuration points which corresponding to times t_i which are equal to ±1 are "at infinity".

Proposition 4.11 *For disjoint intervals L_1 and L_2, the multiplication $\mu_{L_1,L_2}(\varphi,\psi)$ is homotopic to the multiplication $\mu'_{L_1,L_2}(\varphi,\psi)$.*

Proof Since configurations of points along the x–axis are all equivalent in $\widetilde{C}_n\langle \mathbb{I}^3\rangle$, the resulting aligned maps differ only in the subsets of $\mathbb{I}$ on which they are constant. The aligned map resulting from μ' sends all t_j between L_1 and L_2 to the same configuration point, while the same is true for μ and L_1° and L_2°. These are related by a homotopy which "reparametrizes the domains" of φ and ψ from L_1° and L_2° to L_1 and L_2, respectively. □

Recall from Definition 3.7 our transformation ι from the spatial to the operadic mapping space models, which is an equivalence.

Proposition 4.12 *The $\mathcal{C}_1$–actions on $\mathrm{AM}_n^{\mathrm{fr}}$ and $\widetilde{\mathrm{AM}}_n^{\mathrm{fr}}$ are compatible. That is, the diagram*

$$\begin{tikzcd} \mathcal{C}_1(k)\times(\mathrm{AM}_n^{\mathrm{fr}})^k \arrow[r] \arrow[d, "\iota^k"'] & \mathrm{AM}_n^{\mathrm{fr}} \arrow[d, "\iota"] \\ \mathcal{C}_1(k)\times(\widetilde{\mathrm{AM}}_n^{\mathrm{fr}})^k \arrow[r] & \widetilde{\mathrm{AM}}_n^{\mathrm{fr}} \end{tikzcd}$$

commutes up to homotopy, where the top horizontal map is the action given in Definition 4.3 and the bottom horizontal map is the action that generalizes the multiplication μ' given in Definition 4.10.

Proof Recall that the $\mathcal{C}_1$ action on $\mathrm{AM}_n^{\mathrm{fr}}$ is defined by taking the union of the images of $\widehat{L}_i\circ\varphi_i|_{L_i}$, along with points along the x–axis. We will perform a homotopy with three steps, parametrized by $s\in[0,3]$, from the composite through the upper-right corner to the composite through the lower-left corner.

The first step, as s varies in $[0,1]$, is a straightforward homotopy from $(L_1,\dots,L_k)\cdot(\varphi_1,\dots,\varphi_k)$ to $(L_1^\circ,\dots,L_k^\circ)\cdot(\varphi_1,\dots,\varphi_k)$, where as before $L^\circ := L\bigl(\bigl[-\frac{1}{2},\frac{1}{2}\bigr]\bigr)$. The configuration produced by an aligned map in the image at $s=1$ is the union of configurations resulting from all the $\varphi_i|_{L_i^\circ}$, together with points along the x–axis. Roughly, we are ensuring that each aligned map is standard on some interval between each pair of intervals L_i and L_{i+1}, so that we can push apart and shrink configurations continuously.

Next define $L_i^\times := L_i\setminus L_i^\circ$. In the second step, as s varies in $[1,2]$, we scale the image of each $\varphi_i|_{L_i^\circ}$ to an infinitesimal configuration at $(L_i(0),0,0)$. In the notation of Section 4.1, we follow each aligned map by $(\widehat{L}_i^\circ)^{-1}\circ\widehat{J}_s\circ\widehat{L}_i^\circ$, where we define $J_s := [s-2,2-s]$ for $s\in[1,2]$. This can be done at the level of representatives in $C_n^{\mathrm{fr}}\langle\mathbb{I}^3,\partial\rangle$ coming from the top horizontal map. The configurations become infinitesimal only at $s=2$, so continuity requires us to simultaneously pull the images of $L_i^\times$ toward $(L_i(0),0,0)$, from occupying $L_i^\times\times\{(0,0)\}$ at $s=1$ to occupying all of $L_i\times\{(0,0)\}$ at $s=2$.

At $s = 2$, the elements in $\widetilde{\mathrm{AM}}_n^{\mathrm{fr}}$ produce infinitesimal configurations $\varphi_i|_{L_i^\circ}$, together with points on the x–axis between them (where the distances between them are *not* recorded by the operadic compactification $\widetilde{C}_n^{\mathrm{fr}}\langle \mathbb{I}^3 \rangle$). This aligned map looks almost like the composite through the lower-left corner. The only difference is that this aligned map is constant on each component complementary to the L_i°, rather than each component complementary to the L_i. The last stage of the homotopy thus simply requires "reparametrizing the domain" of each φ_i from L_i° to all of L_i, as in the proof of Proposition 4.11. □

Because ι is a homotopy equivalence, we have the following:

Corollary 4.13 *The $\mathcal{C}_1$ action on $\mathrm{AM}_n^{\mathrm{fr}}$ induces a homotopy-commutative multiplication.*

We have not used that the ambient dimension is three, so these results hold for knots in higher-dimensional Euclidean spaces as well. Similar results were proven in [27] for knots modulo immersions (that is, the homotopy fiber of the inclusion of embeddings into immersions) and would work similarly for framed knots, but only for the limit of the $\mathrm{AM}_n^{\mathrm{fr}}$. Turchin [31], and more recently Boavida and Weiss [1], have established versions of this theorem, along with a group structure as in Theorem 5.16 below, for the stages of the tower for knots modulo immersions. Dwyer and Hess have similar results for the limit [8]. We were not able to show that the structure studied in Turchin's paper is compatible with connected sum, which necessitated the present approach.

Remark 4.14 At this point we can explain the connection between Schubert's elementary geometric result that connected sum of knots is commutative [25] and Steenrod's deep, formal work on commutativity of cup product and operations in cohomology [30]. This connection was implied to us by the work of McClure and Smith [19; 20], as applied in this setting by Sinha [27], whose product structures on totalizations of cosimplicial spaces are related to Steenrod's formulae for higher cup products [19].

In different notation from Definition 4.3, the product of two aligned maps f and g is

$$(t_1, \dots, t_n) \mapsto \bigcup_{t_i \le 0 \le t_{i+1}} \hat{f}(t_1, \dots, t_i) * \hat{g}(t_{i+1}, \dots, t_n).$$

Here the union refers to a decomposition of the domain, the n–simplex; $\hat{f}$ and $\hat{g}$ are obtained from f and g by appending times and rescaling (as we did frequently this section); and $*$ indicates a "stacking product" of configurations in $\mathbb{R}^d$.

This is formally similar to the standard formula for cup product

$$\varphi \cup \psi(\sigma\colon [v_1, \dots, v_n] \to X) = \sum \varphi(\sigma|_{[v_1,\dots,v_i]}) \cdot \psi(\sigma|_{[v_{i+1},\dots,v_n]}).$$

Here φ and ψ are cochains, σ is a chain, and brackets $[\,,]$ around some variables refers to the simplex given as convex linear combinations of those variables. This sum is zero except for one term, but, as McClure and Smith show, it is the correct sum to write down for purposes of generalization.

Recall that Theorem 4.9 gives a homotopy from $\mu(f,g)$ to the multiplication defined by Definition 4.8 with $(L_1, L_2) = ([-1,0],[0,1])$. By choosing a path in $\mathcal{C}_1(1)$ from $([-1,0],[0,1])$ to $([0,1],[-1,0])$ (and applying a homotopy from Proposition 4.11) we ultimately get a homotopy from $\mu(f,g)$ to $\mu(g,f)$. We choose the following path, which in the overlapping intervals setting the first interval always lies above the second: start with $([-1,0],[0,1])$; grow the second interval to obtain $([-1,0],[-1,1])$; then translate $[-1,0]$ to $[0,1]$; finally shrink $[-1,1]$ to $[-1,0]$ to obtain the pair $([0,1],[-1,0])$.

If we apply the formulae of Definition 4.8 for the products of f and g governed by this path of 1–disks, we see a formal analogue for Steenrod's formula for cup-one, namely

$$\varphi \cup_1 \psi(\sigma) = \sum_{i<j} \varphi(\sigma|_{[v_1,\dots,v_i,v_j,\dots,v_n]}) \cdot \psi(\sigma|_{[v_i,\dots,v_j]}).$$

The main difference is that the product rather than using an underlying commutative ring uses operad insertion maps. McClure and Smith show this to be an appropriate extension of Steenrod's formula to the operadic setting.

5 Maps and layers in the tower, and abelian group structure

Our goal is to show that each stage of the Goodwillie–Weiss tower for knots has an abelian group structure compatible with connected sum. This also would follow by adapting recent work of Boavida and Weiss [1], which appeared after the present work, where in Theorem 10.2 they exhibit the stages of a closely related tower as the fiber of a two-fold loop map of spaces whose multiplications are compatible with connected sum.

Though we are primarily interested in mapping space models, we use the cosimplicial models and language around them as a starting point and key organizational tool. Cosimplicial structures are also essential for the study of spectral sequences below. We need develop a variety of models for the maps in the totalization tower of a cosimplicial space, which are of independent interest.

5.1 Maps in the tower through projection and restriction

In the cosimplicial realm the totalization tower is a basic object of study, dual in a sense to the skeletal filtration of a simplicial complex. When using the functors $\mathcal{G}_n$ to pull back a subcubical diagram from a cosimplicial space, the maps from $\widetilde{\mathrm{Tot}}^n$ to $\widetilde{\mathrm{Tot}}^{n-1}$ are induced by inclusions $\mathcal{P}_\nu[n-1] \to \mathcal{P}_\nu[n]$.

In the mapping space models, this inclusion of indexing categories explicitly gives rise to the following:

Definition 5.1 The restriction projection $p_n\colon \mathrm{AM}_n^{\mathrm{fr}} \to \mathrm{AM}_{n-1}^{\mathrm{fr}}$ sends a map $\varphi \in \mathrm{AM}_n^{\mathrm{fr}}$ to the composite

$$\Delta^{n-1} \xrightarrow{d} \Delta^n \xrightarrow{\varphi} C_n^{\mathrm{fr}}\langle \mathbb{I}^3, \partial\rangle \xrightarrow{s} C_{n-1}^{\mathrm{fr}}\langle \mathbb{I}, \partial\rangle,$$

where d and s are the images of a (face, degeneracy) pair whose composite is the identity. As any two such choices of a pair (d,s) yield homotopic projections $\mathrm{AM}_n^{\mathrm{fr}} \to \mathrm{AM}_{n-1}^{\mathrm{fr}}$, we take d to be the map $d_n\colon (t_1,\dots,t_{n-1}) \mapsto (t_1,\dots,t_{n-1},1)$ and s to be the map s_n that forgets the last configuration point and framing.

The restriction projection map $\widetilde{p}_n\colon \widetilde{\mathrm{AM}}_n^{\mathrm{fr}} \to \widetilde{\mathrm{AM}}_{n-1}^{\mathrm{fr}}$ and nonframed versions are defined analogously.

Then p_n is our first model for the standard map $\widetilde{\mathrm{Tot}}^n C_\bullet^{\mathrm{fr}}\langle \mathbb{I}^d, \partial\rangle$ to $\widetilde{\mathrm{Tot}}^{n-1} C_\bullet^{\mathrm{fr}}\langle \mathbb{I}^d, \partial\rangle$, and its main use is the following:

Proposition 5.2 *The restriction projection $p_n\colon \mathrm{AM}_n^{\mathrm{fr}} \to \mathrm{AM}_{n-1}^{\mathrm{fr}}$ is a map of $\mathcal{C}_1$–algebras.*

Proof We need to check that the following diagram commutes:

$$\begin{tikzcd}
\mathcal{C}_1(k) \times (\mathrm{AM}_n^{\mathrm{fr}})^k \arrow[r] \arrow[d] & \mathrm{AM}_n^{\mathrm{fr}} \arrow[d] \\
\mathcal{C}_1(k) \times (\mathrm{AM}_{n-1}^{\mathrm{fr}})^k \arrow[r] & \mathrm{AM}_{n-1}^{\mathrm{fr}}
\end{tikzcd}$$

The composite through the upper-right corner is the map

$$(t_1,\dots,t_{n-1}) \mapsto s_n\Bigl(\bigcup_{i=1}^{k} \widehat{L}_i \circ \varphi_i \circ L_i^{-1}(t_1,\dots,t_{n-1},1) \cup \bigcup \{(t_j,0,0)\}\Bigr),$$

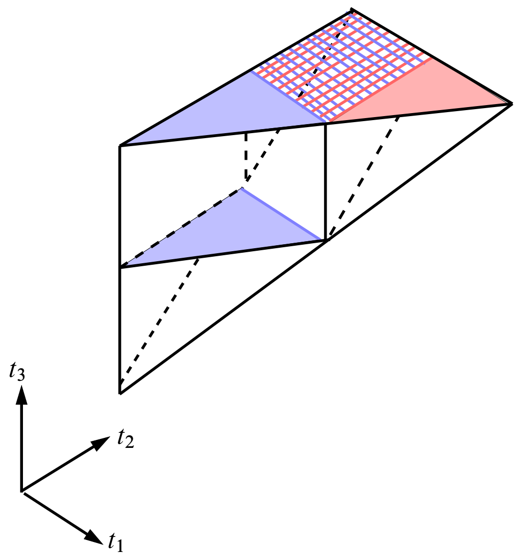

Figure 1: A partial illustration of how the projection $\mathrm{AM}_n \to \mathrm{AM}_{n-1}$ preserves the $\mathcal{C}_1$–action for $n = 3$, $k = 2$, with intervals $(L_-, L_+) = ([-1, 0], [0, 1])$ acting on φ, $\psi \in \mathrm{AM}_n$. The product $(L_-, L_+) \cdot (\varphi, \psi)$ is a map of the whole simplex. The restriction to each light-blue triangular face is (after rescaling in the domain and range, and after forgetting the last configuration point) the projection of φ to AM_2; the blue edge of this triangle is in turn the restriction to the face $t_2 = 1$. Similarly the restriction to the light-red triangular face is the projection of ψ to AM_2; the red edge of this triangle is in turn the restriction to the face $t_1 = 0$. The map on the top face of the whole simplex is the projection of $(L_-, L_+) \cdot (\varphi, \psi)$, which is indeed obtained as a product of the projections of φ and ψ to AM_2.

while the composite through the lower-left corner is the map

$$(t_1, \dots, t_{n-1}) \mapsto \bigcup_{i=1}^{k} \widehat{L}_i \circ s_n \circ (\varphi_i|_{d_n \Delta^{n-1}}) \circ L_i^{-1}(t_1, \dots, t_{n-1}) \cup \bigcup \{(t_j, 0, 0)\}.$$

These maps agree. In either expression, each φ_i is applied to the same configuration, in particular with the same number of t_i equal to 1. As for the resulting configuration in $\mathbb{R}^3$, in the first expression one forgets an extra framed point at $(1, 0, 0)$. In the second expression, one forgets an extra framed point at $(L_i(1), 0, 0)$ for each $i = 1, \dots, k$, yielding the same result. □

A similar argument shows that the projection $\mathrm{AM}_n \to \mathrm{AM}_{n-1}$ is a map of $\mathcal{C}_1$–algebras.

5.2 Maps in the tower through (only) projection

For purposes of connecting our multiplications with those coming from cosimplicial structure, we need another model for the maps and layers in the totalization tower, one which is closely related to p_n in that it uses projection. While the map p_n is defined through including $\mathcal{P}_\nu[n-1]$ in $\mathcal{P}_\nu[n]$, we now proceed by "fibering $\mathcal{P}_\nu[n]$ over $\mathcal{P}_\nu[n-1]$". Consider the functor i_n from $\mathcal{P}_\nu[n]$ to $\mathcal{P}_\nu[n-1]$ which modifies a subset S by identifying $n-1$ and n — that is, changing occurrences of n in S to $n-1$.

Lemma 5.3 *The homotopy limit of $X^\bullet \circ \mathcal{G}_{n-1} \circ i_n$ is homotopy equivalent to $\widetilde{\mathrm{Tot}}^{n-1} X^\bullet$. The n^{th} degeneracy $s^n\colon X^n \to X^{n-1}$ extends to a map*

$$\widehat{s}^n\colon X^\bullet \circ \mathcal{G}_n \to X^\bullet \circ \mathcal{G}_{n-1} \circ i_n.$$

On homotopy limits, the map induced by $\widehat{s}^n$ agrees with the standard map from $\widetilde{\mathrm{Tot}}^n X^\bullet \to \widetilde{\mathrm{Tot}}^{n-1} X^\bullet$.

Proof For the first statement, it suffices to show that the functor i_n is left cofinal [2, Theorem XI.9.2], so that the category $P_\nu[n] \times_{P_\nu[n-1]} (P_\nu[n-1] \downarrow S)$ has contractible nerve for every $S \in P_\nu[n-1]$. But this category has a final object, namely the object corresponding to $S \cup n \in P_\nu[n]$.

It follows from the cosimplicial identities that s^n induces a natural transformation of functors $\widehat{s}^n\colon X^\bullet \circ \mathcal{G}_n \to X^\bullet \circ \mathcal{G}_{n-1} \circ i_n$, where the map on each object is given by either its last degeneracy or the identity map.

Again, the standard map $\widetilde{\mathrm{Tot}}^n X^\bullet \to \widetilde{\mathrm{Tot}}^{n-1} X^\bullet$ is induced by the inclusion functor $j_n\colon \mathcal{P}_\nu[n-1] \to \mathcal{P}_\nu[n]$. Note that the composite $i_n \circ j_n$ is the identity. Consider the following diagram to complete the proof:

$$\begin{tikzcd}
\mathrm{holim}(X^\bullet \circ \mathcal{G}_n) \arrow[d, "- \circ j_n"'] \arrow[dr, "\widehat{s}^n"] & \\
\mathrm{holim}(X^\bullet \circ \mathcal{G}_{n-1}) & \mathrm{holim}(X \circ \mathcal{G}_{n-1} \circ i_n) \arrow[l, "\simeq"', "- \circ j_n"]
\end{tikzcd}$$

The left-hand vertical map is the map $\widetilde{\mathrm{Tot}}^n X^\bullet \to \widetilde{\mathrm{Tot}}^{n-1} X^\bullet$, since $X^\bullet \circ \mathcal{G}_n \circ j_n = X^\bullet \circ \mathcal{G}_{n-1}$. Precomposing $X^\bullet \circ \mathcal{G}_{n-1} \circ i_n$ with j_n induces the horizontal map because $i_n \circ j_n = \mathrm{id}$. This horizontal map is an equivalence because, on these homotopy limits, "precomposing with j_n" is the right-inverse to the equivalence given by "precomposing with i_n". □

By abuse, we let $\widehat{s}^n$ also denote the map it induces on homotopy limits. For purposes of distinction we let $\mathrm{AM}^{\mathrm{fr}}_{n-1}(\Delta^n)$ denote $C^{\mathrm{fr}}_{\bullet}\langle \mathbb{I}^3, \partial\rangle \circ \mathcal{G}_{n-1} \circ i_n$, which by the above is a model for $\mathrm{AM}^{\mathrm{fr}}_{n-1}$. (The notation indicates that we model $\mathrm{AM}^{\mathrm{fr}}_{n-1}$ by maps of Δ^n.) The fact that $\widehat{s}^n\colon \mathrm{AM}^{\mathrm{fr}}_n \to \mathrm{AM}^{\mathrm{fr}}_{n-1}(\Delta^n)$ and p_n both model the structure maps in the totalization tower implies that they are homotopic. We choose $\widehat{s}^n$ for more detailed study of the fibers of these maps, starting with the observation that $\widehat{s}^n$ is a fibration in both the $\widetilde{\mathrm{AM}}^{\mathrm{fr}}_n$ and $\mathrm{AM}^{\mathrm{fr}}_n$ settings, since the projection or identity maps which define it entrywise are fibrations. (See [4, Lemma 3.5] for an explicit proof in this case of a standard result about enriched model structures on diagram categories in general.)

Definition 5.4 Let L_n be the fiber of $\widehat{s}^n\colon \mathrm{AM}^{\mathrm{fr}}_n \to \mathrm{AM}^{\mathrm{fr}}_{n-1}(\Delta^n)$, based at the evaluation map of the unknot. That is, L_n is the space of aligned maps where when one forgets the last point in each configuration in the image one obtains a standard configuration along the x–axis parametrized by the points in the domain simplex. Let $\widetilde{L}_n$ be the fiber of $\widehat{s}^n\colon \widetilde{\mathrm{AM}}^{\mathrm{fr}}_n \to \widetilde{\mathrm{AM}}^{\mathrm{fr}}_{n-1}(\Delta^n)$, which then sits over the constant map at the infinitesimal configuration where all x_{ij} for $i < j$ are equal to $(1,0,0)$.

We will write $L_n(\Delta^n)$ or $\widetilde{L}_n(\Delta^n)$ if we want to emphasize the model we are using for this fiber.

We used both the AM_n and $\widetilde{\mathrm{AM}}_n$ models in the previous section since the former supports an evaluation map and a $\mathcal{C}_1$ structure and the latter has a commutative multiplication. We use both L_n and $\widetilde{L}_n$ in similar fashion here, which means we also need a comparison.

Proposition 5.5 *The map $\iota\colon \mathrm{AM}_n \to \widetilde{\mathrm{AM}}_n$ restricts to a map from L_n to $\widetilde{L}_n$ which is an equivalence and preserves multiplication up to homotopy.*

Proposition 5.6 *The inclusion of L_n into AM_n is a map of $\mathcal{C}_1$–spaces.*

The proofs of all of these are straightforward from the definitions, checking that previous definitions' arguments, such as Proposition 4.12, are compatible with the condition of being a standard configuration but for the last coordinate.

We can use the equivalence $\mathrm{AM}^{\mathrm{fr}}_{n-1}(\Delta^n) \to \mathrm{AM}^{\mathrm{fr}}_{n-1}$ defined by restriction to the last face of Δ^n to define a $\mathcal{C}_1$–structure on $\mathrm{AM}^{\mathrm{fr}}_{n-1}(\Delta^n)$. Combining Propositions 5.2 and 5.6 and composing with the homotopy inverse $\mathrm{AM}^{\mathrm{fr}}_{n-1} \to \mathrm{AM}^{\mathrm{fr}}_{n-1}(\Delta^n)$, which is then also $\mathcal{C}_1$, gives the following:

Corollary 5.7 $L_n \to \mathrm{AM}^{\mathrm{fr}}_n \to \mathrm{AM}^{\mathrm{fr}}_{n-1}(\Delta^n)$ *is a fibration sequence of $\mathcal{C}_1$ spaces.*

5.3 The layers of a totalization tower via cubical diagrams

Following Goodwillie, we use cubical diagrams to develop loop structures on layers in the totalization tower of a cosimplicial space. Some of this material is treated in [22]. We have already seen $\mathcal{P}_\nu[n]$ and $\mathcal{P}_\nu[n-1]$ related by inclusion and "fibering". We next relate them through a Mayer–Vietoris decomposition.

Just as the standard map $\widetilde{\mathrm{Tot}}^n X^\bullet \to \widetilde{\mathrm{Tot}}^{n-1} X^\bullet$ can be defined by an inclusion $\Delta_{n-1} \subset \Delta_n$, it can also be defined by the canonical inclusion $\mathcal{P}_\nu[n-1] \subset \mathcal{P}_\nu[n]$ through the equivalence given by the functor $\mathcal{G}_n$. We will analyze the fibers of these maps — that is, the layers in the totalization tower — in two different ways in this section. First we use a sort of Mayer–Vietoris decomposition of the category $\mathcal{P}_\nu[n]$.

Definition 5.8

- Let $\mathcal{P}_{\neq n} \subset \mathcal{P}_\nu[n]$ be the full subcategory given by all nonempty subsets of $[n]$ except the singleton $\{n\}$.
- Let $\mathcal{P}_{n\in}$ be the (cubical) poset of all subsets of $[n]$ containing n.
- Let $\mathcal{P}_{n+}$ be the (subcubical) poset of subsets of $[n]$ containing n and at least one other element.

The inclusion $\iota\colon \mathcal{P}_\nu[n-1] \hookrightarrow \mathcal{P}_{\neq n}$ is left cofinal, so the map induced on homotopy limits is an equivalence. We can thus replace $\mathrm{holim}_{\mathcal{P}_\nu[n]} X^\bullet \circ \mathcal{G}_n \to \mathrm{holim}_{\mathcal{P}_\nu[n-1]} X^\bullet \circ \mathcal{G}_n$ by an alternate model for the maps in the tower, namely

$$\mathrm{holim}_{\mathcal{P}_\nu[n]} X^\bullet \circ \mathcal{G}_n \to \mathrm{holim}_{\mathcal{P}_{\neq n}} X^\bullet \circ \mathcal{G}_n.$$

The poset $\mathcal{P}_\nu[n]$ can be written as the union of $\mathcal{P}_{\neq n}$ and $\mathcal{P}_{n\in}$ along $\mathcal{P}_{n+}$, yielding the following square:

$$\begin{tikzcd} \mathcal{P}_\nu[n] & \mathcal{P}_{n\in} \arrow[l, hook'] \\ \mathcal{P}_{\neq n} \arrow[u, hook] & \mathcal{P}_{n+} \arrow[l, hook'] \arrow[u, hook] \end{tikzcd} \tag{3}$$

Applying $\mathrm{holim}_{(-)} X^\bullet \circ \mathcal{G}_n$ to the diagram above, we get a pullback square of fibrations [11, Proposition 0.2]. Thus, to study the fiber(s) of the map from $\widetilde{\mathrm{Tot}}^n$ to $\widetilde{\mathrm{Tot}}^{n-1}$, which up to homotopy is the left-hand column of the induced map of homotopy limits of this square, it suffices to study the right-hand column. We say fiber(s) because in our application we study unbased and sometimes disconnected spaces.

Since $\mathcal{P}_{n+}$ is just the cube $\mathcal{P}_{n\in}$ with its initial object removed, the map on homotopy limits induced by the right vertical arrow is just the map from the initial object, at $\{n\}$, to the homotopy limit of the rest of the diagram, which is subcubical. We conclude that the fiber(s) are the total fiber(s) of the cube $\mathcal{P}_{n\in}$ over different possible basepoints in the unbased case. In the based setting the original map is k–connected if and only if the n–cube $X^\bullet \circ \mathcal{G}_n|_{\mathcal{P}_{n\in}}$ is k–Cartesian.

5.4 Loop structure on layers via retracts of cubes

Lemma 5.9 *Let $f\colon C_\bullet \to D_\bullet$ be a map of cubical diagrams and let $r\colon D_\bullet \to C_\bullet$ be a retraction objectwise. Then the total fiber of the cube $C_\bullet \to D_\bullet$ is the loopspace of the total fiber of $D_\bullet \to C_\bullet$.*

Proof Consider the square of cubes, which is itself a cubical diagram:

$$\begin{tikzcd} C_\bullet \arrow[r, "f"] \arrow[d, "\mathrm{id}"'] & D_\bullet \arrow[d, "r"] \\ C_\bullet \arrow[r, "\mathrm{id}"] & C_\bullet \end{tikzcd}$$

We find the total fiber of these cubes in two ways, first internally to the $C_\bullet$ and $D_\bullet$ subcubes followed by taking horizontal then vertical fibers. This yields the total fiber of $C_\bullet \to D_\bullet$. If we first take internal fibers, then vertical, then horizontal, we see loops on the total fiber of $D_\bullet \to C_\bullet$. □

Cosimplicial identities imply that the codegeneracy maps of a cosimplicial space can be used to define retractions of cubes. For example, at the first two levels of a cosimplicial space the codegeneracy map is a retract for either coface map. Since the first layer in the Tot tower is the total fiber of $\mathcal{P}_{1\in}$, which is just a coface map $X^0 \to X^1$, this lemma shows that is loops on the fiber of the codegeneracy $X^1 \to X^0$. More generally we have the following:

Definition 5.10

- For an inclusion of ordered sets $i\colon S \hookrightarrow S'$, we define the dual surjection $i^!\colon S' \to S$ to be the order-preserving retraction which sends each element of S' to the maximal value of S possible among such retractions.
- Let $\mathcal{P}^!_{n\in}$ be the category whose objects are subsets of $[n]$ containing n and where morphisms are all the dual surjections.
- For brevity, let $\mathcal{G}^!_n\colon \mathcal{P}^!_{n\in} \to \mathbf{\Delta}_n$ denote the functor $\mathcal{G}_{\mathcal{P}^!_{n\in}}$ (defined in Definition 3.3).

Proposition 5.11 *For a cosimplicial space $X^\bullet$, the homotopy limit of $X^\bullet \circ \mathcal{G}_n|_{\mathcal{P}_{n\in}}$ and thus the fiber of $\widetilde{\mathrm{Tot}}^n X^\bullet \to \widetilde{\mathrm{Tot}}^{n-1} X^\bullet$ is homotopy equivalent to $\Omega^n \mathrm{holim} X^\bullet \circ \mathcal{G}^!_n$.*

The proof of this proposition in [22, Proposition 9.4.10] is essentially correct, though it makes reference to a diagram which is generally not possible to construct for an arbitrary cosimplicial space and in particular for those we use. We only use a subdiagram of that used in [22] which can always be constructed.

Proof We interpolate between $\mathcal{G}_{\mathcal{P}_{n\in}}$ and $\mathcal{G}_{\mathcal{P}^!_{n\in}}$, which share the underlying objects, namely subsets of $[n]$ which contain n. But $\mathcal{P}_{n\in}$ has basic morphisms (of which all others are composites) of $S \mapsto S \cup i$, while $\mathcal{P}^!_{n\in}$ has basic morphisms with some i, $i+1 \mapsto i+1$ and identity otherwise. For $j = 0, 1, \dots, n$, define $\mathcal{P}_{n\in}(j)$ as having this same set of objects but with generating morphisms $S \subset S \cup i$ for $i \leq j$ and i, $i+1 \mapsto i+1$ and identity otherwise for $i > j$.

We view $X^\bullet \circ \mathcal{G}_{\mathcal{P}_{n\in}(j+1)}$ as a map of cubes $C_\bullet \to D_\bullet$, where $C_\bullet$ is the restriction to subsets which do not contain j and $D_\bullet$ is the restriction to those which do, and the map between cubes is defined by all of the $S \mapsto S \cup j$ maps. Then $X^\bullet \circ \mathcal{G}_{\mathcal{P}_{n\in}(j)}$ is a retract $D_\bullet \to C_\bullet$, with morphisms defined by sending j to the next element in the ordering. We deduce from Lemma 5.9 that the total fiber of $X^\bullet \circ \mathcal{G}_{\mathcal{P}_{n\in}(j+1)}$ is the loopspace of the total fiber of $X^\bullet \circ \mathcal{G}_{\mathcal{P}_{n\in}(j)}$. Thus $\operatorname{fib} X^\bullet \circ \mathcal{G}_{\mathcal{P}_{n\in}(n)} \simeq \Omega^n \operatorname{fib} X^\bullet \circ \mathcal{G}_{\mathcal{P}_{n\in}(0)}$. Since $\mathcal{G}_n|_{\mathcal{P}_{n\in}}$ is $\mathcal{G}_{\mathcal{P}_{n\in}(n)}$ while $\mathcal{G}_{\mathcal{P}^!_{n\in}}$ is $\mathcal{G}_{\mathcal{P}_{n\in}(0)}$, we obtain the result. □

Remark 5.12 It is a tautology that the n^{th} layer in the totalization tower of a fibrant cosimplicial space is an n–fold loopspace. But in our setting we could not use standard fibrant replacement to produce a multiplication on the entries of the tower which is compatible with connected sum. Cubical diagram models for the entries and the layers of the totalization tower give a workable alternative to fibrant replacement, which is broadly applicable.

5.5 Surjectivity on components of maps in the tower

In order to inductively establish a group structure on components of stages in the tower, we need a surjectivity result.

Theorem 5.13 *The restriction-projection map* $\widetilde{p}_n\colon \widetilde{\mathrm{AM}}_n \to \widetilde{\mathrm{AM}}_{n-1}$ *induces a surjection on* π_0.

We prove this unframed version first and then use it to prove the desired framed version, as the end of the proof we give here breaks down in the framed setting.

Proof of Theorem 5.13 We extend techniques from the previous subsection. Recall from Section 3 the cosimplicial space $\widetilde{C}'_\bullet\langle \mathbb{I}^3\rangle$, which we abbreviate here as C. This functor defines both our cosimplicial model and, through pullback by the functors $\mathcal{G}_n$, our operadic mapping space model.

We use the square of posets (3) above. We are led to the map on homotopy limits induced by $\mathcal{P}_{n+} \hookrightarrow \mathcal{P}_{n\in}$, which if surjective on components implies the same for the map $\widetilde{\mathrm{AM}}_n \to \widetilde{\mathrm{AM}}_{n-1}$, induced by the left-vertical map in the square (3).

The initial object in $\mathcal{P}_{n\in}$ is $C(\{n\}) = \widetilde{C}_0'\langle\mathbb{R}^3\rangle$, a point. Thus the map on homotopy limits induced by $\mathcal{P}_{n+} \hookrightarrow \mathcal{P}_{n\in}$ is surjective on components if and only if the homotopy limit over $\mathcal{P}_{n+}$ is connected. As before we reindex this diagram, using the isomorphism of $\mathcal{P}_{n+}$ with $\mathcal{P}_\nu[n-1]$ which sends a set containing n to the set obtained by removing n.

We prove the connectedness of this homotopy limit by induction on n. The case $n=1$ is immediate, as $\widetilde{C}_1'\langle\mathbb{R}^3\rangle = S^2$ is connected. For the induction step, we exhibit the homotopy limit $\mathcal{P}_\nu[n-1]$ as a fibration over a connected space with a connected fiber. Consider the reindexed pushout square (3), with n replaced by $n-1$ everywhere. The inclusion $\mathcal{P}_\nu[n-2] \hookrightarrow \mathcal{P}_{\neq n-1}$ is left cofinal, so the left-hand column of the reindexed (3) gives a fibration whose base, the homotopy limit of D, which we can take over $\mathcal{P}_\nu[n-2]$, is connected by induction.

The square of holims induced by (3) is a pullback, so it suffices to establish connectedness of the fiber of the map induced by the right-hand column, taken over the component to which the connected space $\operatorname{holim}_{\mathcal{P}_{\neq n-1}} D$ maps. Here we choose basepoints for D by choosing the basepoint $(1,0,0)$ in $S^2 = \widetilde{C}_1'\langle\mathbb{I}^3\rangle$.

This induced square of holims is of based spaces, we can describe the fiber of this map as the total fiber of the based cube $D(\mathcal{P}_{n-1})$. Apply Proposition 5.11 to deduce that this total fiber is $\Omega^{n-1}\operatorname{tfib} D^!$.

To show that $\Omega^{n-1}\operatorname{tfib} D^!$ is connected or, equivalently, that $D^!$ is $(n-1)$–Cartesian, we use a Blakers–Massey theorem. This is an $(n-1)$–cube of spaces $\widetilde{C}_i'\langle\mathbb{I}\rangle$ for $i \le n-1$ of configurations in $\mathbb{I}^3$ up to scaling and translation, with a tangent vector at each point. The maps forget points and corresponding tangent vectors. We replace this by a homotopy equivalent cube of spaces $C_i'\langle\mathbb{I}^3\rangle$ or even $C_i'(\mathbb{I}^3)$ of configurations in $\mathbb{I}^3$ with a tangent vector at each point. Every map in this cube is a fibration, so we can take the fiber in one direction. The resulting $(n-2)$–cube, which we call $\varphi D^!$, has entries $\mathbb{I}^3 - f([i])$, where the deleted points are images of a fixed embedding $f\colon [n] \hookrightarrow \mathbb{I}^3$. The maps in the cube are inclusions of open submanifolds and are thus cofibrations. Moreover, $\varphi D^!$ is a pushout cube, so it is strongly co-Cartesian.

Each inclusion $\mathbb{I}^3 - f([i+1]) \hookrightarrow \mathbb{I}^3 - f([i])$ is a 2–connected map. The Blakers–Massey theorem [11, Theorem 2.3] (see also [22, Theorem 6.2.1]) applies to give that the total fiber of $\varphi D^!$ is n–connected. Thus its n^{th} loopspace is connected, which yields the result. □

We will say more about this total fiber — and thus the layers in the tower — both below in this section and in Section 7 when we make spectral sequence calculations.

In the proof just given, the analogous $(n-1)$–cube with frames instead of tangent vectors is not $(n-1)$–Cartesian, which is why we prove the framed version separately now.

Theorem 5.14 *In the framed setting, the map $\widetilde{p}_n\colon \widetilde{\mathrm{AM}}_n^{\mathrm{fr}} \to \widetilde{\mathrm{AM}}_{n-1}^{\mathrm{fr}}$ induces a surjection on π_0.*

Proof As $\widetilde{\mathrm{AM}}_n^{\mathrm{fr}}$ is a subspace of $\mathrm{Map}(\Delta^n, \widetilde{C}_n\langle \mathbb{I}^3\rangle \times O(3)^n)$, we consider the diagram:

$$\begin{tikzcd}
\Delta^n \arrow[r] \arrow[d] & \widetilde{C}_n\langle \mathbb{I}^3\rangle \times O(3)^n \arrow[r] \arrow[d] & \widetilde{C}_n\langle \mathbb{I}^3\rangle \times (S^2)^n \arrow[d] \\
\Delta^{n-1} \arrow[r] & \widetilde{C}_{n-1}\langle \mathbb{I}^3\rangle \times O(3)^{n-1} \arrow[r] & \widetilde{C}_{n-1}\langle \mathbb{I}^3\rangle \times (S^2)^{n-1}
\end{tikzcd}$$

Suppose $\Phi \in \widetilde{\mathrm{AM}}_{n-1}^{\mathrm{fr}}$. Let φ be the image of Φ under $\widetilde{\mathrm{AM}}_{n-1}^{\mathrm{fr}} \to \widetilde{\mathrm{AM}}_{n-1}$, which essentially composes a map to $\widetilde{C}_n\langle \mathbb{I}^3\rangle \times O(3)^n$ with the projection onto $\widetilde{C}_n\langle \mathbb{I}^3\rangle \times (S^2)^n$, using the first vector in each frame. By Theorem 5.13, there is a $\psi \in \widetilde{\mathrm{AM}}_n$ whose image in $\widetilde{\mathrm{AM}}_{n-1}$ is in the same component as φ. We lift ψ to a map $\Delta^n \to \widetilde{C}_n\langle \mathbb{I}^3\rangle \times O(3)^{n-1}$, using Φ to define the map to the $O(3)^{n-1}$ factor.

It remains to lift this map to one additional factor of $O(3)$. Compatibility with the cosimplicial structure maps determine the map $\Delta^n \to \widetilde{C}_n\langle \mathbb{I}^3\rangle \times O(3)^n$ on precisely two faces of Δ^n, where the map to the additional factor of $O(3)$ must agree with the map on another factor. Away from these faces, there are no constraints on the map to the additional factor of $O(3)$. Thus topologically the problem is to extend this map from $D^{n-1} \subset \partial D^n$ to D^n, which is immediate. □

5.6 Group structure

Lemma 5.15 *The multiplication on $\widetilde{L}_n$ obtained by restricting the multiplication μ' of Definition 4.10 is homotopic to the one coming from the description of $\widetilde{L}_n$ as the n–fold loopspace of the total fiber of an n–cube of configuration spaces in Proposition 5.11.*

This compatibility is key in proving the following. Recall the multiplication μ from Definition 4.8.

Theorem 5.16 *$\pi_0(\widetilde{\mathrm{AM}}_n^{\mathrm{fr}})$ is an abelian group with the multiplication μ (or μ') on $\widetilde{\mathrm{AM}}_n^{\mathrm{fr}}$.*

Proof By Theorem 4.9, $\pi_0(\widetilde{\mathrm{AM}}_n^{\mathrm{fr}})$ is an abelian monoid. We prove that this monoid is a group by induction on n.

The case $n=0$ is clear, since $\widetilde{\mathrm{AM}}_0^{\mathrm{fr}} = *$. Suppose $n \geq 1$ and $\widetilde{\mathrm{AM}}_{n-1}^{\mathrm{fr}}$ is a group. From Lemma 5.15, the fiber $\widetilde{L}_n$ is a loopspace when $n \geq 1$, so that its components form a group. Moreover, $\pi_0(\widetilde{L}_n) \to \pi_0(\widetilde{\mathrm{AM}}_n)$ is a map of monoids. Thus applying π_0 to the fiber sequence $\widetilde{L}_n \to \widetilde{\mathrm{AM}}_n^{\mathrm{fr}} \to \widetilde{\mathrm{AM}}_{n-1}^{\mathrm{fr}}$ gives an exact sequence of monoids. By Theorem 5.14, $\pi_0(\widetilde{\mathrm{AM}}_n^{\mathrm{fr}} \to \widetilde{\mathrm{AM}}_{n-1}^{\mathrm{fr}})$ is surjective. The lemma now follows from the elementary fact that, if $G \to H \to K \to 0$ is an exact sequence of monoids with G and K groups, then H is a group. $\square$

Corollary 5.17 *The homotopy fibers of* $\mathrm{AM}_n^{\mathrm{fr}}$ *over varying components of* $\mathrm{AM}_{n-1}^{\mathrm{fr}}$ *are homotopy equivalent.*

Proof of Lemma 5.15 Explicitly, $\widetilde{L}_n(\Delta^n)$ is the subspace of maps $\Delta^n \to \widetilde{C}_n'\langle \mathbb{I}^3\rangle$ in $\widetilde{\mathrm{AM}}_n$ whose projection by forgetting the last point in the configuration yields the chosen constant map. This implies that such maps are themselves constant on the $t_n = t_{n-1}$ and $t_n = t_{n+1} (=1)$ faces of Δ^n. We show that μ' and the multiplication from the n–fold loopspace structure both agree with another multiplication, which we now define.

Consider the map $\Delta^{n-1} \times \mathbb{I} \to \Delta^n$ given by

$$(t_1, \dots, t_{n-1}) \times t \mapsto (t_1, \dots, t_{n-1}, (1-t)t_{n-1} + t).$$

By precomposing by this map, we can view an element of $\widetilde{L}_n(\Delta^n)$ as a loop in $\mathrm{Map}(\Delta^{n-1}, \widetilde{C}_n'\langle \mathbb{I}^3\rangle)$, based at the constant standard map. As we use configuration spaces modulo translations and scalings, these loops begin and end at the exact same map. Loop concatenation then defines a multiplication μ_Ω on $\widetilde{L}_n(\Delta^n)$, which of course has homotopy inverses.

We first prove that μ_Ω is homotopic to the multiplication μ' of Definition 4.10 by "homotoping away the mixed terms of μ'". Consider the function $r\colon \Delta^n \times \mathbb{I} \to \Delta^n$ which linearly interpolates between a configuration of times $\vec{t} = (t_1, \dots, t_n)$ and the "constant configuration" at t_n, but only until all the t_i are contained in $[-1,0]$ or $[0,1]$. Let r_s denote its restriction to $\Delta^n \times s$. This function r is not continuous at points with $t_n = 0$. But because a product $\mu(\varphi, \psi)$ of elements $\varphi, \psi \in \widetilde{L}_n(\Delta^n)$ is constant along $t_n = 0$, the map given by

$$\vec{t} \mapsto \begin{cases} \varphi(r_s(\vec{t})) & \text{if } t_n \leq 0, \\ \psi(r_s(\vec{t})) & \text{if } t_n \geq 0, \end{cases}$$

is ultimately continuous. Moreover, for any s the resulting map is again an element of $\widetilde{\mathrm{AM}}_n$ and in fact $\widetilde{L}_n(\Delta^n)$. More specifically, in the image of the product $\mu'(\varphi,\psi)$ in $\widetilde{C}_n\langle\mathbb{I}^d\rangle \subset \prod_{i<j}(S^{d-1})$, the vectors labeled by pairs i, j with $t_i \leq 0$ and $t_j \geq 0$ are all of the form $(1,0,0)$, and this will be preserved throughout the homotopy.

Thus the "mixed stacking terms" in the product are not essential and up to homotopy the product can be "decoupled" to a map which at each fixed time uses only one of φ and ψ.

The multiplication when $s=1$ agrees with μ_Ω, as the map $\Delta^{n-1}\times I \to \Delta^n$ that we used to define μ_Ω sends $\Delta^{n-1}\times[-1,0]$ to $\{t_n \leq 0\}$ and $\Delta^{n-1}\times[0,1]$ to $\{t_n \geq 0\}$. Thus μ' restricted to $\widetilde{L}_n(\Delta^n)$ is homotopic to μ_Ω, which completes the first half of our proof.

Next, Proposition 5.11 in the case of $\widetilde{\mathrm{AM}}_n$ implies that $\widetilde{L}_n(\Delta^n)$ is homotopy equivalent to $\Omega^n\,\mathrm{tfib}(\widetilde{C}'_\bullet\langle\mathbb{I}^3\rangle\circ\mathcal{G}^!_n)$, where $\widetilde{C}'_\bullet\langle\mathbb{I}^3\rangle\circ\mathcal{G}^!_n$ is a cube of configuration spaces with structure maps which forget points. This is established in the proof of Proposition 5.11 through a series of equivalences

$$\mathrm{tfib}(\widetilde{C}'_\bullet\langle\mathbb{I}^3\rangle\circ\mathcal{G}_n) \simeq \Omega\,\mathrm{tfib}(\widetilde{C}'_\bullet\langle\mathbb{I}^3\rangle\circ\mathcal{G}_{\mathcal{P}_{n\in}(n-1)}) \simeq\cdots\simeq \Omega^i\,\mathrm{tfib}(\widetilde{C}'_\bullet\langle\mathbb{I}^3\rangle\circ\mathcal{G}_{\mathcal{P}_{n\in}(n-i)}).$$

The multiplication from the single loopspace structure on $\Omega\,\mathrm{tfib}(\widetilde{C}'_\bullet\langle\mathbb{I}^3\rangle\circ\mathcal{G}_{\mathcal{P}_{n\in}(n-1)})$ agrees up to homotopy with the others coming from i–fold loopspace structure, including $i=n$. (That is, all the equivalences above, except the first one, are loop maps.) We will complete the proof by showing that this single loopspace multiplication coincides with μ_Ω.

By definition $\mathrm{tfib}(\widetilde{C}'_\bullet\langle\mathbb{I}^3\rangle\circ\mathcal{G}_{\mathcal{P}_{n\in}(n-1)})$ is the total fiber of an $(n-1)$–cube of fibers, where each fibration forgets the last point. An element of loops on this total fiber is a loop of maps $\alpha\colon I^{n-1}\to \mathrm{fib}(\widetilde{C}'_n\langle\mathbb{I}^3\rangle\to\widetilde{C}'_{n-1}\langle\mathbb{I}^3\rangle)$, where

- each map α is constant on the face $\{t_i=1\}$ for every i, and
- on each of the remaining $n-1$ faces, the image of α is in the image of the appropriate doubling map.

On the other hand, an element of $\widetilde{L}_n(\Delta^n)$ is a loop of maps $\alpha\colon \Delta^{n-1}\to\widetilde{C}'_n\langle\mathbb{I}^3\rangle$, where

- the image of each α in $\widetilde{C}'_{n-1}\langle\mathbb{I}^3\rangle$ is the standard constant map,
- α is constant on the face $\{t_{n-1}=1\}$, and
- on each of the remaining $n-1$ faces, the image of α is in the image of the appropriate doubling map.

Consider a homeomorphism from I^{n-1} to Δ^{n-1} which identifies $\bigcup_{i=1}^{n}\{t_i = 1\} \subset I^{n-1}$ with $\{t_{n-1} = 1\} \subset \Delta^{n-1}$. Via this homeomorphism, we get a homeomorphism between loops on this total fiber and $\widetilde{L}_n(\Delta^n)$ which is compatible with the multiplications on each. $\square$

5.7 Summary

Putting results together the results, the sequences $L_n \to \mathrm{AM}_n^{\mathrm{fr}} \to \mathrm{AM}_{n-1}^{\mathrm{fr}}$ in the Goodwillie–Weiss tower approximating classical framed knots satisfy the following:

(1) They are fibration sequences of $\mathcal{C}_1$–spaces (Definition 4.3 and Corollary 5.7).

(2) $\mathrm{AM}_n^{\mathrm{fr}}$ receives a multiplication-preserving (in fact, $\mathcal{C}_1$–action-preserving) evaluation map from the knot space, with connected sum as its multiplication (Proposition 4.4).

(3) At the level of components, all multiplications are commutative (Corollary 4.13) and have inverses (Theorem 5.16 and Proposition 4.12). Moreover, $\mathrm{AM}_n^{\mathrm{fr}} \to \mathrm{AM}_{n-1}^{\mathrm{fr}}$ is surjective on components (Theorem 5.14).

6 The homotopy tower is a finite-type invariant

In this section we show that $\pi_0(\mathrm{ev}_n)\colon \pi_0 \operatorname{Emb}^{\mathrm{fr}}(\mathbb{R}, \mathbb{R}^3) \to \pi_0 \mathrm{AM}_n^{\mathrm{fr}}$, which we've shown to be an abelian group-valued invariant compatible with connected sum of knots, is a finite-type invariant of type $n-1$. The main tool is a theorem of Habiro [15], which states that two classical knots share finite-type invariants of degree $\leq n-1$ if and only if they differ by a series of C_n–moves.

We describe these moves in a way that will facilitate our proof. Let E_2 be a copy of $D^2 \times I = D^2 \times [-1, 1]$, with two properly embedded subarcs which clasp in the center as in Figure 2. Iteratively form E_n from E_{n-1} by replacing a regular neighborhood of the top left arc of E_n by a copy of E_2. Thus E_n is $D^2 \times I$ with n properly embedded arcs. (So technically each E_n is a pair of spaces.)

Now a basic C_n–move on a knot K is given by finding an embedding e of E_n into $\mathbb{R}^3$ which meets the knot K as the given collection of arcs, and sliding another subarc of K across the central disk $e(D^2 \times \{0\})$ of the embedded E_n as in Figure 3. In this figure, the almost vertical strand in each left (resp. right) picture is a subarc of the knot which is isotopic to the front (resp. back) half of the boundary of this central disk.

The basic tool we need is the following theorem, which follows directly from work of Habiro [15]:

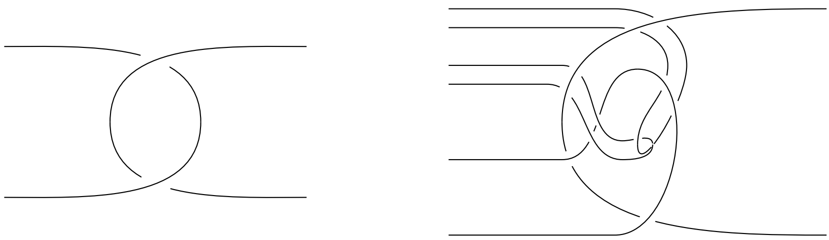

Figure 2: The basic clasp E_2 and the iteratively constructed E_4. The ambient cylinders are not drawn.

Theorem 6.1 *Suppose that ν is an additive invariant of (unframed) knots. It is invariant under C_n–moves if and only if it is a finite-type invariant of degree $n-1$.*

Proof The C_n moves constructed here are an alternate formulation of clasper surgery, and in fact are very close to Habiro's original formulation in his master's thesis. The fact that C_n–moves give the same equivalence relation as arbitrary capped clasper surgery is well known, but we will sketch the argument here. (See eg [21] at the top of page 124.) The first step is to prove that C_n–moves correspond to "linear claspers", where all nodes are directly adjacent to a leaf, while the second step is to show that any capped tree clasper surgery may be represented by surgery on a sequence of linear claspers.

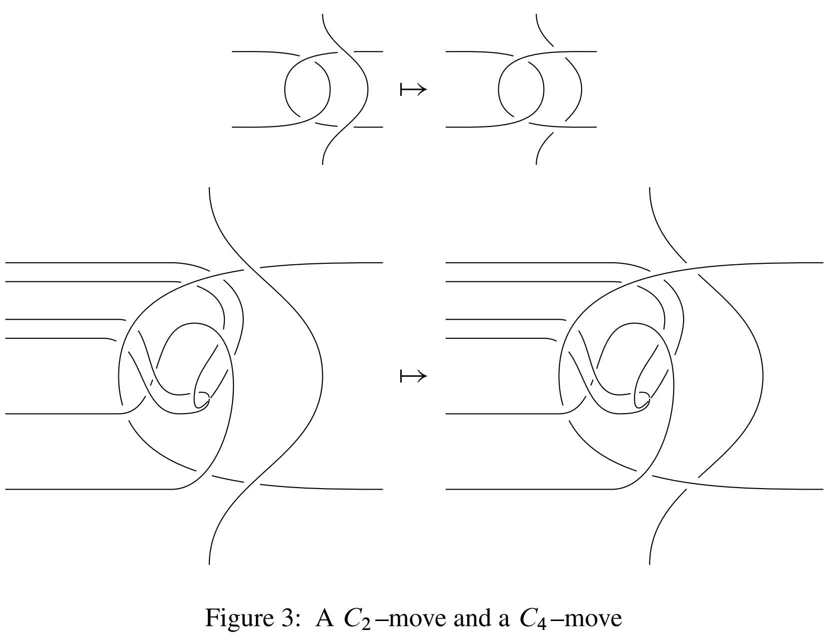

Figure 3: A C_2–move and a C_4–move

Step one can be explicitly proven relatively easily by using the duality between claspers and capped grope cobordism of [7]. One explicitly builds a capped grope cobordism between the before and after knots in a C_n–move. Moreover every capped grope cobordism corresponding to a linear tree can be isotoped to a standard picture of a capped grope embedded in 3–space with a band attached that pierces each cap once. This standard picture is visibly represents the C_n–move.

Step two follows because the topological IHX relation allows us to reduce to the case of linear claspers. (See [7, Corollary 31(a)] or [6, Theorem 7].) The point is that surgery on a capped tree clasper can be replaced by surgery on a sequence of two claspers, each corresponding to the other two trees in an IHX relation. Any tree type can be reduced to linear trees by judicious applications of the IHX relation, as illustrated in the proof of [7, Corollary 31(a)]. So the C_n–moves introduced here are equivalent to the C_n–moves in [15].

The monoid of knots modulo the equivalence relation of n–equivalence is a finitely generated abelian group [14; 15]. Theorem 6.17 of [15] states that the natural projection ψ_{n-1} from knots to this abelian group is a universal additive finite-type invariant of degree $n-1$. So if ν is an additive invariant of knots which also is invariant under C_n moves, it induces a homomorphism on the group of knots modulo n–equivalence. It thus factors as a composition of a degree $n-1$ invariant with a group homomorphism. It is therefore a degree $n-1$ invariant itself. □

In order to move to the framed case, we need the following lemma. For any integer k, let fr_k be the map from unframed knots to framed knots which adds a k–framing.

Lemma 6.2 *Let U_1 represent the $+1$–framed unknot. Let ν be an additive framed knot invariant taking values in an abelian group. Then*

$$\nu \circ \mathrm{fr}_k = k\nu(U_1) + \nu \circ \mathrm{fr}_0 .$$

Proof One can push the twisting of the framing onto a standard subarc of the knot to see that $\mathrm{fr}_k(K) = U_k \# \mathrm{fr}_0(K)$, where U_k is a k–framed unknot. Then one separates each of the twists and uses the fact that ν is additive. □

Corollary 6.3 *Suppose that ν is an additive invariant of framed knots. If it is invariant under C_n–moves then it is a finite-type invariant of degree $n-1$ for $n \geq 2$.*

Proof Note that $\nu \circ \mathrm{fr}_0$ is an additive invariant of unframed knots, and that it is invariant under C_n moves, since ν is. Therefore $\nu \circ \mathrm{fr}_0$ is finite-type of type $n-1$ by Theorem 6.1. On the other hand, by Lemma 6.2, $\nu(K) = \mathrm{fr}(K)\nu(U_1) + \nu \circ \mathrm{fr}_0(K)$, where $\mathrm{fr}(K)$ is the framing number. The invariant fr is known to be of type 1, so we have a linear combination of a type-$(n-1)$ and type-1 invariant, which is therefore of type $n-1$. □

This corollary gives us the main tool we need to show that $\pi_0(\mathrm{ev}_n)$ is of finite type. We also use deformations of evaluation maps of knots.

Definition 6.4 A *σ–deformation* of a knot K is a map obtained by precomposing the adjoint of the evaluation map $\Delta^n \times \mathrm{Emb}^{\mathrm{fr}}(\mathbb{R}, \mathbb{R}^3) \to C_n\langle \mathbb{I}^3, \partial\rangle \times O(3)^n$ with a section $\sigma\colon \Delta^n \to \Delta^n \times \mathrm{Emb}^{\mathrm{fr}}(\mathbb{R}, \mathbb{R}^3)$ of this trivial bundle which maps into the component of K in the embedding space.

Theorem 6.5 *The map $\pi_0(\mathrm{ev}_n)$ is invariant under C_n–moves, and therefore is a type-$(n-1)$ invariant.*

Proof Let K be a framed knot, and let K' be the knot after the C_n–move has been applied. Our strategy is to find σ–deformations of K and K' which we can then show are homotopic. These σ–deformations will have the property that none of the configuration points in $\mathrm{ev}_n(\sigma(\vec{t}))$ meet the central disk whenever strictly fewer than n of those configuration points in $\mathrm{ev}_n(\sigma(\vec{t}))$ lie on the subarcs of E_n. This implies that the σ–deformations are homotopic, since we can just push the arc across the central disk of E_n without ever introducing collisions of configuration points. This is by design for points in Δ^n for which $n-1$ points or fewer are inside E_n. If there are n points in E_n, that means there is no point left over to lie in the exterior arc that we are homotoping.

Both of our σ–deformations will isotop only the copies of E_n in K and K'. Consider a nested copy of E_i inside E_n. It consists of an arc α_i clasping with a copy of E_{i-1}. Let D_i be the (pair of) space(s) given by the intersection of E_i with a small neighborhood $D^2 \times (-\delta, \delta)$ of the central disk $D^2 \times \{0\}$, where δ is large enough so that there is a homeomorphism of pairs $E_i \cong D_i$. The isotopies of E_i will slide D_i along E_i, ie they will take D_i into $D^2 \times (a, b)$ for some $(a, b) \subset [-1, 1]$. Roughly, (a, b) will be a small interval near either -1 or 1 according as which of these endpoints is closer to the center of mass of the configuration points. More specifically, a set of configuration points in the arc α_i will pull D_i toward that side of $D^2 \times [-1, 1]$ in a manner that increases as the minimum distance of these points to the midpoint of the arc α_i decreases. Configuration points in the copy of E_{i-1} exert a similar tug to their end of E_i in a manner which increases as they get closer to the midpoints of their arcs. However the tug of a point is halved in magnitude when you pass to E_{i-1}. This has

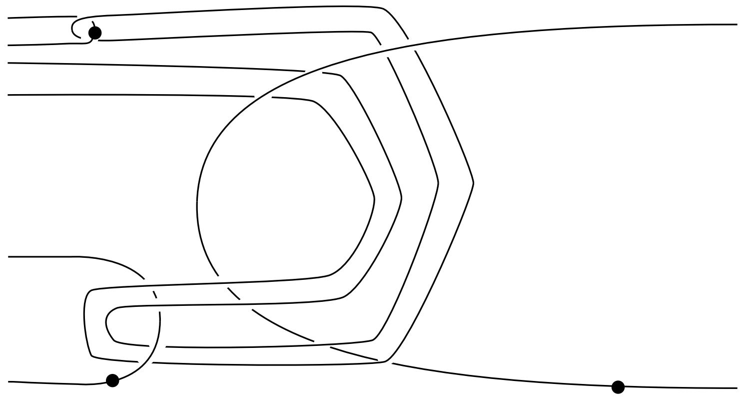

Figure 4: An example of the reembedding of E_4 resulting from a point $(t_1,\dots,t_4)\in C_4\langle\mathbb{I},\partial\rangle$, whose image under ev_4 is also shown. (The image of one of the t_i is not in E_4.) The two left-hand configuration points drag their respective arcs toward their ends of their cylinders. The right-hand configuration point is partially balancing their tug on the central clasp. If the right-hand point moved toward the middle of its arc, the central clasp of E_4 would get tugged close to the right-hand end of the cylinder, although the other configuration points would still be close to the left-hand end. A point added to the arc without a point on it would bring E_4 closer to the original embedding as it got closer to the midpoint of the arc.

been set up so that a point at the midpoint of α_i will always exert a tug that equals or exceeds the collective tug of E_{i-1}.

We also set things up so that if E_{i-1} has fewer than $i-1$ configuration points on it then configuration points in α_i can never get more than ε away from their end disk $D^2\times 0$. The point is that any configuration points in α_i will have a greater tug on the clasp than E_{i-1} when one is at the midpoint of α_i. We just ensure that this tug is strong enough to pull α_i ε–close to its end disk. (We can come up with a uniform ε in this way since there is a discrete gap between the maximal tug that E_{i-1} can exert and the tug it exerts at less than full occupancy.)

Similarly, arrange that if the arc α_i has no configuration points then no configuration points in E_{i-1} can get more than ε away from $D^2\times 1$ inside E_i.

Ultimately, we have an isotopy of E_n parametrized by $C_n\langle\mathbb{I},\partial\rangle$ or, equivalently, a reembedding of E_n for each $(t_1,\dots,t_n)\in C_n\langle\mathbb{I},\partial\rangle$. See Figure 4 for an example.

With such a family of reembeddings, we claim that no configuration point ever passes through the $D^2\times\frac{1}{2}$ disk of E_n, provided that strictly fewer than n configuration points are present inside E_n. Let the arcs in E_n be called α_1, $\alpha_2, \ldots$ arranged in order of decreasing depth.

We know that there is some arc α_i which is not occupied by a configuration point. The points in E_{i-1} stay within ε of the end disk and cannot cross the center disk of E_n. Thus any points that do cross the center must lie on one of the arcs $\alpha_{i+1}, \ldots, \alpha_n$. Now consider such an arc α_j with $j > i$. It links with E_{j-1} which has less than full occupancy. Points in α_j cannot get further than ε away from their end disk, and cannot cross the center disk of E_n, so we are done! □

7 The homotopy spectral sequence for the tower and finite-type knot theory

In this section we further develop the spectral sequence for the homotopy groups and in particular the components of the stages in the Goodwillie–Weiss tower for classical knots and its relationship — both established and conjectural — with finite-type knot theory. Such analysis for knots in higher-dimensional Euclidean space, which are connected, has been covered elsewhere, starting in [28]. We see that at the E^2 stage the entries of the spectral sequence are exactly what one would expect if the tower is to serve as a universal additive finite-type invariant. We in particular see similar structures to what Goodwillie and Weiss [13, Section 5] originally saw in higher dimensions, but can also compare that to newer results on the combinatorics of finite-type invariants [5].

A priori the spectral sequence of a totalization tower, or any other tower of fibrations, is difficult to discern in degree zero. Not only is π_0 only a set-valued functor, but homotopy groups can differ over different components. We saw in Section 5 however that this tower has additional algebraic structure, which leads to the following:

Theorem 7.1 *The spectral sequence for the homotopy groups, and in particular components, of* $\mathrm{AM}_n^{\mathrm{fr}}$ *as a stage in the Goodwillie–Weiss tower is a spectral sequence of abelian groups which converges.*

Proof By Corollary 5.7 and Theorem 5.16 the fiber sequence

$$L_n \to \mathrm{AM}_n^{\mathrm{fr}} \to \mathrm{AM}_{n-1}^{\mathrm{fr}} \tag{4}$$

is a fibration sequence of group-like $\mathcal{C}_1$–spaces. It is thus loops on the fibration sequence defined on their classifying spaces. Its long exact sequence in homotopy groups is then a degree shift of that for the classifying spaces, starting with π_1 of the classifying spaces being π_0 of these spaces. These exact sequences can be spliced in the usual way to obtain a spectral sequence, which by Theorem 5.16 is one of abelian groups.

Convergence follows from the fact that this tower is finite when we truncate it at $\mathrm{AM}_n^{\mathrm{fr}}$, and the groups are finitely generated by Proposition 7.2 below along with the fact that homotopy groups of spheres are finitely generated. □

We can now analyze the spectral sequence in further detail. By Proposition 5.11, L_n is equivalent to Ω^n of the total fiber of the n–cube $C_\bullet^{\mathrm{fr}}\langle \mathbb{I}^3, \partial\rangle \circ \mathcal{G}_n^!$. Unraveling definitions, this is an n–cube whose entries are spaces of configurations of at most n points, together with at most n frames, where each map in the cube forgets a point and a frame. We can express this cube as an entrywise product of the cubes $S \mapsto C_{\underline{n}\setminus S}\langle \mathbb{I}^3, \partial\rangle$ and $S \mapsto O(3)^{\underline{n}\setminus S}$.

The total fiber of the product is the product of the total fibers, and for $n \geq 2$ the total fiber of the cube of powers of $O(3)$ is a point. Thus it suffices to consider the total fiber of $S \mapsto C_{\underline{n}\setminus S}\langle \mathbb{I}^3, \partial\rangle$. Furthermore, we can switch to open configuration spaces, for which these forgetting maps are well known to be fibrations. We take fibers in one direction and consider the resulting $(n-1)$–cube instead. Here we see the entries as $\mathbb{I}^3$ with a finite set of points removed, and maps which are inclusions. Up to homotopy, this is a cube of spaces $\bigvee_T S^2$ indexed by subsets $T \subset \underline{n-1}$, where each map projects off a wedge factor. Call this cube $\mathcal{P}(\underline{n-1})\big(\bigvee S^2\big)$.

By Hilton's theorem [16], the homotopy groups of a wedge $\bigvee_n S^2$ is a direct sum of homotopy groups of higher-dimensional spheres. To elaborate, let $\mathcal{L}_n$ be the free graded Lie algebra (working over the integers for the rest of this section) on n odd-graded generators in degree one. Let $\mathcal{B}_{n,\mathrm{full}}$ be a basis for $\mathcal{L}_n$, choosing these consistently as n varies. For example, we could use a graded version of Hall bases.

Hilton's theorem states that $\pi_*\big(\bigvee_n S^2\big)$ is a direct sum $\bigoplus_{W\in\mathcal{B}_{n,\mathrm{full}}} \pi_* S^{|W|-1}$, where W is the degree or word length of W. The theorem is functorial if we use bases for free Lie algebras of different ranks which extend one another, since the Whitehead products used to define the elements of homotopy are functorial. Because the projection maps between wedge products of S^2 are split, these different bases split off. An immediate inductive calculation of the homotopy groups of the total fiber (as an iterated fiber of fibers) shows that what is left for homotopy groups is indexed by a basis $\mathcal{B}_n$ of the submodule of the $\mathcal{L}_n$ spanned by brackets in which all generators occur.

Proposition 7.2 *The spectral sequence for the homotopy groups (including π_0) of AM_n has as $E^1_{-p,*}$ the module $\bigoplus_{W\in\mathcal{B}_{p-1}} \pi_* S^{|W|-1}$.*

By Lemma 5.15, the abelian group structure in the spectral sequence agrees with the usual abelian group structure on the homotopy groups of spheres.

We now focus on total degree zero. Let $\mathcal{A}_n^I$ be the $\mathbb{Z}$–module of chord diagrams on a line segment with n chords, modulo the usual four-term relation 4T and the relation SEP, which sets every separated (ie nonprimitive) diagram to zero. Alternatively, $\mathcal{A}_n^I$ is the $\mathbb{Z}$–module of trivalent diagrams, modulo antisymmetry, the IHX relation and SEP. See [5] for more details.

Theorem 7.3 *The group $E^2_{-n,n}$ is isomorphic to $\mathcal{A}^I_{n-1}$.*

Proof By Proposition 7.2 the group $E^1_{-n,n}$ is isomorphic to the submodule of the free Lie algebra on $n-1$ generators generated by $(n-1)$–fold brackets where each generator appears exactly once. This module is $\mathcal{L}ie(n-1)$, the $n-1^{\mathrm{st}}$ space in the Lie operad, which well known to be free of rank $(n-1)!$.

Next, we consider the 1-line of the E^1 page. Under the identification of Proposition 7.2 these groups decompose into a free summand and two-torsion. The free summands are indexed by n–fold brackets in the free Lie algebra on $n-1$ generators, again in which all generators occur. The two-torsion summands occur as composites of $S^n \xrightarrow{\eta} S^{n-1}$ with $(n-1)$–fold Whitehead products from $S^{n-1} \to \mathcal{P}(\underline{n-1})(\bigvee S^2)$. This summand is thus isomorphic to $\mathcal{L}ie(n-1) \otimes \mathbb{Z}/2$.

The differential d_1 must be zero on the torsion summand. We claim that on the free summand the differential is the integral version of the differential defined in [24]. There, in Theorem 2.1, through the tower of fibrations

$$C_n(\mathbb{R}^d) \to C_{n-1}(\mathbb{R}^d) \to \cdots \to C_0(\mathbb{R}^d) \tag{5}$$

the well-known rational homotopy Lie algebra of the configuration space $C_n(\mathbb{R}^d)$ is calculated as generated by classes b_{ij} in degree $d-1$. Under the map from the total fiber of $\mathcal{P}(\underline{n-1})(\bigvee S^2)$ to $C_n(\mathbb{R}^3)$ the basis for the free Lie algebra on generators, say x_i, sends a bracket to a corresponding brackets in the b_{in}. Because the projections in the tower (5) are split, these brackets are integral generators of free summands as well. (In fact, one can use the splitting of the tower (5) and the Hilton–Milnor theorem to express homotopy groups of configuration spaces as a direct sum of homotopy groups of spheres.) The formulas for the differential given in [24] are given in terms of these integral generators, so they hold for the spectral sequence over the integers as well.

In [5], the cokernel of the rational d^1 is computed to be $\mathcal{A}^I_{n-1} \otimes \mathbb{Q}$. While the result is stated rationally (which is where the conjecture was made), all of the calculations involve only integer coefficients. □

At the E^2–level the components of AM_n thus look like they should receive a universal additive finite type-$(n-1)$ invariant over the integers, which was established for $n=3$ as the main result of [4].

We elaborate our conjecture as follows. Because the map ev_n from the knot space to this Goodwillie–Weiss tower, sometimes called the "homotopy tower", factors the map to the variant that Volić considers in [32], sometimes called the "rational homology tower", we already knew that it encodes every rational finite type-$\frac{n}{2}$ invariant. By Theorem 6.5, we now know that any invariant which factors through $\pi_0(\mathrm{ev}_n)$ is of type at most $n-1$. Standard finite-type theory then gives a map from $\mathcal{A}^I_{n-1}$ to $\pi_0(\mathrm{AM}_n)$, as follows. For a diagram with $n-1$ chords, choose a knot K with $n-1$ double points prescribed by the chords; by the well known Vassiliev skein relation, K can be rewritten as an alternating sum of 2^{n-1} knots where all the singularities are resolved. This alternating sum can then be interpreted inside the abelian group $\pi_0(\mathrm{AM}_n)$, which defines the map. Although the choice of K is not determined up to isotopy, any two choices will have the same type-$(n-1)$ invariants, and will yield the same image under $\pi_0(\mathrm{ev}_n)$. We conjecture that this is an isomorphism at E^2, which collapses to E^∞. This would imply by Theorem 7.3 that all weight systems lift to finite-type invariants over the integers. That is, it would establish $\pi_0(\mathrm{ev}_\infty)$ as a refinement of the Kontsevich integral, defined over the integers.

In the framed setting we have some additional low-dimensional calculations. Namely, $\mathrm{AM}_1^{\mathrm{fr}} \simeq \Omega\mathrm{SO}(3)$, implying its components are isomorphic to $\mathbb{Z}/2$. The evaluation map calculates the parity of the framing, which is indeed the only additive type-1 invariant. Next, Theorem 3.6 of [4] states that AM_2 is contractible. The subcubical diagram which defines $\mathrm{AM}_2^{\mathrm{fr}}$ fibers over that which defines AM_2, with the maps in this fibering built from the standard fibration of SO(3) over S^2 as the unit tangent bundle. The fiber is a subcubical model for $\Omega S^1 \simeq \mathbb{Z}$ (pulled back from the cosimplicial model for ΩS^1 via $\mathcal{G}_2$), and the map from the framed knot space to $\mathrm{AM}_2^{\mathrm{fr}}$ classifies framing number.

More generally, the subcubical diagram which defines $\mathrm{AM}_n^{\mathrm{fr}}$ fibers over that which defines AM_n with fiber given by a n–subcubical diagram which models ΩS^1. Because $\pi_0(\mathrm{AM}_n^{\mathrm{fr}})$ projects surjectively onto $\pi_0(\mathrm{AM}_2^{\mathrm{fr}})$, compatibly with the identification of this fiber with ΩS^1, on components this yields a splitting $\pi_0(\mathrm{AM}_n^{\mathrm{fr}}) \cong \pi_0(\mathrm{AM}_n) \times \mathbb{Z}$. (Note that this would follow immediately from elementary results in Goodwillie–Weiss calculus if the splitting of spaces of framed knots can be made functorial.) Thus these Goodwillie–Weiss models reflect the fact that the usual and framed finite-type theories differ only by the framing number invariant.

One key step towards establishing this conjecture would be the collapse of the spectral sequence, which is now of a tower of fibrations amenable to tools from algebraic topology. This is in contrast to Vassiliev's approach, where the limiting process of unstable spectral sequences is not well understood. (See [10] for this limiting process in a piecewise-linear setting.) The Goodwillie–Weiss tower is built from maps of spaces, in particular the sequences of (4) and the diagrams which define AM_n, which have been analyzed to great effect for knots in higher dimensions [17].

Another key intermediate step would be the surjectivity on components of ev_n. This statement would follow from deep connectivity results of Goodwillie and Klein [12] if those applied in codimension two, but it may be approachable more directly in this case.

We suspect, however, that direct analysis of the invariants which arise from $\pi_0(\mathrm{ev}_n)$ will be most fruitful. They have already led to new geometric insight in degrees three and four [4; 9]. Sinha and Walter's Hopf invariants [29] can fully be applied by the calculations of Proposition 7.2. They seem to lead to Goussarov–Polyak–Viro formulae, which is a promising sign.

Acknowledgments The authors thank the referee for a careful reading of the paper and useful comments. Budney acknowledges support from NSERC and PIMS. Koytcheff was supported partially by NSF grant DMS-1004610 and partially by a PIMS post-doctoral fellowship. He thanks Tom Goodwillie for useful conversations.

References

[1] **P Boavida de Brito**, **M S Weiss**, *Spaces of smooth embeddings and configuration categories*, preprint (2015) arXiv

[2] **A K Bousfield**, **D M Kan**, *Homotopy limits, completions and localizations*, Lecture Notes in Mathematics 304, Springer, Berlin (1972) MR

[3] **R Budney**, *Little cubes and long knots*, Topology 46 (2007) 1–27 MR

[4] **R Budney**, **J Conant**, **K P Scannell**, **D Sinha**, *New perspectives on self-linking*, Adv. Math. 191 (2005) 78–113 MR

[5] **J Conant**, *Homotopy approximations to the space of knots, Feynman diagrams, and a conjecture of Scannell and Sinha*, Amer. J. Math. 130 (2008) 341–357 MR

[6] **J Conant**, **R Schneiderman**, **P Teichner**, *Jacobi identities in low-dimensional topology*, Compos. Math. 143 (2007) 780–810 MR

[7] **J Conant**, **P Teichner**, *Grope cobordism of classical knots*, Topology 43 (2004) 119–156 MR

[8] **W Dwyer**, **K Hess**, *Long knots and maps between operads*, Geom. Topol. 16 (2012) 919–955 MR

[9] **G Flowers**, *Satanic and Thelemic circles on knots*, J. Knot Theory Ramifications 22 (2013) art. id. 1350017 MR

[10] **C D Giusti**, *Plumbers' knots and unstable Vassiliev theory*, PhD thesis, University of Oregon (2010) MR Available at `http://search.proquest.com/docview/749774372`

[11] **T G Goodwillie**, *Calculus, II: Analytic functors*, K–Theory 5 (1992) 295–332 MR

[12] **T G Goodwillie**, **J R Klein**, *Multiple disjunction for spaces of Poincaré embeddings*, J. Topol. 1 (2008) 761–803 MR

[13] **T G Goodwillie**, **M Weiss**, *Embeddings from the point of view of immersion theory, II*, Geom. Topol. 3 (1999) 103–118 MR

[14] **M N Goussarov**, *Interdependent modifications of links and invariants of finite degree*, Topology 37 (1998) 595–602 MR

[15] **K Habiro**, *Claspers and finite type invariants of links*, Geom. Topol. 4 (2000) 1–83 MR

[16] **P J Hilton**, *On the homotopy groups of the union of spheres*, J. London Math. Soc. 30 (1955) 154–172 MR

[17] **P Lambrechts**, **V Turchin**, **I Volić**, *The rational homology of spaces of long knots in codimension* > 2, Geom. Topol. 14 (2010) 2151–2187 MR

[18] **J E McClure**, **J H Smith**, *A solution of Deligne's Hochschild cohomology conjecture*, from "Recent progress in homotopy theory" (D M Davis, J Morava, G Nishida, W S Wilson, N Yagita, editors), Contemp. Math. 293, Amer. Math. Soc., Providence, RI (2002) 153–193 MR

[19] **J E McClure**, **J H Smith**, *Multivariable cochain operations and little n–cubes*, J. Amer. Math. Soc. 16 (2003) 681–704 MR

[20] **J E McClure**, **J H Smith**, *Cosimplicial objects and little n–cubes, I*, Amer. J. Math. 126 (2004) 1109–1153 MR

[21] **J-B Meilhan**, **A Yasuhara**, *On C_n–moves for links*, Pacific J. Math. 238 (2008) 119–143 MR

[22] **B A Munson**, **I Volić**, *Cubical homotopy theory*, New Mathematical Monographs 25, Cambridge University Press (2015) MR

[23] **P Salvatore**, *Knots, operads, and double loop spaces*, Int. Math. Res. Not. 2006 (2006) art. id. 13628 MR

[24] **K P Scannell**, **D P Sinha**, *A one-dimensional embedding complex*, J. Pure Appl. Algebra 170 (2002) 93–107 MR

[25] **H Schubert**, *Die eindeutige Zerlegbarkeit eines Knotens in Primknoten*, S.-B. Heidelberger Akad. Wiss. Math.-Nat. Kl. 1949 (1949) 57–104 MR

[26] **D P Sinha**, *Manifold-theoretic compactifications of configuration spaces*, Selecta Math. 10 (2004) 391–428 MR

[27] **D P Sinha**, *Operads and knot spaces*, J. Amer. Math. Soc. 19 (2006) 461–486 MR

[28] **D P Sinha**, *The topology of spaces of knots: cosimplicial models*, Amer. J. Math. 131 (2009) 945–980 MR

[29] **D Sinha**, **B Walter**, *Lie coalgebras and rational homotopy theory, II: Hopf invariants*, Trans. Amer. Math. Soc. 365 (2013) 861–883 MR

[30] **N E Steenrod**, *Products of cocycles and extensions of mappings*, Ann. of Math. 48 (1947) 290–320 MR

[31] **V Turchin**, *Delooping totalization of a multiplicative operad*, J. Homotopy Relat. Struct. 9 (2014) 349–418 MR

[32] **I Volić**, *Finite type knot invariants and the calculus of functors*, Compos. Math. 142 (2006) 222–250 MR

Mathematics and Statistics, University of Victoria
PO Box 1700 STN CSC, Victoria, BC V8W 2Y2, Canada

Department of Mathematics, University of Tennessee
227 Ayres Hall, 1403 Circle Dr, Knoxville, TN 37996, United States

Department of Mathematics and Statistics, University of Massachusetts-Amherst
Leaderless Graduate Research Tower, Amherst, MA 01003, United States

Department of Mathematics, University of Oregon
Eugene, OR 97403, United States

ryan.budney@gmail.com, jconant@utk.edu, koytcheff@math.umass.edu, dps@uoregon.edu

Received: 15 February 2016 Revised: 16 August 2016

Algebraic & Geometric Topology 17 (2017) 1743–1770

Affine Hirsch foliations on 3–manifolds

BIN YU

This paper is devoted to discussing affine Hirsch foliations on 3–manifolds. First, we prove that up to isotopic leaf-conjugacy, every closed orientable 3–manifold M admits zero, one or two affine Hirsch foliations. Furthermore, every case is possible.

Then we analyze the 3–manifolds admitting two affine Hirsch foliations (we call these *Hirsch manifolds*). On the one hand, we construct Hirsch manifolds by using exchangeable braided links (we call such Hirsch manifolds *DEBL Hirsch manifolds*); on the other hand, we show that every Hirsch manifold virtually is a DEBL Hirsch manifold.

Finally, we show that for every $n \in \mathbb{N}$, there are only finitely many Hirsch manifolds with strand number n. Here the strand number of a Hirsch manifold M is a positive integer defined by using strand numbers of braids.

57M50, 57R32; 37E10, 57M25

1 Introduction

In 1975, Hirsch [8] constructed an analytic 2–foliation on a closed 3–manifold such that the foliation contains exactly one exceptional minimal set. Let's briefly recall his construction here.

The foliation is constructed by starting with a solid torus and removing from the interior another solid torus which wraps around the original solid torus twice. This gives us a manifold, foliated by 2–punctured disks, with two transverse tori as boundary components. We then glue the exterior boundary component to the interior boundary component to obtain a foliated manifold without boundary. Hirsch chose a gluing map carefully so that the 2–punctured fibration structure induces a foliation and the induced foliation is analytic and contains exactly one exceptional minimal set.

There are many variations of Hirsch's construction in the literature, for instance, Ghys [7], Biś, Hurder and Shive [4]:

- Ghys [7] considered a variant of Hirsch's construction: Hirsch's gluing map is changed to a map that is "affine" in some sense. In [4], the authors call these foliations *affine Hirsch foliations*.

Published: 17 July 2017 DOI: 10.2140/agt.2017.17.1743

• In [4], the authors generalize Hirsch's construction in many cases. In the three-dimensional case, they generalize Hirsch's construction by starting with a solid torus V and removing from V a small solid torus V_0 so that V_0 can be regarded as a tubular neighborhood of a closed twisted braid in V.

Actually, it is natural to further generalize these foliations by using braids:

• For every n–braid b whose closure is a knot, starting with a solid torus V and removing from its interior a small solid torus V_0 which is a small tubular neighborhood of the closure of b, we get a compact 3–manifold, foliated by n–punctured disks, with two boundary components transverse to the n–punctured disk fibration.

• Then we glue the exterior boundary component to the interior boundary component to obtain a foliated manifold induced by the n–punctured disk fibration.

For simplicity, we still call the new foliations *Hirsch foliations*, which are the main objects in this paper. Similarly, if the gluing map is "affine" in some sense, we call the Hirsch foliation *affine*. More precise definitions can be found in Section 2.

There are several kinds of discussions about Hirsch foliations in the literature:

• Biś, Hurder and Shive [4] generalized Hirsch's construction to construct analytic foliations of arbitrary codimension with exotic minimal sets.

• Alvarez and Lessa [1] considered the Teichmüller space of a Hirsch foliation.

• Shive in his thesis [12] considered the following conjugacy question: fixing two Hirsch foliations $(M_1, \mathcal{H}_1)$ and $(M_2, \mathcal{H}_2)$, a C_r leaf-conjugacy diffeomorphism $H\colon M_1 \to M_2$ and an integer $k \in \mathbb{N}$, how does one find conditions on the foliations and the map H which ensure that the map H is $C_{k+\lambda}$?

In this paper, we also would like to discuss a conjugacy question. In contrast to what Shive did, we hope to understand the leaf-conjugacy classes of Hirsch foliations. We say two foliations $\mathcal{H}_1$ and $\mathcal{H}_2$ on a closed 3–manifold M are *isotopically leaf-conjugate* if there exists a homeomorphism $h\colon M \to M$ which maps every leaf of $\mathcal{H}_1$ to a leaf of $\mathcal{H}_2$ and is isotopic to the identity map on M. We say that $\mathcal{H}_1$ and $\mathcal{H}_2$ are the same up to *isotopic leaf-conjugacy* if $\mathcal{H}_1$ and $\mathcal{H}_2$ are isotopically leaf-conjugate. In this paper, we will restrict ourselves to affine Hirsch foliations. The reasons why we focus on affine Hirsch foliations are the following:

• A Hirsch foliation always can be easily rebuilt (see Remark 4.4) by modifying the gluing map of an affine Hirsch foliation.

• Affine Hirsch foliations are natural objects in dynamical systems: the projection of the stable manifolds of a Smale solenoid attractor on the orbit space of the wandering set (by a Smale solenoid mapping on a solid torus) is an affine Hirsch foliation. Our forthcoming paper [16] will focus on this topic.

Now we can naturally ask the following question as the main motivation for this paper.

Question 1.1 For a closed 3–manifold M, can we classify all affine Hirsch foliations up to isotopic leaf-conjugacy?

Alvarez and Lessa [1, Section 1.3] have discussed this question on the 3–manifolds constructed by Hirsch. As a first step toward answering Question 1.1, we have:

Theorem 1.2 *Let M be a closed orientable 3–manifold. Then M admits 0, 1 or 2 affine Hirsch foliations up to isotopic leaf-conjugacy.*

Then one naturally would like to answer:

Question 1.3 (1) Which 3–manifolds admit a Hirsch foliation?

(2) Which 3–manifolds admit two nonisotopically leaf-conjugate affine Hirsch foliations and what are the relations between these two foliations?

Actually, to the first item of Question 1.3, on the one hand, which manifolds these are is very clear, ie each one is precisely determined by a braid and a gluing map; on the other hand, it is not easy to describe all of these manifolds in a straightforward way. Nevertheless, we would like to give some characterizations of these 3–manifolds.

Proposition 1.4 *Let M be a closed orientable 3–manifold which admits an (affine) Hirsch foliation. Then*

(1) *M is a toroidal 3–manifold whose JSJ diagram is cyclic;*

(2) *each JSJ piece is either hyperbolic or a $S(0,2;q/p)$–type Seifert manifold where p and q ($0 < q < p$) are coprime.*

This proposition is a consequence of Lemma 3.7 and Corollary 3.8.

We are more interested in the second item of Question 1.3. We call a 3–manifold M a *Hirsch manifold* if M admits two nonisotopically leaf-conjugate Hirsch foliations. Notice that the 3–manifold constructed by Hirsch in [8] actually is a Hirsch manifold.

Actually, there are many Hirsch manifolds; see Section 4.2 and Proposition 4.3. The following are the reasons why we are interested in Hirsch manifolds:

- A Hirsch manifold has some nice symmetric structures.

- Hirsch manifolds and their two affine Hirsch foliations will play a central role in a class of dynamical systems: in [16], the author will use Hirsch manifolds and affine

Hirsch foliations to discuss a kind of Ω–stable diffeomorphism on 3–manifolds whose nonwandering set is the union of a Smale solenoid attractor and a Smale solenoid repeller.

The exchangeably braided links introduced by Morton [10] will play a crucial role in describing Hirsch manifolds. An *exchangeably braided link* is a two-component link $L = K_1 \cup K_2$ in S^3 such that each component is braided relative to the other one. More details about exchangeably braided links can be found in Section 2.

Motivated by the second item of Question 1.3, we will give two observations to describe the relationships between exchangeably braided links and Hirsch manifolds. The first observation is that for every exchangeably braided link $L = K_1 \cup K_2$, one can build a (unique) Hirsch manifold following a series of standard combinatorial surgeries (see Section 4). Such a Hirsch manifold is called a *Hirsch manifold derived from an exchangeably braided link* (abbreviated as a *DEBL Hirsch manifold*). The second observation is that every Hirsch manifold virtually is a DEBL Hirsch manifold. More precisely:

Theorem 1.5 *Let M be a Hirsch manifold. Then there exists a q_2–covering space of M, denoted by $\widetilde{M}$, such that $\widetilde{M}$ is a Hirsch manifold derived from an exchangeably braided link (a DEBL Hirsch manifold). Moreover, q_2 can be divided by n^2-1 where n is the strand number of M.*

Here, the *strand number* of a Hirsch manifold M (see Definition 4.2) is defined to be the strand number of a braid which can be used to build the Hirsch manifold M.

Hirsch manifolds have the following finiteness property.

Proposition 1.6 *For every $n \in \mathbb{N}$, there are only finitely many Hirsch manifolds with strand number n.*

In the final section (Section 5), we will build an example to show:

Proposition 1.7 *There exists a 3–manifold which admits a Hirsch foliation but is not a Hirsch manifold.*

Proposition 1.4, Proposition 4.3, the examples in Section 4 and Proposition 1.7 imply that there exist closed oriented 3–manifolds M_0, M_1 and M_2 such that

- M_0 doesn't admit any affine Hirsch foliations;
- M_1 admits exactly one affine Hirsch foliation;
- M_2 is a Hirsch manifold, ie M_2 admits two nonisotopically leaf-conjugate affine Hirsch foliations.

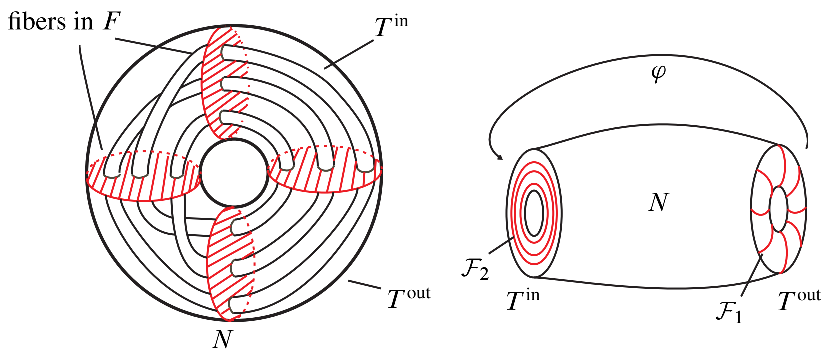

Figure 1: Constructing a Hirsch foliation

This means that every case in Theorem 1.2 can be realized (in some closed 3–manifold).

We can see that the results in this paper (in particular, Proposition 4.3 and Theorem 1.2) give a satisfying response to the problem of classifying all Hirsch foliations. They allow us to reduce classifying all Hirsch foliations to a classical problem in one-dimensional dynamical systems: classifying degree-n ($n \geq 2$) endomorphisms[1] on S^1 up to conjugacy. More details can be found in Remark 4.4.

2 Preliminaries

Definition 2.1 Let $\mathcal{H}$ be a codimension-1 foliation on a closed oriented 3–manifold M. $\mathcal{H}$ is called a *Hirsch foliation* if there exists a torus T embedded into M such that

(1) the path closure of $M - T$, denoted by N, is a compact oriented 3–manifold with two tori T^{out} and T^{in} as its boundary;

(2) $\mathcal{H}|_N$ is an n–punctured disk fibration on N such that each fiber is transverse to ∂N;

(3) every leaf in $\mathcal{H}$ is orientable.

By Definition 2.1, a Hirsch foliation $\mathcal{H}$ on a closed oriented 3–manifold M can be constructed as follows (see Figure 1 for an illustration):

- Choose an n–braid b whose closure is a knot; b also can be used to represent a diffeomorphism on an n–punctured disk Σ.

- We denote the mapping torus of (Σ, b) by N. Notice that $F = \{\Sigma \times \{\star\}\}$ provides a natural n–punctured disk fibration on N, which provides T^{in} and T^{out} two S^1–fibration structures $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively.

[1] An endomorphism on S^1 means a monotonic continuous map on S^1.

- We suppose that Σ is oriented. Then Σ naturally induces an orientation on each fiber of $\mathcal{F}_1$ and $\mathcal{F}_2$. We also give an orientation on N which naturally induces two orientations on T^{out} and T^{in}, respectively.

- Build an orientation-preserving homeomorphism $\varphi\colon T^{\text{out}} \to T^{\text{in}}$ which maps every fiber of $\mathcal{F}_1$ to a fiber of $\mathcal{F}_2$ and preserves the corresponding orientations.[2] Let $M = N \setminus x \sim \varphi(x)$ $(x \in T^{\text{out}}N)$. Then the n–punctured disk fibration F on N naturally induces a Hirsch foliation $\mathcal{H}$ on M by φ.

There are some further comments about Hirsch foliations which will be useful:

- N also can be obtained by removing a small solid torus V_0 from a solid torus V where V_0 is a small tubular neighborhood of the closure of a braid b.

- There is a natural quotient map $P\colon N \to S^1$ where S^1 is the fiber quotient space of F.

- φ induces a map $\varphi_2\colon S^1 \to S^1$, which is called the *projective holonomy map* of $\mathcal{F}$ relative to the embedded torus T.

Definition 2.2 Let $\mathcal{H}$ be a Hirsch foliation on a closed 3–manifold M. $\mathcal{H}$ is called an *affine Hirsch foliation* if the projective holonomy map of $\mathcal{F}$ relative to an embedded torus T transverse to $\mathcal{H}$ is topologically conjugate to the map z^n on S^1 for some $n \in \mathbb{N}$ satisfying $n \geq 2$. Here we can parametrize S^1 by $S^1 = \{z \in \mathbb{C} : |z| = 1\}$.

In 1985, Morton [10] introduced exchangeably braided links. An *exchangeably braided link* is a two-component link $L = K_1 \cup K_2$ which admits a kind of very nice symmetry: each component is braided relative to the other one, ie K_1 is a closed braid $\widetilde{b}_1$ in the solid torus $S^3 - K_2$ and K_2 is a closed braid $\widetilde{b}_2$ in the solid torus $S^3 - K_1$. Such a braid b_1 is called an *exchangeable braid*. Automatically, every exchangeably braided link L can be regarded as the union of the closure of an exchangeable braid and an axis of the closed braid.

Morton [10] showed many nice properties of exchangeably braided links. For instance, he built some necessary and some sufficient conditions for exchangeability. For instance, he showed that the exchangeable braids belong to a family of braids introduced by Stallings [13].

Let's briefly introduce Stallings braids and the relationships between Stallings braids and exchangeable braids. Certainly, the closure of an exchangeable braid is a trivial

[2] $\varphi\colon T^{\text{out}} \to T^{\text{in}}$ preserves the orientations since the glued manifold M should be orientable. φ preserves the corresponding orientations of the fibers of $\mathcal{F}_1$ and $\mathcal{F}_2$ since every leaf in the glued foliation $\mathcal{H}$ should be orientable.

knot. But the converse is not true, ie if the closure of a braid b is a trivial knot, b is not necessarily an exchangeable braid. Actually, Stallings [13] introduced a family of braids in which every braid b satisfies the following:

(1) $\widetilde{b}$ is a trivial knot.

(2) There is a disk D spanning $\widetilde{b}$ which intersects the axis at exactly n points.

Morton called these *Stallings braids*. The set of Stallings braids is a proper subset of the union of the braids whose closure is a trivial knot. In [10], Morton constructed a braid $\omega = \sigma_3\sigma_2\sigma_3^{-1}\sigma_2\sigma_1^{-1}\sigma_2\sigma_1 \in B_4$[3] which is a Stallings braid but not an exchangeable braid. Therefore, the union of the exchangeable braids is a proper subset of the union of Stallings braids.

Stallings braids have a very nice characterization.[4] Under this characterization, it is easy to obtain the following finiteness property.

Proposition 2.3 *For a given $n \in \mathbb{N}$, up to conjugacy, there are finitely many Stallings braids with strand number n.*

The following corollary is an immediate consequence of this proposition (this corollary exactly is [10, Corollary 1.2]).

Corollary 2.4 (1) *Up to conjugacy, there are finitely many exchangeable braids with strand number n.*

(2) *Up to isotopy, there are finitely many exchangeably braided links with linking number n.*

3 Proof of Theorem 1.2

This section is devoted to proving Theorem 1.2. We will prove an equivalent form of the theorem: if a closed 3–manifold M admits a Hirsch foliation $\mathcal{F}$, then up to isotopic leaf-conjugacy, M admits at most two affine Hirsch foliations.

First, we give more notation and parameters (see Figure 2 as an illustration):

- Assume that $i_1\colon T^{\text{out}} \to N$ and $i_2\colon T^{\text{in}} \to N$ are the associated embedding maps and $i_{1,\star}\colon H_1(T^{\text{out}}) \to H_1(N)$ and $i_{2,\star}\colon H_1(T^{\text{in}}) \to H_1(N)$ are the corresponding induced homomorphisms.

- We denote the oriented simple closed curve $\Sigma \cap T^{\text{out}}$ by m_1 and denote the oriented simple closed curves $\Sigma \cap T^{\text{in}}$ by $m_2^1, \dots, m_2^n$. Here, the orientations of the simple closed curves are induced by Σ. Sometimes we also use m_2 to represent m_2^1.

[3] Here and below, the notation for braids is standard in braid theory (see, for instance, Birman [3]).

[4] A careful reader can find the characterization in the beginning of [10, Section 2].

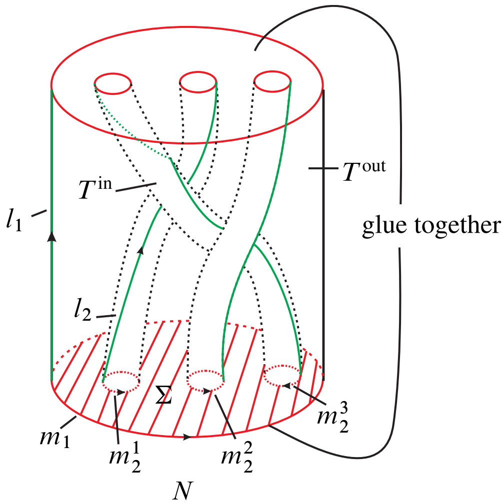

Figure 2: Notation used in the proof of Theorem 1.2

- l_1 is chosen to be an oriented simple closed curve in T^{out} which intersects m_1 at one point such that (m_1, l_1) endows T^{out} with an orientation which is consistent with the orientation of T^{out}.

- l_2 is chosen to be the unique (up to isotopy in T^{in}) oriented simple closed curve in T^{in} such that

  (1) l_2 intersects m_2 at one point;

  (2) (m_2, l_2) endows T^{in} with an orientation which is consistent with the orientation of T^{in};

  (3) $i_{2,\star}([l_2]) = n \cdot i_{1,\star}([l_1])$.

- If there are two oriented simple closed curves m and l on a torus T^2 which intersect at one point, we will use $pm + ql$ (p and q are coprime) to represent an oriented simple closed curve on T^2 which wraps p times around m and q times around l.

The existence and the uniqueness of l_2 can be shown by a short computation on homology, as follows. If we choose a simple closed curve $l_2' \in T^{\text{in}}$ so that l_2' intersects m_2 at one point and the orientation given by (m_2, l_2') is consistent with the orientation of T^{in} given by N, then $i_{2,\star}([l_2']) = n \cdot i_{1,\star}([l_1]) + x \cdot i_{2,\star}([m_2])$ for some $x \in \mathbb{Z}$. Since $i_{2,\star}([m_2])$ is nonzero in $H_1(N)$, there is, up to isotopy, a unique simple closed curve $l_2 = l_2' - x \cdot m_2$ in T^{in} such that $i_{2,\star}([l_2]) = n \cdot i_{1,\star}([l_1])$. For simplicity, we will use $[m_j]$ and $[l_j]$ ($j = 1, 2$) to represent the corresponding elements in $H_1(N)$.

We collect some information about $H_1(N)$ as follows, which can be obtained by Alexander duality.

Lemma 3.1 *$H_1(N) \cong \mathbb{Z} \oplus \mathbb{Z}$, and it is generated by $[l_1]$ and $[m_2]$. Moreover, $[m_1] = n[m_2]$ and $[l_2] = n[l_1]$.*

The following lemma shows that the set of all punctured disk fibrations on N is quite limited.

Lemma 3.2 *Let F be an s–punctured disk fibration on N. Assume that Σ is a fiber of F whose boundary is the union of a simple closed curve $c_1 \in T^{\mathrm{out}}$ and s pairwise parallel and pairwise disjoint simple closed curves $c_2^1, \dots, c_2^s$ in T^{in} (sometimes we also use c_2 to represent c_2^1). Assume that $c_i = p_i m_i + q_i l_i$ $(i = 1, 2)$ where p_i and q_i are coprime. Then there exists an orientation on Σ which induces an orientation on c_1 and an orientation on c_2 such that $s = n$, $p_1 = p_2 = 1$ and $q_1 = n^2 q_2$.*

Proof First let us prove that $p_2 = 1$. If we glue a solid torus V to N by a gluing map $\psi\colon \partial V \to T^{\mathrm{in}}$ so that c_2 bounds a disk in V, then the glued 3–manifold U is homeomorphic to a solid torus. On the one hand, it is obvious that $H_1(U) = \langle [l_1] \rangle \cong \mathbb{Z}$. On the other hand, $H_1(U) = \langle [m_2], [l_1] : p_2[m_2] + q_2[l_2] = p_2[m_2] + n q_2 [l_1] = 0 \rangle$. Therefore, $p_2 = \pm 1$.

Furthermore, we can endow Σ with an orientation which induces two orientations on c_1 and c_2, respectively, so that $p_2 = 1$. These orientations satisfy the requirements in the lemma and will be used in the remainder of the proof.

To conclude, we will prove that $p_1 = 1$ and $s = n$. Since the union of $c_2^1, \dots, c_2^s$ and c_1 bound an s–punctured disk Σ, we have $[c_1] = s[c_2]$. Equivalently, $p_1[m_1] + q_1[l_1] = s([m_2] + n q_2 [l_1])$, and so $q_1 = s q_2 n$ and $p_1 n = s$. We have $q_1 = p_1 q_2 n^2$. Recall that p_1 and q_1 are coprime, and therefore $p_1 = 1$, $s = n$ and $q_1 = n^2 q_2$. □

Remark 3.3 Actually, for every $q_2 \in \mathbb{Z}$, there always exists an associated punctured disk fibration F on N. One can construct it by a standard surgery in low-dimensional topology (for the surgery, see, for instance, Jaco [9, III.14]).

From now on, c_1, c_2 and Σ are oriented as Lemma 3.2.

Lemma 3.4 *Let $\varphi\colon T^{\mathrm{out}} \to T^{\mathrm{in}}$ be a diffeomorphism such that $\varphi(m_1) = m_2$ and $\varphi(l_1) = l_2 + y m_2$ $(y \in \mathbb{Z})$. If $\varphi(c_1)$ is isotopic to c_2 in T^{in}, then c_1 is isotopic to m_1 in T^{out} and c_2 is isotopic to m_2 in T^{in}.*

Proof On the one hand,

$$\begin{aligned}\varphi_\star([c_1]) &= \varphi_\star([m_1]+q_1[l_1]) = \varphi_\star([m_1]) + q_1\varphi_\star([l_1]) \\ &= m_2 + q_1(l_2 + ym_2) = (1+q_1y)m_2 + q_1 l_2;\end{aligned}$$

on the other hand,

$$\varphi_\star([c_1]) = [c_2] = [m_2] + q_2[l_2].$$

Therefore, $1+q_1y=1$ and $q_1=q_2$. By Lemma 3.2, $q_1=n^2q_2$ and $|n|\geq 2$. Then $q_1=q_2=0$. Also notice that $p_1=p_2=1$ (Lemma 3.2). The conclusions of the lemma follow. □

Lemma 3.5 *If F_1 and F_2 are two n–punctured disk fibrations on N with two fibers Σ_1 and Σ_2, respectively, such that $\partial\Sigma_1=\partial\Sigma_2$ is the union of m_1 and n simple closed curves $m_2^1,\dots,m_2^n$ which are pairwise isotopic, then Σ_1 is isotopic to Σ_2 relative to $\partial\Sigma_1=\partial\Sigma_2$ in N.*

Proof Up to isotopy, we can assume that $\mathrm{int}(\Sigma_1)\cap\mathrm{int}(\Sigma_2)$ is the union of finitely many pairwise disjoint simple closed curves $\alpha_1,\dots,\alpha_m$. Here $\mathrm{int}(\Sigma_i)$ $(i=1,2)$ is defined to be the interior of Σ_i. Moreover, we assume that $m\geq 1$ and m is minimal up to isotopy relative to $\partial\Sigma_1=\partial\Sigma_2$.

First, we will show that every α_i $(i\in\{1,\dots,m\})$ is essential in Σ_2. Otherwise, some α_i bounds a disk D_2 in Σ_2. Notice that Σ_1 is incompressible in N, and the union of D_1 and D_2, denoted by S, is a 2–sphere embedded in N. Since N is an irreducible 3–manifold, S bounds a 3–ball in N. This means that we can do a surgery on Σ_2 in a small neighborhood of the 3–ball to obtain Σ_2' so that Σ_2' is isotopic to Σ_2 and the number of connected components of $\Sigma_2'\cap\Sigma_1$ is smaller than m. This contradicts the assumption.

Then there is a nested k–punctured disk $D_1^k\subset\Sigma_1$ with boundary $\alpha_j\cup(m_2^{s_1}\cup\cdots\cup m_2^{s_k})$ for some $j\in\{1,\dots,m\}$, where α_j is an essential simple closed curve in the interior of Σ_1. Here the fact that D_1^k is a nested disk means that the interior of D_1^k is disjoint from Σ_2. We cut N along Σ_1 to obtain a 3–manifold N_0 which is homeomorphic to $\Sigma_1\times[0,1]$. Because $\partial\Sigma_1=\partial\Sigma_2$ and D_1^k is a nested k–punctured disk, by a simple argument on N_0, one can obtain that ∂D_1^k also bounds a nested k–punctured disk D_2^k in N_0. We define Σ_3 to be $(\Sigma_1-D_1^k)\cup D_2^k$, which is an incompressible k–punctured disk. Since N_0 is homeomorphic to $\Sigma_1\times[0,1]$, we have that Σ_3 is isotopic to Σ_1 relative to $\Sigma_1-D_1^k$ in N_0. We can push Σ_3 a little into the interior of N_0 to Σ_3' so that the intersection number of Σ_3' and Σ_2 is strictly smaller than the intersection number of Σ_1 and Σ_2. This contradicts the minimality. □

Now we deal with the trouble that maybe there are many incompressible tori in a Hirsch manifold. For this purpose, we should observe more topological information about N.

First, we recall some classical facts about the geometry and topology of surface bundles. The Nielsen–Thurston theorem (see, for instance, Fathi, Laudenbach and Poenaru [6]) states that a homeomorphism f on a compact surface Σ is isotopic to one of three types according to its dynamics: periodic, reducible and pseudo-Anosov. The Thurston geometrization theorem for surface bundles (see Thurston [14]) implies that the Nielsen–Thurston theorem deeply involves the geometric structure of three-dimensional manifolds as follows: the mapping torus $M_f = \Sigma \times I/(s,1) \sim (f(s),0)$ is an irreducible 3–manifold, and moreover,

1. M_f is hyperbolic if and only if f is pseudo-Anosov;
2. M_f is Seifert-fibered if and only if f is periodic;
3. M_f contains an essential torus (hence we can perform JSJ decomposition) if and only if f is reducible.

In particular, in the third case, there exists a collection of essential simple closed curves in Σ so that the suspension of these curves can be glued up by a map isotopic to f to give a collection of essential tori and Klein bottles which are the collection of JSJ tori and Klein bottles. In the following lemma, we formalize some facts about the geometry and topology of surface bundles which will be very useful.

Lemma 3.6 *Let $M_f = \Sigma \times I/(s,1) \sim (f(s),0)$ be a mapping torus where Σ is a compact orientable surface and f is an orientation-preserving homeomorphism on Σ. Then*

1. *M_f is an irreducible 3–manifold and every JSJ piece of M_f is either hyperbolic or Seifert;*
2. *every JSJ torus T of M_f corresponds to an essential simple closed curve c in Σ which is periodic up to isotopy under f.*

Now, we come back to observing some topological information about N.

Lemma 3.7 *N is an irreducible 3–manifold such that*

1. *every JSJ piece of N is either hyperbolic or Seifert;*
2. *the JSJ diagram of N is a path;*
3. *every Seifert piece is homeomorphic to $S(0,2;q/p)$ where p and q $(0<q<p)$ are coprime and $S(0,2;q/p)$ represents the Seifert manifold whose base orbifold is a 2–punctured sphere with a (q/p)–singularity.*

Proof Recall that N can be defined to be the mapping torus of (Σ, b), where Σ is an n–punctured compact disk and b is a homeomorphism on Σ. Then by item (1) of Lemma 3.6, N is an irreducible 3–manifold such that every JSJ piece is either hyperbolic or Seifert.

By item (2) of Lemma 3.6, every JSJ torus T of N corresponds to an essential simple closed curve c in Σ which is periodic up to isotopy under b. On the one side, notice that every simple closed curve in Σ is separating, and so every JSJ torus of N is separating. This implies that the JSJ diagram of N is a tree. On the other side, ∂N is the union of two tori, T^{out} and T^{in}. Combing these two observations, one could immediately obtain that the JSJ diagram of N is a path.

Let N_0 be a Seifert piece of N. Then N_0 is homeomorphic to a solid torus minus a small open tubular neighborhood of a closed braid $\widetilde{b}_0$. Since N_0 is Seifert, b_0 should be a periodic braid. Since every periodic homeomorphism on a disk is conjugate to a rotation (see Constantin and Kolev [5]), up to conjugacy, b_0 should be a twisted braid. This implies that N_0 is homeomorphic to a Seifert manifold $S(0,2;q/p)$. □

Recall that $M = N \setminus x \sim \varphi(x)$ $(x \in T^{\mathrm{out}}N)$. By Lemma 3.7, the gluing map φ glues the two JSJ pieces corresponding to the two ends of the JSJ diagram of N (notice that the two JSJ pieces might be the same), and the two JSJ pieces should belong to one of the following three cases:

(1) Both of them are hyperbolic.

(2) One of them is hyperbolic and the other one is Seifert.

(3) Both of them are Seifert.

In the first two cases, it is obvious that the glued torus T is a JSJ torus in M. In the third case, since $\varphi(m_1) = m_2$, one can easily check that up to isotopy, $\varphi\colon T^{\mathrm{out}} \to T^{\mathrm{in}}$ doesn't map a regular fiber on T^{out} to a regular fiber on T^{in} induced by the associated Seifert pieces. Therefore, T is also a JSJ torus in M. Now naturally we have the following corollary.

Corollary 3.8 *Let M be a closed orientable 3–manifold which admits a Hirsch foliation. Then every incompressible torus T embedded in M is a JSJ torus and the JSJ diagram of M is cyclic.*

Lemma 3.9 *Let M be a closed 3–manifold which admits an affine Hirsch foliation. We have the following conclusions.*

(1) *M is the union of n JSJ pieces $M_1, M_2, \dots, M_n$ by the gluing maps*

$$\varphi_1\colon \partial^{\mathrm{out}} M_1 \to \partial^{\mathrm{in}} M_2, \quad \dots, \quad \varphi_{n-1}\colon \partial^{\mathrm{out}} M_{n-1} \to \partial^{\mathrm{in}} M_n \quad \text{and} \quad \varphi_n\colon \partial^{\mathrm{out}} M_n \to \partial^{\mathrm{in}} M_1.$$

Here the union of T_i^{out} and T_i^{in} is the boundary of M_i ($i \in \{1, \dots, n\}$).

(2) *Let $\{T_1, \dots, T_n\}$ be a union of the maximal pairwise disjoint, pairwise nonparallel JSJ tori of M and $\mathcal{H}$ be a Hirsch foliation on M. Then $\mathcal{H}$ can be isotopically leaf-conjugate to $\mathcal{H}'$ so that every T_i ($i \in \{1, \dots, n\}$) is transverse to $\mathcal{H}'$.*

Proof Item (1) of the lemma is a direct consequence of Corollary 3.8. We only need to prove item (2).

Without loss of generality, we can suppose that $T_i = \partial^{\mathrm{out}} M_i$ ($i \in \{1, \dots, n\}$) and $\mathcal{H}$ is transverse to T_n. Let N be the union of $M_1, M_2, \dots, M_n$ by the gluing maps $\varphi_1, \dots, \varphi_{n-1}$. The Hirsch foliation $\mathcal{H}$ restricted to N is an m–punctured disk fibration, denoted by F. Since N admits an m–punctured disk fibration F, by Corollary 3.8, every incompressible torus T in the interior of N is a JSJ torus. Moreover, by item (2) of Lemma 3.6, T can be isotopic to T' relative to ∂N so that T' is transverse to F. Then by an easy inductive argument, $T_1, \dots, T_{n-1}$ in N can be isotopic to $T_1', T_2', \dots, T_{n-1}'$ relative to ∂N, respectively, so that every T_i' is transverse to F. Equivalently, we can perturb F in N relative to ∂N to F' which is transverse to every T_i. Then F' naturally induces a foliation $\mathcal{H}'$ in M such that

- $\mathcal{H}'$ is isotopically leaf-conjugate to $\mathcal{H}$;
- $\mathcal{H}'$ is transverse to every T_i. □

Lemma 3.10 *Let M be a closed 3–manifold which admits an affine Hirsch foliation $\mathcal{H}$. Let T_1 and T_2 be two incompressible tori in M each of which is transverse to $\mathcal{H}$. We denote the path closure of $M - T_i$ by N_i ($i = 1, 2$) and denote the restriction of $\mathcal{H}$ to N_i, which is an n_i–punctured disk fibration on N_i, by $\mathcal{F}_i$. Then $n_1 = n_2$.*

Proof Without loss of generality, we can suppose that T_1 and T_2 are disjoint and nonparallel. The path closure of $M - T_1 \cup T_2$ is the union of two compact 3–manifolds W_1 and W_2. Actually, $N_1 = W_1 \cup_{T_2} W_2$ and $N_2 = W_2 \cup_{T_1} W_1$. We denote $\mathcal{H}$ restricted to W_i ($i = 1, 2$), which is an m_i–punctured disk fibration on W_i, by H_i. Notice that every fiber of $\mathcal{F}_1$ is the union of one fiber of H_1 and m_1 fibers of H_2. Therefore, every fiber of $\mathcal{F}_1$ is an $m_1 \cdot m_2$–punctured disk. Equivalently, $n_1 = m_1 \cdot m_2$. Similarly, $n_2 = m_2 \cdot m_1$. In summary, $n_1 = n_2$. □

Definition 3.11 Let M be a closed 3–manifold which admits an affine Hirsch foliation $\mathcal{F}$ and T be an incompressible torus which is transverse to $\mathcal{F}$. We denote by N

the path closure of $M-T$ and denote $\mathcal{F}$ restricted to N, which is an n–punctured disk fibration, by F. We call n the *strand number* of $\mathcal{F}$.

Remark 3.12 Lemma 3.9 and Lemma 3.10 imply that the strand number of $\mathcal{F}$ doesn't depend on the choice of T. Furthermore, by Lemma 3.2, the strand number of an affine Hirsch foliation is invariant under isotopic leaf-conjugacy.

The following lemma explains that the "affine" property of an affine foliation is independent of the choices of T and the foliations which are isotopically leaf-conjugate to the original affine foliation.

Lemma 3.13 *Let M be a closed 3–manifold which admits an affine Hirsch foliation $\mathcal{H}_1$. Let $\mathcal{H}_2$ be a Hirsch foliation such that*

- *$\mathcal{H}_2$ is isotopically leaf-conjugate to $\mathcal{H}_1$;*
- *$\mathcal{H}_2$ is transverse to an incompressible torus T in M and N is the path closure of $M-T$.*

Let $\mathcal{F}_2$ be the punctured disk fibration on N and F_2 be the circle fibration of $\mathcal{H}_2$ restricted to T. Then the projective holonomy map of $\mathcal{F}$ relative to an embedded torus T transverse to $\mathcal{H}$ is topologically conjugate to the map z^n on S^1 where n is the strand number of $\mathcal{H}_1$ and $\mathcal{H}_2$.

To show Lemma 3.13, by item (2) of Lemma 3.9, we only need to prove the following claim.

Claim 3.14 *Let M be a closed 3–manifold which admits an affine Hirsch foliation $\mathcal{H}$. Let T_1 and T_2 be two incompressible tori in M. Let N_i $(i=1,2)$ be the path closure of $M-T_i$, let $\mathcal{F}_i$ be the punctured disk fibration on N_i, and let F_i be the circle fibration of $\mathcal{H}$ restricting to T_i. Suppose that φ_2^1, the projective holonomy map of $\mathcal{F}_1$ relative to T_1, is topologically conjugate to the map z^n on S^1. Then φ_2^2, the projective holonomy map of $\mathcal{F}_2$ relative to T_2, is also topologically conjugate to the map z^n on S^1.*

Proof By Lemma 3.9, we can suppose that T_1 and T_2 are disjoint and nonparallel. Let the path closure of $M-T_1\cup T_2$ be the union of two compact 3–manifolds M_1 and M_2 such that

1. $\partial M_i=\partial^{\mathrm{out}}M_i\cup\partial^{\mathrm{in}}M_i$ for $i=1,2$;
2. M is the union of M_1 and M_2 by the gluing maps $\varphi^1\colon \partial^{\mathrm{out}}M_1\to\partial^{\mathrm{in}}M_2$ and $\varphi^2\colon \partial^{\mathrm{out}}M_2\to\partial^{\mathrm{in}}M_1$;
3. $\partial^{\mathrm{out}}M_1$ and $\partial^{\mathrm{in}}M_2$ correspond to T_1 and $\partial^{\mathrm{out}}M_2$ and $\partial^{\mathrm{in}}M_1$ correspond to T_2.

Under this notation, $N_1 = M_2 \cup_{\varphi_2} M_1$ and $N_2 = M_1 \cup_{\varphi_1} M_2$. We denote by $P_i\colon N_i \to S^1_i$ the quotient map of the fiber quotient space of $\mathcal{F}_i$.

We can define an m–covering map $\pi\colon S^1_2 \to S^1_1$ ($m \in \mathbb{N}$) as follows. For every $z_2 \in S^1_2$, since S^1_2 can be regarded as the quotient space of the circle fibration F_2 on $\partial^{\mathrm{out}} M_2$, we can regard z_2 as a fiber of F_2. Also notice that $\partial^{\mathrm{out}} M_2$ is embedded into N_1, and so the fiber z_2 is in some punctured disk fiber of $\mathcal{F}_1$. Therefore, the quotient map $P_1\colon N_1 \to S^1_1$ naturally induces a map $\pi\colon S^1_2 \to S^1_1$. One can easily check that π is an m–covering map.

We claim that $\varphi^1_2 \circ \pi = \pi \circ \varphi^2_2$, which is the key observation for the proof. Now let's check this claim. For every point $x_i \in N_i$ ($i = 1, 2$), we denote by $\langle x_i \rangle_i \in S^1_i$ the fiber of $\mathcal{F}_i$ where x_i lies. Let x_2 be a point in $\partial^{\mathrm{out}} M_2 \subset N_2$ and x_1 be a point in $\partial^{\mathrm{out}} M_1 \subset N_1$ such that $\langle x_1 \rangle_1 = \pi(\langle x_2 \rangle_2)$. Then one can easily show that $P_1 \circ \varphi^1(x_1) = \pi \circ P_2 \circ \varphi^2(x_2)$ by following the definitions of P_i and φ^i ($i = 1, 2$) and π. Note $P_1 \circ \varphi^1(x_1) = \varphi^1_2(\langle x_1 \rangle_1)$ and $\pi \circ P_2 \circ \varphi^2(x_2) = \pi \circ \varphi^2_2(\langle x_2 \rangle_2)$. By these equalities, we have $\varphi^1_2 \circ \pi(\langle x_2 \rangle_2) = \pi \circ \varphi^2_2(\langle x_2 \rangle_2)$ for every $\langle x_2 \rangle_2 \in S^1_2$.

Since φ^1_2 is affine, we can endow S^1_1 with a suitable metric such that $S^1_1 = \{z \in \mathbb{C} : |z| = 1\}$ and $\varphi^1_2 = z^n$ for some $n \in \mathbb{N}$ ($n \geq 2$). Since $\pi\colon S^1_2 \to S^1_1$ is an m–covering map, we also can endow S^1_2 with a metric such that $S^1_2 = \{z \in \mathbb{C} : |z| = 1\}$ and $\pi(z) = z^m$ for every $z \in S^1_2$. Furthermore, by the fact that $\pi \circ \varphi^2_2 = \varphi^1_2 \circ \pi$, we have $\varphi^2_2 = z^n\colon S^1_2 \to S^1_2$. □

Proposition 3.15 *Let T be an incompressible torus on a closed 3–manifold M. We denote by N the path closure of $M - T$, so that ∂N is the union of T^{out} and T^{in}. Then up to isotopic leaf-conjugacy, there exists at most one affine Hirsch foliation $\mathcal{H}$ such that $\mathcal{H}$ is transverse to T and $\mathcal{H}|_N$ is a punctured disk fibration such that each fiber of $\mathcal{H}|_N$ intersects T^{out} in one connected component.*

Proof We assume that $\mathcal{H}_1$ and $\mathcal{H}_2$ are two affine Hirsch foliations on M which satisfy the conditions in the proposition. Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be the punctured disk fibrations induced on N by $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. Suppose $\varphi\colon T^{\mathrm{out}} \to T^{\mathrm{in}}$ is the gluing map so that $M = N \setminus x \sim \varphi(x)$ ($x \in T^{\mathrm{out}}$).

$F^{\mathrm{out}}_i = \mathcal{F}_i \cap T^{\mathrm{out}}$ ($i = 1, 2$) is an S^1–fibration on T^{out}. Similarly, $F^{\mathrm{in}}_i = \mathcal{F}_i \cap T^{\mathrm{in}}$ is an S^1–fibration on T^{in}. We denote a fiber of F^{out}_1 (resp. F^{in}_1, F^{out}_2, F^{in}_2) by m_1 (resp. m_2, c_1, c_2). Then, up to isotopy, $\varphi(m_1) = m_2$ and $\varphi(c_1) = c_2$. By Lemma 3.4, c_1 is isotopic to m_1 in T^{out} and c_2 is isotopic to m_2 in T^{in}. Then we can suppose that $\mathcal{H}_1 \cap T = \mathcal{H}_2 \cap T$, which we denote by F. Here F is a circle fibration on T.

Since each of $\mathcal{H}_1$ and $\mathcal{H}_2$ is an affine Hirsch foliation, by Lemma 3.13, the projective holonomy maps $\varphi^1_2\colon S^1 \to S^1$ of $\mathcal{H}_1$ and $\varphi^2_2\colon S^1 \to S^1$ of $\mathcal{H}_2$ relative to T

are conjugated by an orientation-preserving homeomorphism $g\colon S^1 \to S^1$, that is, $\varphi_2^2 \circ g = g \circ \varphi_2^1$.

Recall that $P\colon N \to S^1$ is a natural quotient map where S^1 is the fiber quotient space of F. One can lift g to a homeomorphism $G_T\colon T \to T$ such that

- G_T is isotopic to the identity map on T;
- $P \circ G_T = g \circ P$.

Since G_T is isotopic to the identity map on T, we can extend G_T to a homeomorphism $G\colon M \to M$ which is isotopic to the identity map on M. Assume that $\mathcal{H}_1' = G(\mathcal{H}_1)$ is also an affine Hirsch foliation on N. Let $\mathcal{F}_1'$ be the punctured disk fibrations induced by $\mathcal{H}_1'$ on N. By $P \circ G_T = g \circ P$ and $\varphi_2^2 \circ g = g \circ \varphi_2^1$, one can quickly check that the boundaries of $\mathcal{F}_1'$ and $\mathcal{F}_2$ are coherent, ie for every fiber $\Sigma_1 \subset \mathcal{F}_1'$, there exists a fiber Σ_2 such that $\partial\Sigma_1 = \partial\Sigma_2$. Then by Lemma 3.5, one can build a homeomorphism $\phi\colon N \to N$ such that

- ϕ is isotopic to the identity map on N relative to ∂N;
- $\phi(\mathcal{F}_1') = \mathcal{F}_2$.

ϕ can automatically induce a homeomorphism Φ on M such that

- $\Phi(x) = \phi(x)$ for every x in the interior of N;
- Φ is isotopic to the identity map on M;
- $\Phi(\mathcal{H}_1') = \mathcal{H}_2$.

In summary, $\Phi \circ G$ is a homeomorphism on M such that

- $\Phi \circ G$ is isotopic to the identity map on M;
- $\Phi \circ G(\mathcal{H}_1) = \mathcal{H}_2$. □

Now we can finish the proof of Theorem 1.2, ie we can show that up to isotopic leaf-conjugacy, a closed orientable 3–manifold admits at most two affine Hirsch foliations.

Proof of Theorem 1.2 Let $\mathcal{H}$ be an affine Hirsch foliation and T be an incompressible torus in M. By Lemma 3.9, we can suppose that $\mathcal{H}$ is transverse to T. We denote the path closure of $M - T$ by N and the boundary of N by the union of T^{out} and T^{in}. Then $F = \mathcal{H}|_N$ is a punctured disk fibration on N. There are two possibilities for F:

(1) Each leaf of F intersects T^{out} in one connected component.

(2) Each leaf of F intersects T^{in} in one connected component.

In both cases, by Proposition 3.15, up to isotopic leaf-conjugacy, there exists at most one affine Hirsch foliation. The conclusion of the theorem follows. □

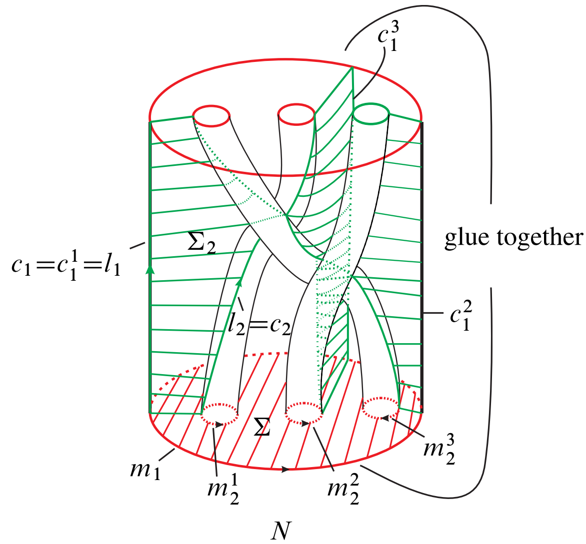

Figure 3: Notation in the case of N associated to the braid $\sigma_1\sigma_2^{-1}$

4 Hirsch manifolds and exchangeably braided links

In this section, we will focus on the study of Hirsch manifolds, ie the closed 3–manifolds which admit two nonisotopically leaf-conjugate affine Hirsch foliations. First, we will introduce or recall some useful notation (see Figure 3 as an illustration[5]):

$\mathcal{H}_1$	an affine Hirsch foliation transverse to T in M
$N, T^{\mathrm{out}}, T^{\mathrm{in}}, \varphi$	$M = N \setminus x \sim \varphi(x)$, $\partial N = T^{\mathrm{out}} \cup T^{\mathrm{in}}$, and $\varphi\colon T^{\mathrm{out}} \to T^{\mathrm{in}}$ is the gluing homeomorphism
$m_1, l_1; m_2, l_2$	$\mathcal{H}_1$ induces oriented simple closed curves m_1, l_1 in T^{out} and m_2, l_2 in T^{in} which are defined at the beginning of Section 3
c_2	$p_2m_2 + q_2l_2$ ($q_2 > 0$), an oriented simple closed curve in T^{in}
c_1	$p_1m_1 + q_1l_1$, an oriented simple closed curve in T^{out}
$c_1^1, \dots, c_1^s$	s pairwise disjoint oriented simple closed curves which are parallel to c_1 in T^{out}
Σ_2	an oriented punctured disk in N such that $\partial\Sigma_2$ is the union of $c_1^1, \dots, c_1^s$ and c_2
F_2	an oriented punctured disk fibration on N with a fiber Σ_2
$\varphi\colon T^{\mathrm{out}} \to T^{\mathrm{in}}$	$\varphi(m_1) = m_2$ and $\varphi(l_1) = l_2 + km_2$

[5] In the case of the figure, $c_1 = c_1^1 = l_1$. To avoid misunderstanding, we should point out that generally we can think $c_1 = c_1^1$, but l_1 may not be isotopic to c_1.

4.1 Homology and Hirsch manifolds

In this subsection, for every two nonzero integers m and n, we will use $[m,n]$ to represent their greatest positive common divisor.

Lemma 4.1 *Suppose F_2 also induces an affine Hirsch foliation $\mathcal{H}_2$ on M under φ. Then $p_1 = k/[n^2-1,k]$, $q_1 = (n^2-1)/[n^2-1,k]$, $p_2 = n^2 p_1$, $q_2 = q_1$, and $s = n$.*

Proof Since $\partial(\Sigma_2) = c_2 \cup c_1^1 \cup \cdots \cup c_1^s$, we have $[c_2] = s[c_1]$ in $H_1(N)$. Equivalently, $p_2[m_2] + q_2[l_2] = sp_1[m_1] + sq_1[l_1]$. Recall that $[m_1] = n[m_2]$ and $[l_2] = n[l_1]$, and so $(snp_1 - p_2)[m_2] + (nq_2 - sq_1)[l_1] = 0$. Recall that $H_1(N) = \langle [m_2],[l_1]\rangle \cong \mathbb{Z}\oplus\mathbb{Z}$. Then

$$p_2 = snp_1 \quad \text{and} \quad nq_2 = sq_1. \tag{1}$$

By filling a solid torus to N along T^{out}, we obtain a new compact 3–manifold V so that c_1 bounds a disk in V, and then V is homeomorphic to a solid torus. Following the gluing surgery, we have

$$\begin{aligned} H_1(V) &= \langle [m_2],[l_1] : p_1[m_1] + q_1[l_1] = 0\rangle \\ &= \langle [m_2],[l_1] : np_1[m_2] + q_1[l_1] = 0\rangle \\ &\cong \mathbb{Z}. \end{aligned}$$

Then we have

$$np_1 \text{ and } q_1 \text{ are coprime.} \tag{2}$$

Define $\varphi_\star$: $H_1(T^{\text{out}}) \to H_1(T^{\text{in}})$ to be the homomorphism induced by φ: $T^{\text{out}} \to T^{\text{in}}$. Notice that F_2 also induces an affine Hirsch foliation $\mathcal{H}_2$ on M. Then, on the one hand,

$$\begin{aligned} \varphi_\star([c_1]) &= [c_2] \\ &= p_2[m_2] + q_2[l_2], \end{aligned}$$

and on the other hand,

$$\begin{aligned} \varphi_\star([c_1]) &= \varphi_\star(p_1[m_1] + q_1[l_1]) \\ &= p_1[m_2] + q_1(k[m_2] + [l_2]) \\ &= (p_1 + q_1 k)[m_2] + q_1[l_2]. \end{aligned}$$

Therefore,

$$p_1 + q_1 k = p_2 \quad \text{and} \quad q_1 = q_2. \tag{3}$$

Now, the lemma is a direct consequence of (1), (2), (3) and the fact that p_i and q_i ($i = 1, 2$) are coprime. □

We can use the strand number of a braid which builds an (affine) Hirsch foliation on M to be an invariant of M. The strand number of M is well defined. Let us explain a little bit more. Suppose there are two braids which build the same Hirsch manifold M; by Lemma 3.10, Definition 3.11 and Lemma 4.1, the strand numbers of the braids are the same.

Definition 4.2 Let M be a Hirsch manifold. The strand number of a braid b which builds an (affine) Hirsch foliation on M is called the *strand number* of M.

Proposition 4.3 *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two affine Hirsch foliations defined as above on a Hirsch manifold M. Then $\mathcal{H}_1$ and $\mathcal{H}_2$ are not isotopically leaf-conjugate.*

Proof Otherwise, we assume that there exists a homeomorphism $h: M \to M$ which maps every leaf of $\mathcal{H}_1$ to a leaf of $\mathcal{H}_2$ and is isotopic to the identity map on M. One can check that every leaf on $\mathcal{H}_1$ is homeomorphic to either a sphere minus a Cantor set or a torus minus a Cantor set. We choose a leaf ℓ_1 on $\mathcal{H}_1$ which is homeomorphic to a sphere minus a Cantor set. We denote $f(\ell_1)$, which is a leaf on $\mathcal{H}_2$, by ℓ_2.

Let $Q: N \to M$ be the natural quotient map. By the construction of $\mathcal{H}_1$, without loss of generality, we can assume that $b_1 = Q(m_1) = Q(m_2)$ is an oriented simple closed curve on ℓ_1. The curve b_1 is homotopically nontrivial in M because of the compressibility of T in M. Since h is isotopic to the identity map on M, we also have that $b_2 = h(b_1) \subset \ell_2$ is homotopically nontrivial in M. By the construction of $\mathcal{H}_2$, b_2 is homotopic to λc in ℓ_2 for some nonzero integer λ. Here $c = Q(c_1) = Q(c_2)$ is a simple closed curve in T. We choose an oriented closed curve c_λ in T which is homotopic to λc in T. Then b_1 and c_λ are homotopic in M. This means that there exists an immersion map $F: A = S^1 \times [0,1] \to M$ and an orientation on A such that

- $F(S^1 \times \{0\}) = l_1$ and $F(S^1 \times \{1\}) = c_\lambda$, where $S^1 \times \{0\}$ and $S^1 \times \{1\}$ are oriented consistently with the orientation of A;
- $F(\mathrm{int}(A))$ is transverse to T, where $\mathrm{int}(A)$ is the interior of A.

Moreover, under some perturbation of F close to ∂A if necessary, we can assume there exists a neighborhood of ∂A, denoted by $N(A)$, satisfying $F^{-1}(T) \cap N(\partial A) = \partial A$. Then $F^{-1}(T) \cap \mathrm{int}(A)$ is the union of finitely many pairwise disjoint oriented simple closed curves $s_0, s_1, \dots, s_m$ where $s_0 = S^1 \times \{0\}$ and $s_m = S^1 \times \{1\}$. Here the orientation of s_i $(i \in \{0, 1, \dots, m\})$ is consistent with the orientation of s_0 in A. We can assume that m is minimal in the following sense: let $F: A \to M$ be an immersion which satisfies the conditions above; then $F^{-1}(T) \cap \mathrm{int}(A)$ contains at least m connected components. If some s_i is inessential in A, then s_i bounds a disk D_i in A. This means that $F(s_i)$ is homotopically trivial in M. Since $F(s_i) \subset T$ and T is incompressible in M, we have that $F(s_i)$ is homotopically trivial in T. Then by some standard surgery, one can build another $F'\colon A \to M$ which satisfies the conditions above and whose intersection circle number is less than m. This contradicts the assumption for m. Therefore, from now on, we can suppose that each s_i is essential in A. Since A is an annulus, $s_0, s_1, \dots, s_m$ are pairwise parallel in A.

Without loss of generality, we can assume that the union of $s_0, s_1, \dots, s_m$ cuts A into m open annuli $A_1, \dots, A_m$ such that

- $\partial A_i = s_{i-1} \cup s_i$ for $i \in \{1, \dots, m\}$;
- $A_i \cap F^{-1}(T) = \varnothing$.

Therefore, $F(s_{i-1})$ and $F(s_i)$ are homotopic in N for every $i \in \{1, \dots, m\}$. We choose a very small tubular neighborhood of s_0 in A. Then $F(N(s_0))$ belongs to one of the two sides of T in M. The two cases for the position of $F(N(s_0))$ and the relations above induce two kinds of "homotopy chain relations". We denote $Q^{-1}(c_\lambda)$ by $c_\lambda^1 \cup c_\lambda^2$ where $c_\lambda^1 \subset T^{\text{in}}$ and $c_\lambda^2 \subset T^{\text{out}}$. In both cases, we can assume there exist $2m$ oriented closed curves $s_1^1, s_2^1, \dots, s_m^1$ in T^{out} and $s_0^2, s_1^2, \dots, s_{m-1}^2$ in T^{in} such that

- $Q(s_i^1) = Q(s_i^2) = F(s_i)$ and $s_i^2 = \varphi(s_i^1)$ for $i \in \{1, \dots, m-1\}$;
- s_{i-1}^2 and s_i^1 are homotopic in N for $i \in \{1, \dots, m\}$.

In one case, $s_0^2 = c_\lambda^2$ and $s_m^1 = m_1$; in the other case, $s_0^2 = m_2$ and $s_m^1 = c_\lambda^1$.

We will get contradictions in both cases by using homology theory. For every oriented closed curve α in N, we will use $[\alpha]$ to represent the corresponding homological element in $H_1(N)$. Recall that $H_1(N) \cong \mathbb{Z} \oplus \mathbb{Z}$, and it is generated by $[l_1]$ and $[m_2]$ (Lemma 3.1). Moreover, $[l_2] = n[l_1]$ and $[m_1] = n[m_2]$. These facts will be used several times in the following.

In the first case, on the one hand, since $s_0^2 = m_2$ and s_1^1 are homotopic in N, we have $[m_2] = [s_1^1]$ in $H_1(N)$; on the other hand, since s_1^1 is an oriented closed curve in T^{out}, we have $[s_1^1] = r[m_1] + t[l_1]$ for two integers r and t. These two sides imply that $[s_1^1] = nr[m_2] + t[l_1] = [m_2]$ in $H_1(N)$. Notice that $n > 1$, and so the equality is impossible. Therefore, we obtain a contradiction.

In the second case, since s_{i-1}^2 and s_i^1 are homotopic in N $(i \in \{1, \dots, m\})$, we have $[s_{i-1}^2] = [s_i^1]$. In particular, $[s_{m-1}^2] = [s_m^1] = [m_1]$. Since s_{m-1}^2 is an oriented closed curve in T^{in}, we have $[s_{m-1}^2] = r_{m-1}[m_2] + t_{m-1}[l_2]$ for two integers r_{m-1} and t_{m-1}. We also have $[s_{m-1}^2] = r_{m-1}[m_2] + nt_{m-1}[l_1] = n[m_2]$. Therefore, $r_{m-1} = n$ and $t_{m-1} = 0$. This implies that s_{m-1}^2 and nm_2 are homotopic in T^{in}. Notice that $s_{m-1}^1 = \varphi^{-1}(s_{m-1}^2)$ and

$\varphi(m_1)=m_2$, and so s^1_{m-1} and nm_1 are homotopic in T^{out}. By some similar arguments, we have that s^1_i and $n^{m-i}m_1$ are homotopic in N for every $i \in \{1,\dots,m-1\}$. Also notice that s^1_1, s^2_0 and c_λ are pairwise homotopic, and so $n^{m-1}m_1$ and c_λ are homotopic in N. This implies that $n^{m-1}[m_1]=[c_\lambda]=n^{m-1+1}[m_2]=n^m[m_2]$. By Lemma 4.1, $[c_\lambda]=\lambda[c_2]=\lambda\big(n^2(k/[n^2-1,k])[m_2]+((n^2-1)/[n^2-1,k])[l_2]\big)$. Since $((n^2-1)/[n^2-1,k])[l_2]=(n(n^2-1)/[n^2-1,k])[l_1]$ is nonzero, $[c_\lambda]\neq n^m[m_2]$. We obtain a contradiction. Then the proposition follows. □

Remark 4.4 By Definition 2.1 and Definition 2.2, we can see that for a given 3–manifold M,

- on the one hand, every Hirsch foliation can be obtained from a unique affine Hirsch foliation by replacing the projective holonomy map $\varphi_2=z^n$ on S^1 by another degree-n endomorphism φ_2' on S^1;
- on the other hand, for every affine Hirsch foliation and every degree-n endomorphism φ_2' on S^1, one can build a Hirsch foliation with the projective holonomy map φ_2'.

Moreover, by Proposition 4.3 and Theorem 1.2, one can classify all of the affine Hirsch foliations on a given 3–manifold M.

Therefore, our results reduce the question of classifying all Hirsch foliations to a classical problem in one-dimensional dynamical systems: classifying degree-n ($n\geq 2$) endomorphisms on S^1 up to conjugacy.

4.2 DEBL Hirsch manifolds

To aid understanding of the materials in this subsection, we suggest the reader look at Figure 4.

Let $L=K_1\cup K_2$ be an exchangeably braided link in S^3. We choose two disjoint small open tubular neighborhoods V_1 and V_2 of K_1 and K_2, respectively. N is defined to be $S^3-V_1\cup V_2$. Its boundary ∂N satisfies $\partial N=T^{\text{out}}\cup T^{\text{in}}$ with $T^{\text{out}}=\partial\overline{V_1}$ and $T^{\text{in}}=\partial\overline{V_2}$. The linking number of K_1 and K_2 is denoted by n. K_1 is a closed n–braid $\widetilde{b}_1$ relative to K_2, and K_2 is a closed n–braid $\widetilde{b}_2$ relative to K_1.

Up to isotopy, there is a unique way to choose a simple closed curve m_1 in T^{out} and n simple closed curves $m_2^1,\dots,m_2^n$ in T^{in} such that

- $m_2^1,\dots,m_2^n$ each bound a disk in $\overline{V_2}$ and m_1 is isotopic to K_1 in $\overline{V_1}$;
- $m_2^1,\dots,m_2^n$ and m_1 bound an n–punctured disk Σ_1 in N.

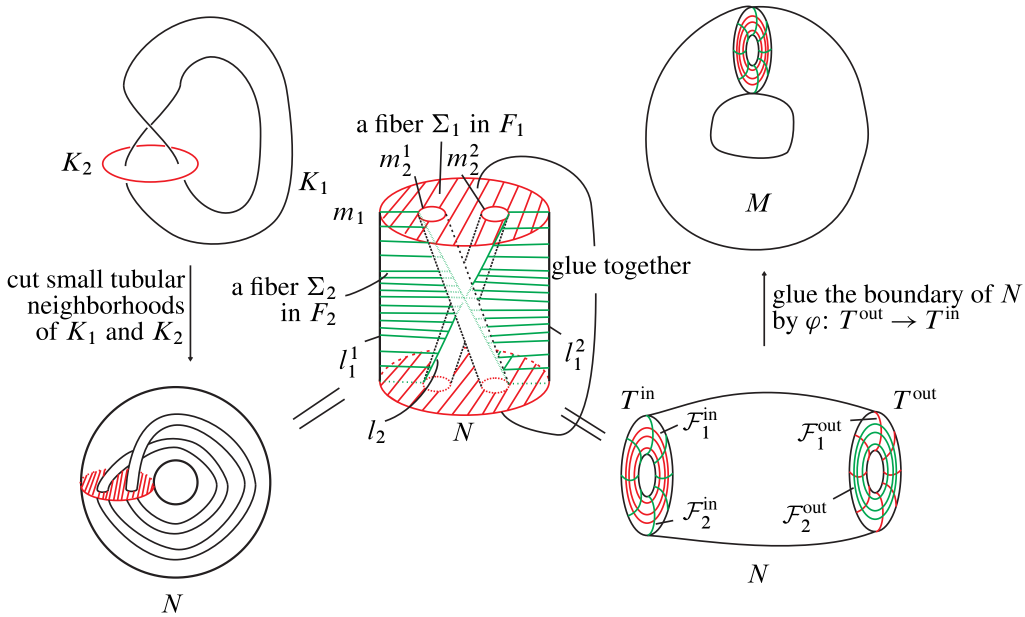

Figure 4: An example DEBL Hirsch manifold

Similarly, up to isotopy, there is a unique way to choose a simple closed curve l_2 in T^{in} and n simple closed curves $l_1^1, \dots, l_1^n$ in T^{out} such that

- $l_1^1, \dots, l_1^n$ each bound a disk in $\overline{V_1}$ and l_2 is isotopic to K_2 in $\overline{V_2}$;
- $l_1^1, \dots, l_1^n$ and l_2 bound an n–punctured disk Σ_2 in N.

Moreover, up to isotopy, we can extend Σ_i $(i = 1, 2)$ to an n–punctured disk fibration F_i on N such that

- $\mathcal{F}_i^{\text{out}} = F_i \cap T^{\text{out}}$ and $\mathcal{F}_i^{\text{in}} = F_i \cap T^{\text{in}}$ are two S^1–fibrations on T^{out} and T^{in};
- $\mathcal{F}_1^{\text{out}}$ and $\mathcal{F}_2^{\text{out}}$ transversely intersect everywhere on T^{out};
- $\mathcal{F}_1^{\text{in}}$ and $\mathcal{F}_2^{\text{in}}$ transversely intersect everywhere on T^{in}.

We can suppose that the intersection number of m_1 and l_1^i is 1 and the intersection number of m_2^i and l_2, for every $i \in \{1, \dots, n\}$, is 1. Similar to the beginning of Section 2 and Section 3, we would like to provide some orientations on these objects:

- We give N an orientation.
- T^{in} and T^{out} are oriented by the orientation of N.

- We give the leaves of F_1 and F_2 orientations continuously so that
  (1) every fiber of $\mathcal{F}_i^{\text{out}}$ and $\mathcal{F}_i^{\text{in}}$ $(i=1,2)$ is oriented as induced by the orientation of F_1 and F_2;
  (2) the orientation of T^{out} is consistent with that of (m_1, l_1^1) and the orientation of T^{in} is consistent with that of (m_2^1, l_2).

Then one can build an orientation-preserving homeomorphism $\varphi\colon T^{\text{out}} \to T^{\text{in}}$ such that

- $\varphi(m_1) = m_2^1$ and $\varphi(l_1^1) = l_2$;
- φ maps every fiber of $\mathcal{F}_1^{\text{out}}$ to a fiber of $\mathcal{F}_1^{\text{in}}$;
- φ maps every fiber of $\mathcal{F}_2^{\text{out}}$ to a fiber of $\mathcal{F}_2^{\text{in}}$;
- F_1 and F_2 induce two affine Hirsch foliations $\mathcal{H}_1$ and $\mathcal{H}_2$ on M under φ where $M = N \setminus x \sim \varphi(x)$ $(x \in T^{\text{out}})$.

Notice that to ensure the glued manifold M can admit two Hirsch foliations induced by F_1 and F_2, up to isotopy, we can suppose that $\varphi(m_1) = m_2^1$ and $\varphi(l_1^1) = l_2$. This implies that under this restriction, the glued manifold M is unique up to homeomorphism. Therefore, we can say that an exchangeably braided link determines a unique Hirsch manifold. Every Hirsch manifold built in this way is called a *Hirsch manifold derived from an exchangeably braided link* (abbreviated as a *DEBL Hirsch manifold*).

By the second item of Corollary 2.4, we have the following consequence.

Corollary 4.5 *For every $n \in \mathbb{N}$, there are only finitely many DEBL Hirsch manifolds with strand number n.*

4.3 A virtual property of Hirsch manifolds

Let M be a Hirsch manifold. By the definition of Hirsch manifold and Lemma 3.9, there exist two affine Hirsch foliations $\mathcal{H}_1$ and $\mathcal{H}_2$ on M and a JSJ torus T in M which satisfy the propositions in Lemma 3.9. Let N be the path closure of $M - T$. Then:

- N admits two n–punctured disk fibrations F_1 and F_2 and we can parametrize some subsets of N as in Section 2.
- The two Hirsch foliations $\mathcal{H}_1$ and $\mathcal{H}_2$ can be induced, respectively, by F_1 and F_2 and a gluing map $\varphi\colon T^{\text{out}} \to T^{\text{in}}$.

$H_1(N) \cong \mathbb{Z} \oplus \mathbb{Z}$, and it is generated by $[m_2]$ and $[l_1]$. We denote the abelianization homomorphism from $\pi_1(N)$ to $H_1(N)$ by ψ_1 and denote by ψ_2 the quotient homomorphism from $H_1(N)$ to $\mathbb{Z}_{q_2}$ such that $\psi_2([m_2]) = 0$ and $\psi_2([l_1]) = 1$. The kernel of $\psi_N = \psi_2 \circ \psi_1\colon \pi_1(N) \to \mathbb{Z}_{q_2}$, which we denote by G, is a normal subgroup of $\pi_1(N)$. Here, the definitions and properties of m_i and l_i $(i = 1, 2)$ and q_2 can be found in the beginning of Section 4 and Lemma 4.1.

As a subgroup of $\pi_1(N)$, the kernel G induces a q_2–covering space of N by a covering map $P\colon \widetilde{N} \to N$. We collect some useful properties in the following proposition, which one can prove by some routine checks. We omit the details here.

Proposition 4.6 *The following properties hold for $i = 1, 2$:*

1. $P^{-1}(F_i) = \widetilde{F}_i$ *is an n–punctured disk fibration on* $\widetilde{N}$.
2. *Let $\widetilde{\Sigma}_1$ be a connected component of $P^{-1}(\Sigma_1)$. Then $P\colon \widetilde{\Sigma}_1 \to \Sigma_1$ is a homeomorphism satisfying $P(\widetilde{m}_i) = m_i$.*
3. $P\colon \widetilde{l}_1 \to l_1$ *is a q_2–covering map.*
4. *Let $\widetilde{\Sigma}_2$ be a connected component of $P^{-1}(\Sigma_2)$. Then $P\colon \widetilde{\Sigma}_2 \to \Sigma_2$ is a homeomorphism satisfying $P(\widetilde{c}_i) = c_i$.*
5. $\widetilde{c}_i$ *intersects* $\widetilde{m}_i$ *at one point.*

Lemma 4.7 *There is a homeomorphism $\widetilde{\varphi}\colon \widetilde{T}^{\mathrm{out}} \to \widetilde{T}^{\mathrm{in}}$ such that*

1. $P \circ \widetilde{\varphi} = \varphi \circ P\colon \widetilde{T}^{\mathrm{out}} \to T^{\mathrm{in}}$;
2. $\widetilde{\varphi}(\widetilde{m}_1) = \widetilde{m}_2$ *and* $\widetilde{\varphi}(\widetilde{c}_1) = \widetilde{c}_2$.

Proof By Proposition 4.6, $\widetilde{m}_i \cap \widetilde{c}_i$ $(i = 1, 2)$ is one point, which we will denote by $\widetilde{x}_i$. Denoting $P(\widetilde{x}_i)$ by x_i, we have $\varphi(x_1) = x_2$,

$$(\varphi \circ P)_\star(\pi_1(\widetilde{T}^{\mathrm{out}}, \widetilde{x}_1)) = \langle \varphi_\star([m_1]), \varphi_\star([c_1]) \rangle = \langle [m_2], [c_2] \rangle \lhd \pi_1(T^{\mathrm{in}}, x_2),$$

and

$$P_\star(\pi_1(\widetilde{T}^{\mathrm{in}}, \widetilde{x}_2)) = \langle [m_2], [c_2] \rangle \lhd \pi_1(T^{\mathrm{in}}, x_2).$$

Then, by the classical homotopy lifting theorem, we can construct a unique map $\widetilde{\varphi}\colon \widetilde{T}^{\mathrm{out}} \to \widetilde{T}^{\mathrm{in}}$ such that

1. $\widetilde{\varphi}(\widetilde{x}_1) = \widetilde{x}_2$ and $\widetilde{\varphi}(\widetilde{m}_1) = \widetilde{m}_2$;
2. $P \circ \widetilde{\varphi} = \varphi \circ P\colon \widetilde{T}^{\mathrm{out}} \to T^{\mathrm{in}}$.

Now it is routine to check that $\widetilde{\varphi}$ is a homeomorphism. □

We denote $\widetilde{N} \setminus y \sim \widetilde{\varphi}(y)$ $(y \in \widetilde{T}^{\mathrm{out}})$ by $\widetilde{M}$. The corresponding quotient maps are $Q\colon \widetilde{N} \to \widetilde{M}$ and $q\colon N \to M$.

Lemma 4.8 *There is a unique map $\pi\colon \widetilde{M} \to M$ so that $\pi \circ Q(y) = q \circ P(y)$ for every $y \in \widetilde{N}$. Furthermore, $\pi\colon \widetilde{M} \to M$ is a q_2–covering map.*

Proof For every $\tilde{x} \in \widetilde{M}$, we can define $\pi(\tilde{x})$ as follows. Since $Q\colon \widetilde{N} \to \widetilde{M}$ is surjective, there exists $\tilde{y} \in \widetilde{N}$ so that $\tilde{x} = Q(\tilde{y})$. Define $\pi(\tilde{x}) = q \circ P(\tilde{y})$. The first item of Lemma 4.7 ensures that π is well defined. π is unique because it has no freedom in $\widetilde{M} - \widetilde{T}$ where $\widetilde{T} = \pi^{-1}(T)$.

Finally, since $P\colon \widetilde{N} \to N$ is a q_2–covering map, $\pi\colon \widetilde{M} \to M$ is also a q_2–covering map. □

Lemma 4.9 *$\widetilde{M}$ is a Hirsch manifold which admits two affine Hirsch foliations $\widetilde{\mathcal{H}}_1$ and $\widetilde{\mathcal{H}}_2$ such that $\widetilde{\mathcal{H}}_i$ $(i = 1, 2)$ is induced by $\mathcal{H}_i$ under π, ie π maps each leaf of $\widetilde{\mathcal{H}}_i$ to a leaf of $\mathcal{H}_i$.*

Proof Assume that $\mathcal{F}_i^{\text{out}} = F_i \cap T^{\text{out}}$ and $\mathcal{F}_i^{\text{in}} = F_i \cap T^{\text{in}}$ are two S^1–fibrations on T^{out} and T^{in}, respectively. Since F_i induces $\mathcal{H}_i$ on M under the gluing homeomorphism $\varphi\colon T^{\text{out}} \to T^{\text{in}}$, we know φ maps every fiber of $\mathcal{F}_i^{\text{out}}$ to a fiber of $\mathcal{F}_i^{\text{in}}$.

Suppose $\widetilde{\mathcal{F}}_i^{\text{out}}$ and $\widetilde{\mathcal{F}}_i^{\text{in}}$ are the lifted fibrations of $\mathcal{F}_i^{\text{out}}$ and $\mathcal{F}_i^{\text{in}}$ on $\widetilde{T}^{\text{out}}$ and $\widetilde{T}^{\text{in}}$ under the covering map P, respectively. $\widetilde{\varphi}\colon \widetilde{T}^{\text{out}} \to \widetilde{T}^{\text{in}}$ is the lifted map of $\varphi\colon T^{\text{out}} \to T^{\text{in}}$, ie $P \circ \widetilde{\varphi} = \varphi \circ P\colon \widetilde{T}^{\text{out}} \to T^{\text{in}}$. Therefore, $\widetilde{\varphi}$ maps every fiber of $\widetilde{\mathcal{F}}_i^{\text{out}}$ to a fiber of $\widetilde{\mathcal{F}}_i^{\text{in}}$. Then $\widetilde{F}_i$ induces a Hirsch foliation $\widetilde{\mathcal{H}}_i$ on $\widetilde{M}$.

To finish the proof, now we only need to check that $\widetilde{\mathcal{H}}_i$ is an affine Hirsch foliation. This actually is a consequence of the following facts:

- $\widetilde{\varphi}$ is the lifted map of φ.
- $\mathcal{H}_i$ is an affine Hirsch foliation.
- Every expanding map on S^1 is topologically conjugate to an affine map on S^1 with the same degree. □

Lemma 4.10 *$\widetilde{M}$ is a DEBL Hirsch manifold.*

Proof We glue two solid tori $\widetilde{V}_1$ and $\widetilde{V}_2$ to $\widetilde{N}$ along its boundary $\widetilde{T}^{\text{in}} \cup \widetilde{T}^{\text{out}}$ by the gluing maps $\phi_1\colon \partial\widetilde{V}_1 \to \widetilde{T}^{\text{in}}$ and $\phi_2\colon \partial\widetilde{V}_2 \to \widetilde{T}^{\text{out}}$, respectively, so that $\widetilde{m}_2$ bounds a disk in $\widetilde{V}_2$ and $\widetilde{c}_1$ bounds a disk in $\widetilde{V}_1$. Then the glued manifold is homeomorphic to S^3.

Let K_i $(i = 1, 2)$ be a simple closed curve in $\widetilde{V}_i$ such that $\widetilde{V}_i$ is a tubular neighborhood of K_i.

Since $\widetilde{F}_2$ is a punctured disk fibration structure on $\widetilde{N}$ and $\widetilde{m}_2$ bounds a disk in $\widetilde{V}_2$, the union of $\widetilde{V}_2$ and $\widetilde{N}$, denoted by $\widetilde{U}_2$, is also homeomorphic to a solid torus. Obviously, K_2 is a closed braid in $\widetilde{U}_2$. Since $S^3 = \widetilde{V}_2 \cup \widetilde{N} \cup \widetilde{V}_2$, automatically, K_2 is a closed braid relative to K_1, ie K_2 is a closed braid in $S^3 - K_1$.

Similarly, one can show that K_1 is a closed braid relative to K_2. Therefore, $L = K_1 \cup K_2$ is an exchangeably braided link. Now it is routine to build the Hirsch manifold derived from L and check that the Hirsch manifold is homeomorphic to $\widetilde{M}$. □

Proof of Theorem 1.5 The first part of Theorem 1.5 is a direct consequence of Lemma 4.8 and Lemma 4.10. Moreover, by Lemma 4.1, q_2 can be divided by n^2-1. □

4.4 Finiteness of Hirsch manifolds with strand number *n*

We will use the following theorem of Wang [15].

Theorem 4.11 *Let M be a closed irreducible 3–manifold which is nonorientable or Seifert fibered or has a nontrivial torus decomposition (ie there is a JSJ torus). Then M covers infinitely many nonhomeomorphic 3–manifolds if and only if M is an orientable Seifert fiber space with nonzero Euler number.*

Proof of Proposition 1.6 On the one hand, by Proposition 1.4, an n–strand Hirsch manifold is an irreducible orientable closed 3–manifold with some JSJ tori. By Theorem 4.11, for a given DEBL Hirsch manifold $\widetilde{M}$, there are only finitely many Hirsch manifolds with $\widetilde{M}$ as a finite covering space.

On the other hand, Corollary 2.4 says that for a positive integer n, up to isotopy, there are only finitely many exchangeably braided links with strand number n. Recall that an exchangeably braided link determines a DEBL Hirsch manifold. Therefore, there are only finitely many DEBL Hirsch manifolds with strand number n.

Let M be a Hirsch manifold with strand number n. Then, by Theorem 1.5, $\widetilde{M}$, a finite covering space of M, is a DEBL Hirsch manifold with strand number n. Combining the two sides above, up to homeomorphism, there are only finitely many Hirsch manifolds with strand number n. □

5 Proof of Proposition 1.7

In this section, we will construct an example to prove Proposition 1.7, which says there exists a 3–manifold which admits an affine Hirsch foliation but is not a Hirsch manifold. We will use the following inequality by Bennequin [2]:

Lemma 5.1 (Bennequin inequality) *Let L be a nonseparating link of μ components, presented by a closed braid with l strands and c_+ (c_-) positive (negative) crossings. Then $g(L)$, the genus of L, is bounded as follows:*

$$\frac{|c_+ - c_-| - l - \mu}{2} + 1 \le g(L) \le \frac{|c_+ + c_-| - l - \mu}{2} + 1.$$

Proof of Proposition 1.7 Let $b = (\sigma_1\sigma_2^{-1})^2$ be a 3–strand braid. Now we can follow the beginning of Section 2 to build an affine Hirsch foliation $\mathcal{H}$ on a closed 3–manifold M. We briefly recall the construction here:

- b also can be used to represent a diffeomorphism on a 3–punctured disk Σ, and we denote the mapping torus of (Σ, b) by N.
- $F = \{\Sigma \times \{\star\}\}$ provides a 3–punctured disk fibration on N, which provides T^{in} and T^{out} two S^1–fibration structures $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively.
- After carefully choosing orientations on the objects above, we can build an orientation-preserving homeomorphism $\varphi\colon T^{\text{out}} \to T^{\text{in}}$ which maps every fiber of $\mathcal{F}_1$ to a fiber of $\mathcal{F}_2$ and preserves the corresponding orientations.
- Let $M = N \setminus x \sim \varphi(x)$ $(x \in T^{\text{out}}N)$. Then F naturally induces a Hirsch foliation $\mathcal{H}$ on M by φ. If we choose φ suitably, $\mathcal{F}$ is an affine Hirsch foliation.

Now we assume that M is a Hirsch manifold. Following the arguments in Section 4.3, there exists some integer p so that the braid $b^{q_2}\tau^p$ is an exchangeable braid where τ is a 3–strand full-twist braid. This means that the knot $K = \widehat{b^{q_2}\tau^p}$, the closed braid of $b^{q_2}\tau^p$, is a trivial knot. In the following, we will show that $g(K)$, the genus of K, is nonzero. Then K isn't a trivial knot. We obtain a contradiction. Then M isn't a Hirsch manifold.

Using the notation of Lemma 5.1 in our case, $L = K = \widehat{b^{q_2}\tau^p}$, $l = 3$, $\mu = 1$ and $|c_+ - c_-| = 6|p|$. By Lemma 5.1, $g(K) \geq 3|p| - 1$. Therefore, if $g(K) = 0$, then $p = 0$. In the case $p = 0$, we have $K = \widehat{b^{q_2}}$. By Lemma 4.1, q_2 is nonzero. Actually, it is well known that in this case, $\widehat{b^{q_2}}$ is a genus-1 fiber knot (see, for instance, Rolfsen [11, Chapter 10]). Therefore, K isn't a trivial knot. □

Acknowledgements The author is grateful to Sébastien Alvarez for useful discussions. The author is supported by the National Natural Science Foundation of China (grant no. 11471248).

References

[1] **S Alvarez**, **P Lessa**, *The Teichmüller space of the Hirsch foliation*, preprint (2015) arXiv

[2] **D Bennequin**, *Entrelacements et équations de Pfaff*, from "Third Schnepfenried geometry conference", Astérisque 107, Soc. Math. France, Paris (1983) 87–161 MR

[3] **J S Birman**, *Braids, links, and mapping class groups*, Annals of Mathematics Studies 82, Princeton University Press (1974) MR

[4] **A Biś**, **S Hurder**, **J Shive**, *Hirsch foliations in codimension greater than one*, from "Foliations 2005" (P Walczak, R Langevin, S Hurder, T Tsuboi, editors), World Sci., Hackensack, NJ (2006) 71–108 MR

[5] **A Constantin**, **B Kolev**, *The theorem of Kerékjártó on periodic homeomorphisms of the disc and the sphere*, Enseign. Math. 40 (1994) 193–204 MR

[6] **A Fathi**, **F Laudenbach**, **V Poénaru**, *Travaux de Thurston sur les surfaces*, Astérisque 66, Soc. Math. France, Paris (1979) MR

[7] **E Ghys**, *Topologie des feuilles génériques*, Ann. of Math. 141 (1995) 387–422 MR

[8] **M Hirsch**, *A stable analytic foliation with only exceptional minimal sets*, from "Dynamical systems" (A Manning, editor), Lecture Notes in Math. 468, Springer (1975) 8–9

[9] **W Jaco**, *Lectures on three-manifold topology*, CBMS Regional Conference Series in Mathematics 43, Amer. Math. Soc., Providence, RI (1980) MR

[10] **H R Morton**, *Exchangeable braids*, from "Low-dimensional topology" (R Fenn, editor), London Math. Soc. Lecture Note Ser. 95, Cambridge Univ. Press (1985) 86–105 MR

[11] **D Rolfsen**, *Knots and links*, corrected reprint of 1st edition, Math. Lect. Ser. 7, Publish or Perish, Houston, TX (1990) MR

[12] **J H Shive**, *Conjugation problems for Hirsch foliations*, PhD thesis, University of Illinois at Chicago (2005) MR Available at `http://search.proquest.com/docview/305371555`

[13] **J R Stallings**, *Constructions of fibred knots and links*, from "Algebraic and geometric topology, 2" (R J Milgram, editor), Proc. Sympos. Pure Math. XXXII, Amer. Math. Soc., Providence, RI (1978) 55–60 MR

[14] **W P Thurston**, *Hyperbolic structures on* 3*–manifolds, II: Surface groups and* 3*–manifolds which fiber over the circle*, preprint (1998) arXiv

[15] **S C Wang**, 3*–manifolds which cover only finitely many* 3*–manifolds*, Quart. J. Math. Oxford Ser. 42 (1991) 113–124 MR

[16] **B Yu**, *Smale solenoid attractors and affine Hirsch foliations*, Ergodic Theory and Dynamical Systems (online publication May 2017) 1–23

School of Mathematical Sciences, Tongji University
200092 Shanghai, China

`binyu1980@gmail.com`

Received: 16 April 2016 Revised: 12 October 2016 msp

Algebraic & Geometric Topology 17 (2017) 1771–1783

Equivariant corks

Dave Auckly
Hee Jung Kim
Paul Melvin
Daniel Ruberman

For any finite subgroup G of $\mathrm{SO}(4)$, we construct a contractible 4–manifold C with a G–action on its boundary that can be embedded in a closed 4–manifold so that cutting C out and regluing using distinct elements of G will always yield distinct smooth 4–manifolds. If we simply require G to be a subgroup of the mapping class group of the boundary, then such examples exist for groups that cannot act on any homology sphere.

57M99; 57R55

0 Introduction

A *cork* is a smooth, compact, contractible 4–manifold with an involution on its boundary that does not extend to a diffeomorphism of the full manifold. Akbulut [1] discovered this phenomenon for the classical Mazur manifold $\mathbb{W}$ [18] with the boundary involution τ shown in Figure 1, proving that $\mathbb{W}$ embeds in a 4–manifold X so that the result of removing $\mathbb{W}$ and regluing it using τ is not diffeomorphic to X.

This operation is called *cork twisting*, and it is now known (see Curtis, Freedman, Hsiang and Stong [9] and Matveyev [17]) that any two smooth, closed, simply connected 4–manifolds that are homeomorphic differ by a single cork twist. It is not known whether the same cork can be used in all situations, ie whether there exists a *universal cork*; it is indeed conceivable, though unlikely, that the Mazur cork is universal.

The property that the cork twist τ is an *involution* is interesting, indeed inherent in most constructions of corks to date, but it is not clear that it is fundamental to the

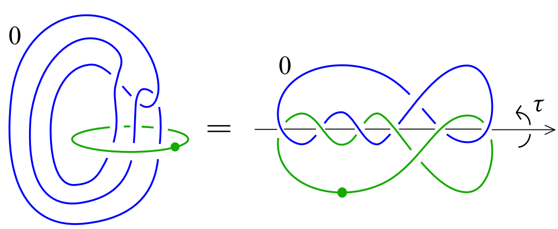

Figure 1: The Mazur cork

Published: 17 July 2017 DOI: 10.2140/agt.2017.17.1771

relation between cork twists and other smooth 4–manifold constructions. It is therefore natural to ask whether cutting and gluing by higher order diffeomorphisms of the boundary of a contractible submanifold of a 4–manifold can change the underlying smooth structure. In this note, we give an affirmative answer, producing examples of embeddings of contractible 4–manifolds with twists of arbitrary finite order that alter the ambient smooth structure; it follows that none of those twists extend over the contractible manifold. A different construction of such nonextending twists was given in a recent preprint of Tange [19].

In fact we show more: for suitable finite groups G, there exist contractible 4–manifolds with effective G–actions on the boundary that embed in closed 4–manifolds so that twists corresponding to distinct elements of G yield distinct smooth structures. We call such a gadget an *equivariant cork*, or G–cork if we want to specify the group.

Theorem A *There exist G–corks for any finite subgroup G of* $\mathrm{SO}(4)$.

If the action of G on S^3 is free, then the action of G on the boundary of the cork constructed in the theorem is free; this seems to be a new phenomenon, even for $G = \mathbb{Z}_2$. The notion of an equivariant cork can be extended to a *weak equivariant cork* where the relevant group is a subgroup of the mapping class group of the boundary; see the end of Section 1 for details. In the final section of the paper, we give an example of a *weak G–cork* in this sense, where G is a group that does not act effectively on any homology 3–sphere.

Theorem B *There are groups G that do not act effectively on any homology sphere, but for which there exist weak equivariant G–corks.*

The boundaries of the corks constructed in the proof of Theorem A are reducible. In a sequel we will prove the following theorem, using rather different techniques from those in the current paper.

Theorem C *Given an oriented 3–manifold Y with an effective, orientation-preserving, smooth action of a finite group G, there is an equivariant invertible $\mathbb{Z}[\pi_1(Y)]$–homology cobordism from it to a hyperbolic manifold.*

As in Akbulut and Ruberman [2], this immediately implies:

Corollary D *For any given finite subgroup G of* $\mathrm{SO}(4)$, *there exists a G–cork with hyperbolic boundary.*

Some experimentation with SnapPy [8] suggests that the simplest corks in Tange's paper [19] have hyperbolic boundaries, but a proof in general would require different techniques.

Acknowledgements The construction of G–corks and their hyperbolization were worked out by the authors at the American Institute of Mathematics (AIM) at our SQuaRE meeting in July 2015. We thank AIM for its support for this and future endeavors. Our results were announced at the 2016 Joint Mathematics Meeting; see [6].

Kim was supported by NRF grants 2015R1D1A1A01059318 and BK21 PLUS SNU Mathematical Sciences Division. Ruberman was supported by the NSF grants DMS-1105234 and DMS-1506328, and by NSF FRG grant 1065827.

1 Preliminaries and statement of results

In this section, we lay the groundwork for our proof of the existence of equivariant corks. Most of the ideas discussed here are well known, but since we will use "corks" in a broader sense than usual, and employ cork twists on multiple copies of boundary sums of embedded copies of the Mazur cork, we must give careful definitions of the relevant notions.

Corks and boundary equivalence

Extending the usual terminology, a *cork* will refer to any pair (C, g) where C is a smooth, compact, contractible 4–manifold, and g is an *arbitrary* diffeomorphism of ∂C. In particular, g need not be an involution, nor even of finite order, and C need not be Stein (as is often assumed; see Akbulut and Yasui [3]). But if g *is* a *special involution* (meaning orientation preserving with nonempty fixed point set, as with the Mazur twist τ) then we also refer to (C, g) as a *special* 2*–cork*.

In general, we call a cork (C, g) *trivial* if g extends to a diffeomorphism of C (it always extends to a homeomorphism by Freedman [11]) and *nontrivial* otherwise; with this convention, (B^4, g) is a trivial cork for any g, whereas the Mazur cork $(\mathbb{W}, \tau)$ is nontrivial. These notions induce an equivalence relation on corks associated with the same underlying manifold: (C, g) and (C, h) are *boundary equivalent* if and only if $(C, g^{-1}h)$ is trivial, ie $g^{-1}h$ extends over C.

Boundary sums of corks

The *boundary sum* operation $\natural$ is well defined on boundary equivalence classes of corks, as follows: Given corks (C_1, g_1) and (C_2, g_2), choose (for $i = 1, 2$) diffeomorphisms h_i isotopic (and thus boundary equivalent) to g_i that are the identity on 3–balls $B_i \subset \partial C_i$. Form $C_1 \natural C_2$ by identifying the C_i along the B_i so that h_1 and h_2 glue together to form $h_1 \sharp h_2$. The result

$$(C_1, g_1) \natural (C_2, g_2) := (C_1 \natural C_2, g_1 \sharp g_2)$$

may depend on the choices of h_i and B_i, but its boundary equivalence class does not. Note however that $\natural$ *is* well defined for special 2–corks *without* imposing boundary equivalence; just choose the B_i to be g_i–invariant 3–balls centered at fixed points, and then $g_1 \sharp g_2$ is a well-defined involution, independent of the choices up to equivariant diffeomorphism.

Cork embeddings

A *cork embedding* of (C, g) in a 4–manifold X is a smooth embedding $e: C \hookrightarrow X$ together with the induced map $\overline{g} = ege^{-1}$ on the boundary of its image $\overline{C} = e(C)$. The associated *cork twist* X^e_g is obtained by removing $\overline{C}$ from X and regluing using $\overline{g}$:

$$X^e_g = (X - \operatorname{int} \overline{C}) \cup_{\overline{g}} \overline{C}.$$

The embedding is *trivial* if X^e_g is diffeomorphic to X, and it is otherwise *nontrivial* or *effective*; note that this definition depends on both e and g. Thus the nontriviality of (C, g) can be verified by producing a nontrivial embedding, rather than trying to show directly that g does not extend smoothly across C.

Note that the definition of boundary equivalence of cork maps is compatible with the use of such maps in changing smooth structures, because the result of twisting by g is the same as the result of twisting by h when $g^{-1}h$ extends across C. Conversely, given any nontrivial cork (C, g), Akbulut and Ruberman [2] construct a pair of absolutely exotic structures on a contractible manifold related by twisting (C, g). It follows that for any two boundary inequivalent diffeomorphisms g and h, there is a 4–manifold X and an embedding $e: C \hookrightarrow X$ such that X^e_g is not diffeomorphic to X^e_h. Akbulut has made a similar observation.

Boundary sums of cork embeddings

Given any pair of embeddings $e_i: C_i \hookrightarrow X$ (for $i = 1, 2$) of corks (C_i, g_i) with disjoint images $\overline{C}_i = e_i(C_i)$ and induced boundary maps $\overline{g}_i: \partial\overline{C}_i \to \partial\overline{C}_i$, both twists can be performed simultaneously to produce the 4–manifold

$$X^{e_1 e_2}_{g_1 g_2} = (X - \operatorname{int}(\overline{C}_1 \sqcup \overline{C}_2)) \cup_{\overline{g}_1 \sqcup \overline{g}_2} (\overline{C}_1 \sqcup \overline{C}_2).$$

Alternatively, $\overline{C}_1$ and $\overline{C}_2$ can be joined by an embedded 1–handle in X, the thickening of an arc α in $X - \operatorname{int}(\overline{C}_1 \sqcup \overline{C}_2)$ from $\overline{C}_1$ to $\overline{C}_2$. The result is an embedding $e_1 \natural e_2$ of the single cork $(C_1, g_1) \natural (C_2, g_2) = (C_1 \natural C_2, g_1 \sharp g_2)$ (where, as noted above, the map $g_1 \sharp g_2$ is only defined up to boundary equivalence unless the g_i are special involutions) whose cork twist is independent of α. Indeed, it is readily seen that the single cork twist $X^{e_1 \natural e_2}_{g_1 \sharp g_2}$ is diffeomorphic to the pair of cork twists $X^{e_1 e_2}_{g_1 g_2}$.

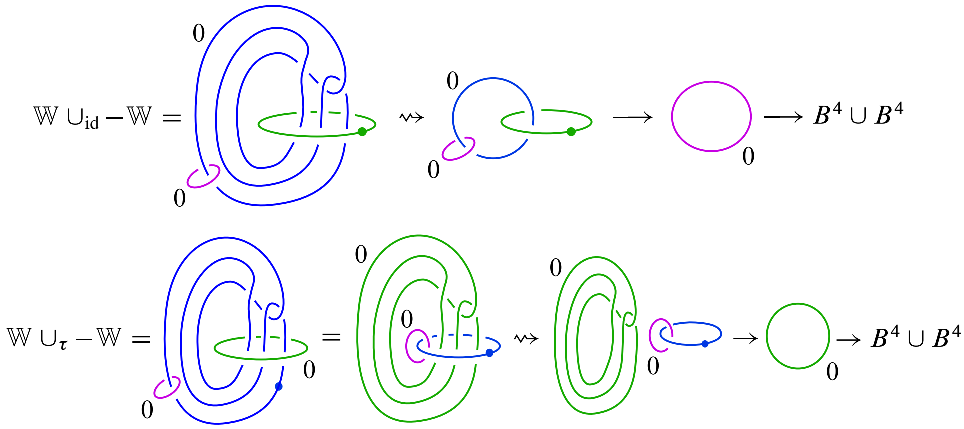

Figure 2: Trivial embedding of the Mazur cork in S^4

This process can be iterated to construct the *multiple cork twist* $X^{e_1\cdots e_n}_{g_1\cdots g_n}$ of a family $e_1, \dots, e_n$ of disjoint embeddings of corks $(C_1, g_1), \dots, (C_n, g_n)$ in X, or a single cork twist $X^{e_1 \natural\cdots\natural e_n}_{g_1 \sharp\cdots\sharp g_n}$ of an embedding of the boundary sum of the (C_i, g_i). Both twists produce the same smooth 4–manifold. This construction will play a key role in what follows.

Trivial cork embeddings

Most explicit corks (C, g) in the literature can be shown to have trivial embeddings in the 4–ball, and thus in every 4–manifold. In particular, it suffices to prove that the double $C \cup_{\mathrm{id}} -C$ and twisted double $C \cup_g -C$ are both diffeomorphic to the 4–sphere, often accomplished by an elementary Kirby calculus argument; cf Akbulut and Yasui [5, Section 2.6]. This is illustrated for the Mazur cork $(\mathbb{W}, \tau)$ in Figure 2, where the squiggly and straight arrows represent handle slides and cancellations, respectively, and as usual, the 3 and 4–handles are not drawn.

Equivariant corks

If G is a subgroup of the diffeomorphism group of ∂C with (C, g) nontrivial for all $g \neq 1$ in G, then (C, G) is called a *G–cork*. For cyclic G of finite order n, we refer to the corks (C, g) for generators g of G as *n–corks*. All explicit corks that have appeared in the literature prior to [19] are special 2–corks; recently, Gompf [13; 14] has shown how to construct $\mathbb{Z}$–corks.

There is a more general notion, which we call a *weakly equivariant cork*, in which the group G is a subgroup of the mapping class group of the boundary, ie the group of isotopy classes of diffeomorphisms. In this situation, it is more appropriate to use the relation of isotopy, rather than boundary equivalence, because the subgroup of diffeomorphisms of the boundary that extend across the cork need not be normal. Hence the set of boundary equivalent diffeomorphisms does not in general form a group in any natural way. In the last section, we give a construction of weakly equivariant corks for many groups G that are not subgroups of $\mathrm{SO}(4)$, and in fact that do not act effectively on any homology 3–sphere.

In general, if C is a cork with an effective G–action on ∂C, then an embedding $e\colon C \hookrightarrow X$ will be said to be *G–effective* if $X^e_{g_1}$ and $X^e_{g_2}$ are smoothly distinct for any $g_1 \neq g_2$ in G. Thus the existence of such embeddings shows that (C, G) is a G–cork. In this case, one has a G–action on the set of 4–manifolds $\{X^e_g \mid g \in G\}$ in the sense that $(X^e_{g_1})^{\overline{e}}_{g_2} = X^e_{g_1 g_2}$ for any two elements $g_1, g_2 \in G$, where $\overline{e}\colon C \to X^e_{g_1}$ is the obvious embedding induced by e.

For the reader's convenience, we repeat the statement of our main result:

Theorem A *There exist G–corks for any finite subgroup G of* $\mathrm{SO}(4)$.

Addenda (1) The proof will show that if $|G| = n$, then the boundary sum $\natural_{n^2}(\mathbb{W}, \tau)$ of n^2 copies of the Mazur cork can be given a G–cork structure that has G–effective embeddings in any blown-up elliptic surface $E(2k) \# m\overline{\mathbb{C}\mathrm{P}}^2$ for $k, m \geq n(n-1)/2$.

(2) More generally, if G is any finite group that acts effectively on the boundary of a compact, contractible submanifold of $\mathbb{R}^4$, then essentially the same proof shows that there is a G–cork with an effective embedding into a closed manifold; Theorem C can then be used to construct such corks with hyperbolic boundary.

2 Construction of equivariant corks

Our proof of Theorem A relies on the existence of certain embeddings e_i of the Mazur cork $(\mathbb{W}, \tau)$ in the blown-up Kummer surface

$$\mathbb{E} := E(2) \# \overline{\mathbb{C}\mathrm{P}}^2.$$

Here $E(2)$ is the minimal elliptic surface of Euler characteristic 24 (or Kummer surface; see for example [15]). The key input from Seiberg–Witten theory is the count of the number of basic classes in the associated cork twists $\mathbb{E}^{e_i}_\tau$.

Definition 2.1 Let X be a smooth, closed, simply connected 4–manifold. If $b_2^+(X)$ is odd and greater than 1, then $\mathcal{N}(X)$ will denote the number of Seiberg–Witten basic classes of X, and otherwise $\mathcal{N}(X) = 0$. For example, $\mathcal{N}(\mathbb{E}) = 2$ (the basic classes are $\pm\overline{\mathbb{C}\mathrm{P}}^1$).

Akbulut [1] established the nontriviality of $(\mathbb{W},\tau)$ by constructing a nontrivial embedding $e_0\colon \mathbb{W}\hookrightarrow\mathbb{E}$ with reducible cork twist $\mathbb{E}^{e_0}_\tau\cong 3\mathbb{C}\mathrm{P}^2\,\#\,20\overline{\mathbb{C}\mathrm{P}}^2$, so in particular, $\mathcal{N}(\mathbb{E}^{e_0}_\tau)=0$. It was later observed [7] that such an embedding could be chosen with image in the complement $\mathbb{E}^\bullet$ of a *nucleus* in $\mathbb{E}$; see [12].

More recent work of Akbulut and Yasui [4] shows that $(\mathbb{W},\tau)$ has another nontrivial embedding $e_2\colon \mathbb{W}\hookrightarrow\mathbb{E}^\bullet$ with $\mathcal{N}(\mathbb{E}^{e_2}_\tau)\neq 0$. The nontriviality of e_2 was proved by showing that $\mathbb{E}^{e_2}_\tau$ results from a rational blow-down of $\mathbb{E}$ [10], leaving $\mathcal{N}$ unchanged, followed by an honest blow-up, doubling $\mathcal{N}$, so $\mathcal{N}(\mathbb{E}^{e_2}_\tau)=4$. (In particular, this follows from Theorem 4.1 for $p=2$, Proposition 5.1 for $n=1$ and $p_1=2$, and Lemma 6.6 in [4].)

As noted in the last section, $(\mathbb{W},\tau)$ also embeds *trivially* into any 4–manifold. Choose one such embedding $e_1\colon \mathbb{W}\hookrightarrow\mathbb{E}^\bullet$. Thus e_0, e_1 and e_2 are numbered so that $\mathcal{N}(\mathbb{E}^{e_i}_\tau)=i\mathcal{N}(\mathbb{E})$. Only e_1 and e_2 are needed to prove the following key result, which is a strengthening of an analogous *noncompact* embedding theorem of Akbulut and Yasui [5, Theorem 1.5].

Lemma 2.2 *For each $n>0$, there exists a 2–cork $(\mathbb{S},\sigma)$ that has n disjoint embeddings $s_1,\dots,s_n$ in some closed 4–manifold X, with distinct cork twists*

$$X^{s_1}_\sigma\cong X,\ X^{s_2}_\sigma,\dots,X^{s_n}_\sigma.$$

For example, the boundary sum $(\mathbb{S},\sigma)=\natural_n(\mathbb{W},\tau)$ has n such embeddings in the blown-up elliptic surface $X=E(2k)\,\#\,m\overline{\mathbb{C}\mathrm{P}}^2$ for any $k,m\ge n(n-1)/2$.

Proof It suffices to prove the last statement. First consider the case $k=m=n^2$, and view $X=E(2n^2)\,\#\,n^2\overline{\mathbb{C}\mathrm{P}}^2$ as the fiber sum of n^2 copies of the blown-up Kummer surface $\mathbb{E}=E(2)\,\#\,\overline{\mathbb{C}\mathrm{P}}^2$ along regular torus fibers in a chosen nucleus. Denote the copies of $\mathbb{E}$ by $\mathbb{E}_{ij}$ for $1\le i,j\le n$. Choose an embedding e_{ij} of $(\mathbb{W},\tau)$ in each summand $\mathbb{E}^\bullet_{ij}$, with $e_{ij}=e_1$ if $i\le j$ and $e_{ij}=e_2$ if $i>j$. For $1\le i\le n$, let s_i be the boundary sum $e_{i1}\natural\cdots\natural e_{in}$ of all the embeddings in the "i^{th} row". Then the s_i are distinct embeddings of $(\mathbb{S},\sigma)=\natural_n(\mathbb{W},\tau)$ and can be chosen with disjoint images by choosing the 1–handles that join the summands to be disjoint. Furthermore, s_i has $i-1$ nontrivial summands and $n-i+1$ trivial ones, and so $\mathcal{N}(X^{s_i}_\sigma)=2^{i-1}\mathcal{N}(X)$. Since $\mathcal{N}(X)\neq 0$, the $X^{s_i}_\sigma$ are pairwise distinct.

Of course, one can be more efficient by using only the "nontrivial" copies of $\mathbb{E}$, ie $\mathbb{E}_{ij}$ for $i>j$, and putting all the trivial embeddings of the Mazur cork inside one of these. This handles the smallest case $k=m=n(n-1)/2$, and the fiber sum and blow-up formulas for Seiberg–Witten invariants show that k and m can be increased at will. □

Proof of Theorem A

Given a finite subgroup G of SO(4) of order n, apply Lemma 2.2 to produce n disjoint embeddings s_g of a cork $(\mathbb{S},\sigma)$ in a closed 4–manifold X, indexed by the elements of G, with distinct cork twists $X_\sigma^{s_g}$. Using these cork embeddings, we construct a G–cork $(\mathbb{T},G)$ and a G–effective embedding $t\colon \mathbb{T}\to X$, as follows.

The underlying contractible manifold $\mathbb{T}$ is the boundary sum $\natural_n\mathbb{S}$ of n copies of $\mathbb{S}$. To define the G action on $\partial\mathbb{T}$, it is convenient to represent $\mathbb{T}$ as a cork twist on a diffeomorphic copy $\overline{\mathbb{T}}$ of itself that supports a natural G–action, namely the equivariant boundary sum

$$\overline{\mathbb{T}} = B^4 \natural (G\times\mathbb{S})$$

taken along a *principal* orbit $\{b_g \mid g\in G\}$ of the linear G action on ∂B^4, where G acts on $G\times\mathbb{S}$ by left multiplication on the first factor and trivially on the second. In other words, $\overline{\mathbb{T}}$ is obtained from a disjoint union of the 4–ball and n copies $\mathbb{S}_g$ of $\mathbb{S}$ (indexed by $g\in G$) by adding 1–handles joining $b_g\in\partial B^4$ to $x_g\in\partial\mathbb{S}_g$, where the $x_g\in\partial\mathbb{S}_g$ correspond to a chosen point $x\in\partial\mathbb{S}$. The G action is linear on B^4, and permutes the copies of $\mathbb{S}_g$ by left multiplication on the subscript (since the boundary sum is along a principal orbit).

Now the embeddings s_g of $\mathbb{S}$ can be used to define an embedding

$$\bar{t}\colon \overline{\mathbb{T}}\hookrightarrow X$$

by identifying $\mathbb{S}_g$ with the image $s_g(\mathbb{S})$ in X, B^4 with a small 4–ball B disjoint from the $\mathbb{S}_g$, and the 1–handles joining B^4 to the $\mathbb{S}_g$ with embedded 1–handles.

To obtain $\mathbb{T}$, we twist a shrunken copy of the cork $1\times\mathbb{S}$ in $\overline{\mathbb{T}}$. To make this precise, recall that $\overline{\mathbb{T}}$ contains n copies $\mathbb{S}_g = g\times\mathbb{S}$ of $\mathbb{S}$, the images of the embeddings $e_g\colon \mathbb{S}\hookrightarrow\overline{\mathbb{T}}$ sending x to (g,x). Consider an embedding $s\colon\mathbb{S}\hookrightarrow\mathbb{S}$ that shrinks $\mathbb{S}$ inside itself; that is, s is the identity off of a boundary collar $\partial\mathbb{S}\times[0,1)$, and maps (x,t) to $(x,(t+1)/2)$ inside the collar. Then $e=e_1\circ s$ embeds $\mathbb{S}$ onto a shrunken copy of $\mathbb{S}_1$. We define $\mathbb{T}$ to be the cork twist associated with this embedding:

$$\mathbb{T} = \overline{\mathbb{T}}{}_\sigma^e.$$

Since the $\partial\mathbb{T}=\partial\overline{\mathbb{T}}$, there is still a G–action on $\partial\mathbb{T}$, and this defines our cork $(\mathbb{T},G)$. Note that $\mathbb{T}$ is actually diffeomorphic to $\overline{\mathbb{T}}$, and thus to $\natural_n\mathbb{S}$, since $\natural$ is a well defined operation, but for our purposes it is most convenient to describe $\mathbb{T}$ as a cork twist of $\overline{\mathbb{T}}$.

Now observe that the embedding $\bar{t}\colon\overline{\mathbb{T}}\hookrightarrow X$ above induces an embedding

$$t\colon \mathbb{T}\hookrightarrow X_\sigma^{s_1}$$

since $\mathbb{T} = \overline{\mathbb{T}}^e_\sigma$. Furthermore, twisting this embedding of $\mathbb{T}$ by an element $g \in G$ just transfers the cork twist from $\mathbb{S}_1$ to $\mathbb{S}_g$; that is,

$$(X^{s_1}_\sigma)^t_g = X^{s_g}_\sigma.$$

Since the smooth 4–manifolds $X^{s_g}_\sigma$ are distinct for $g \in G$, this shows that t is a G–effective embedding, and so $(\mathbb{T}, G)$ is a G–cork. This completes the proof of Theorem A. $\square$

Remark Even in the case $G = \mathbb{Z}_2$ this result can give something new. Applying the construction from Theorem A to the free $\mathbb{Z}_2$ action on S^3 extended across B^4 we get a 2–cork with free action on the boundary.

Proof of the addenda to Theorem A

The first addendum to the theorem follows from this proof by using $(\mathbb{S}, \sigma) = \natural_n(\mathbb{W}, \tau)$ and $X = E(2k)\,\#\, m\overline{\mathbb{C}\mathrm{P}}^2$, as provided by the lemma. Note that in the proof, $X^{s_1}_\sigma$ is diffeomorphic to X since s_1 is a trivial cork embedding, so t can be viewed as an embedding of $\natural_{n^2}\mathbb{W} \hookrightarrow X$.

With regard to the second addendum, if a finite group G acts on a compact contractible submanifold of $\mathbb{R}^4$, we may repeat the argument replacing B^4 by the contractible submanifold to produce a G–cork $\mathbb{T}$. To build a G–cork with hyperbolic boundary, let $\mathbb{U}$ be an invertible cobordism from $\partial\mathbb{T}$ to a hyperbolic 3–manifold M with inverse $\mathbb{V}$ as given by Theorem C. Then

$$\mathbb{T} \cup_{\partial\mathbb{T}} \mathbb{U} \ \subset\ \mathbb{T} \cup_{\partial\mathbb{T}} \mathbb{U} \cup_M \mathbb{V} \ \cong\ \mathbb{T},$$

and $\mathbb{T} \cup_{\partial\mathbb{T}} \mathbb{U}$ inherits a G action so twisting it via g has the same effect as twisting $\mathbb{T}$ since g extends across $\mathbb{V}$. $\square$

Remark From the construction, we see that our G–corks are boundary-connected sums of Stein manifolds, and hence are Stein. In contrast to the argument in [19], this fact does not play any role in our verification that our corks are effective.

3 Weakly equivariant corks

In this section, we construct examples of weakly equivariant corks for certain finite groups that are not subgroups of $\mathrm{SO}(4)$. In fact, these groups cannot act on any homology sphere, so there are no corresponding equivariant corks. This will prove Theorem B.

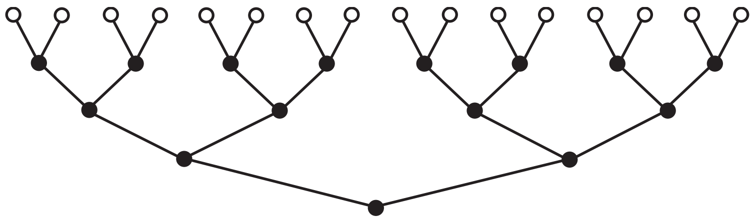

Figure 3: A weak C_2^4–cork

Proof of Theorem B

Fix $n \geq 4$, and let $G = C_2^n$, the product of n copies of the cyclic group C_2. It is known that G does not act effectively on any homology 3–sphere [20, Proposition 3]. In this proof, we show how to construct a nontrivial weak G–cork $\mathbb{V}$.

Apply Lemma 2.2 to get a 2–cork $(\mathbb{S}, \sigma)$ with 2^n inequivalent embeddings s_g (for $g \in G$) in some 4–manifold X, meaning their cork twists $X_\sigma^{s_g}$ are 2^n distinct smooth 4–manifolds. For convenience, assume that $X_\sigma^1 \cong X$. For example, $\mathbb{S}$ could be the boundary sum of 2^n Mazur corks, with $X = E(2^{2n+1}) \# 2^{2n}\overline{\mathbb{C}\mathrm{P}}^2$; see the proof of Lemma 2.2.

As in the proof of Theorem A, we will define the cork $\mathbb{V}$ to be a suitable cork twist of a diffeomorphic copy $\overline{\mathbb{V}}$ of $\mathbb{V}$. To define $\overline{\mathbb{V}}$, consider a full binary tree T of height n, built from the bottom up, as shown in Figure 3 for the case $n = 4$. Thus T has one vertex at the root, two at the first level, four at the second level, etc. At the top there are 2^n vertices which can be indexed in a natural way by the elements of G (as explained below). To get $\overline{\mathbb{V}}$, replace the black dots by 4–balls, the white dots by copies of the cork $\mathbb{S}$ (referred to as the *leaves* of the cork) and the edges by 1–handles. Also choose an *equatorial* 3*–disk* D for each black 4–ball B that separates the 1–handle attached to B below D (if any) from the two attached above; D splits $\overline{\mathbb{V}}$ into two components with closures D^+ (locally above D) and D^- (locally below D).

Let $\tau_0, \ldots, \tau_{n-1}$ denote the generators of the C_2 factors in $G = C_2^n$, and let τ_k act on $\overline{\mathbb{V}}$ by performing half Dehn twists on all the level k equatorial 3–disks. Here a *half Dehn twist* about such a disk D is the diffeomorphism of $\overline{\mathbb{V}}$ that leaves D^- fixed, sends a collar neighborhood $D \times [0, \pi]$ of D in D^+ to itself by the map $(x, \theta) \mapsto (\mathrm{rot}_\theta(x), \theta)$, and sends the rest of D^+ to itself in the obvious way, reversing the order of the leaves above D. Thus, for example, τ_0 reverses the order of all the leaves at the top, τ_1 independently reverses the orders of the first and second halves of the leaves, and so forth. Note that a *full Dehn twist* of a 4–manifold X can be defined in a similar way about *any* 3–disk D that is either properly embedded or embedded in ∂X. In either case one uses a collar $D \times [0, 2\pi]$ that restricts to a collar of ∂D in ∂X,

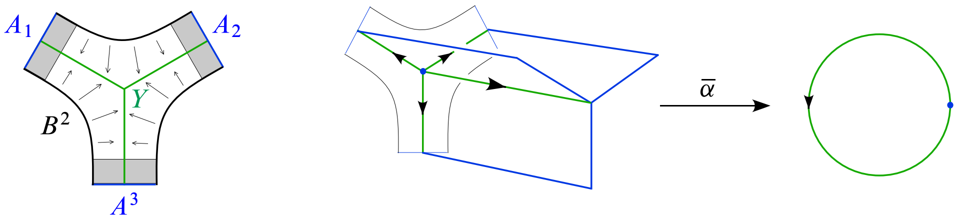

Figure 4: $A_i, Y \subset B^2$ when $n=3$ (left) and the map $\overline{\alpha}\colon Y \times I \to S^1$ (right)

lying to the outside of D when $D \subset \partial X$; the shaded region in Figure 4 (left) illustrates how the collar meets the boundary in this latter case.

Now observe that τ_k is of order 2 in the mapping class group of $\overline{\mathbb{V}}$. This is clear for τ_0, since τ_0^2 is a full Dehn twist about the equatorial disk D_0 that untwists by an isotopy over the 4–ball D_0^- below it, and in general we claim that τ_k^2 is isotopic to τ_{k-1}^2. Indeed, the portion of $\overline{\mathbb{V}}$ lying between level $k-1$ and level k is a union of 4–balls, each containing exactly three equatorial 3–disks in its boundary. Thus it suffices to prove that a full twist about two of these disks is isotopic to a full twist about the third. Since $\pi_1\,\mathrm{SO}(3) = \mathbb{Z}_2$, this is a consequence of the following elementary fact (cf [16, page 190]):

Lemma 3.1 *The composition δ of Dehn twists of a 4–ball B about any finite number of disjoint 3–disks $D_1, \dots, D_n$ in its boundary is isotopic to the identity, leaving the D_i fixed.*

Proof of the Lemma View $B = B^2 \times B^2$ and $D_i = A_i \times B^2$, where the A_i are disjoint arcs in ∂B^2. Let $r\colon B^2 \to Y$ be a deformation retraction that collapses each A_i to its midpoint a_i, where Y is the cone $0 * \{a_1, \dots, a_n\}$. Pictures of the arcs A_i and the graph Y in B^2, and an indication of the retraction r, are shown in Figure 4 (left) for the case $n=3$, with collars corresponding to the shaded regions.

With this parametrization $B = B^2 \times B^2$, we can take

$$\delta(x,y) = (x, \mathrm{rot}_{\alpha(r(x))}(y)),$$

where $\alpha\colon Y \to S^1$ is a map of degree one on each edge $e_i = 0 * a_i$ of Y. Evidently, α extends to a map $\overline{\alpha}\colon Y \times I \to S^1$ that has degree one on each edge $e_i \times 0$ and $0 \times I$, and is constant on each edge $e_i \times 1$ and $a_i \times I$; see Figure 4 (right). This defines the desired isotopy δ_t from $\delta = \delta_0$ to the identity, rel the D_i, given by $\delta_t(x,y) = (x, \mathrm{rot}_{\overline{\alpha}(r(x),t)}(y))$. $\square$

Continuing with the proof of Theorem B, it is clear that the action of the τ_k extends to an embedding of G in the mapping class group of $\overline{\mathbb{V}}$, and that distinct elements of G

carry the first leaf to distinct leaves. This gives a natural way to index the leaves of $\overline{\mathbb{V}}$ by the elements $g \in G$, according to where g carries the first leaf. Thus, for example, the last leaf is indexed by τ_0, while the $(2^{n-1})^{\text{st}}$ leaf is indexed by τ_1.

Now let $\mathbb{V}$ be the cork twist of $\overline{\mathbb{V}}$ along (a shrunken copy of) the first leaf. Then $\partial\mathbb{V}$ is naturally identified with $\partial\overline{\mathbb{V}}$, so there is an induced embedding of G in the mapping class group of $\partial\mathbb{V}$. To see that this defines a weak G–cork structure on $\mathbb{V}$, just choose an embedding $e\colon \mathbb{V} \hookrightarrow X$ that restricts to the embeddings s_g (for $g \in G$) on the leaves of $\mathbb{V}$. Then $X_g^e = X_\sigma^{s_g}$, and so X_g^e and X_h^e are not diffeomorphic unless $g = h$. □

References

[1] **S Akbulut**, *A fake compact contractible* 4*–manifold*, J. Differential Geom. 33 (1991) 335–356 MR

[2] **S Akbulut**, **D Ruberman**, *Absolutely exotic compact* 4*–manifolds*, Comment. Math. Helv. 91 (2016) 1–19 MR

[3] **S Akbulut**, **K Yasui**, *Corks, plugs and exotic structures*, J. Gökova Geom. Topol. 2 (2008) 40–82 MR

[4] **S Akbulut**, **K Yasui**, *Knotting corks*, J. Topol. 2 (2009) 823–839 MR

[5] **S Akbulut**, **K Yasui**, *Stein* 4*–manifolds and corks*, J. Gökova Geom. Topol. 6 (2012) 58–79 MR

[6] **D Auckly**, **H J Kim**, **P Melvin**, **D Ruberman**, *From tangles to equivariant hyperbolic corks*, AMS abstract (2016) `http://tinyurl.com/jmm2181abs1116-57-1143`

[7] **Ž Bižaca**, **R E Gompf**, *Elliptic surfaces and some simple exotic* $\mathbf{R}^4$*'s*, J. Differential Geom. 43 (1996) 458–504 MR

[8] **M Culler**, **N M Dunfield**, **J R Weeks**, *SnapPy, a computer program for studying the topology of* 3*–manifolds* `http://snappy.computop.org`

[9] **C L Curtis**, **M H Freedman**, **W C Hsiang**, **R Stong**, *A decomposition theorem for* h*–cobordant smooth simply-connected compact* 4*–manifolds*, Invent. Math. 123 (1996) 343–348 MR

[10] **R Fintushel**, **R J Stern**, *Rational blowdowns of smooth* 4*–manifolds*, J. Differential Geom. 46 (1997) 181–235 MR

[11] **M H Freedman**, *The topology of four-dimensional manifolds*, J. Differential Geom. 17 (1982) 357–453 MR

[12] **R E Gompf**, *Nuclei of elliptic surfaces*, Topology 30 (1991) 479–511 MR

[13] **R E Gompf**, *Infinite order corks*, preprint (2016) arXiv To appear in Geom. Topol.

[14] **R E Gompf**, *Infinite order corks via handle diagrams*, preprint (2016) arXiv

[15] **R E Gompf**, **A I Stipsicz**, *4–manifolds and Kirby calculus*, Graduate Studies in Mathematics 20, Amer. Math. Soc., Providence, RI (1999) MR

[16] **H Hendriks**, *Applications de la théorie d'obstruction en dimension 3*, Bull. Soc. Math. France Mém. 53 (1977) 81–196 MR

[17] **R Matveyev**, *A decomposition of smooth simply-connected h–cobordant 4–manifolds*, J. Differential Geom. 44 (1996) 571–582 MR

[18] **B Mazur**, *A note on some contractible 4–manifolds*, Ann. of Math. 73 (1961) 221–228 MR

[19] **M Tange**, *Finite order corks*, preprint (2016) arXiv

[20] **B Zimmermann**, *On the classification of finite groups acting on homology 3–spheres*, Pacific J. Math. 217 (2004) 387–395 MR

Department of Mathematics, Kansas State University
Manhattan, KS 66506, United States

Department of Mathematical Sciences, Seoul National University
Seoul 151-747, South Korea

Department of Mathematics, Bryn Mawr College
Bryn Mawr, PA 19010, United States

Department of Mathematics, Brandeis University, MS 050
Waltham, MA 02454, United States

dav@math.ksu.edu, heejungorama@gmail.com, pmelvin@brynmawr.edu, ruberman@brandeis.edu

www.math.ksu.edu/~dav/, www.math.uga.edu/directory/hee-jung-kim, www.brynmawr.edu/math/people/melvin/, people.brandeis.edu/~ruberman/

Received: 20 April 2016 Revised: 14 September 2016

Algebraic & Geometric Topology 17 (2017) 1785–1811

A homology-valued invariant for trivalent fatgraph spines

YUSUKE KUNO

We introduce an invariant for trivalent fatgraph spines of a once-bordered surface, which takes values in the first homology of the surface. This invariant is a secondary object coming from two 1–cocycles on the dual fatgraph complex, one introduced by Morita and Penner in 2008, and the other by Penner, Turaev and the author in 2013. We present an explicit formula for this invariant and investigate its properties. We also show that the mod 2 reduction of the invariant is the difference of two naturally defined spin structures on the surface.

20F34, 32G15, 57N05

1 Introduction

Let $\Sigma_{g,1}$ be a once-bordered C^∞–surface of genus $g > 0$, and let $\mathcal{M}_{g,1}$ be the mapping class group of $\Sigma_{g,1}$ relative to the boundary. It is known that the Teichmüller space $\mathcal{T}(\Sigma_{g,1})$ of $\Sigma_{g,1}$ has an $\mathcal{M}_{g,1}$–equivariant ideal simplicial decomposition; see Penner [21]. Taking its dual, one obtains a contractible CW complex $\widehat{\mathcal{G}}(\Sigma_{g,1})$ on which $\mathcal{M}_{g,1}$ acts freely and properly discontinuously. This CW complex is called the dual fatgraph complex of $\Sigma_{g,1}$, since its cells are indexed by fatgraph spines of $\Sigma_{g,1}$, which are graphs embedded in the surface satisfying some conditions. Each 0–cell of $\widehat{\mathcal{G}}(\Sigma_{g,1})$ corresponds to a trivalent fatgraph spine, and by contracting nonloop edges we obtain higher-dimensional cells. In particular, each oriented 1–cell of $\widehat{\mathcal{G}}(\Sigma_{g,1})$ corresponds to a flip (or a Whitehead move) between trivalent fatgraph spines of $\Sigma_{g,1}$.

This combinatorial structure of the Teichmüller space has a number of applications to the cohomology of the mapping class group and the moduli space of Riemann surfaces. See eg Harer [4; 5], Harer and Zagier [6], Penner [20] and Kontsevich [12].

Recently, mainly motivated by the theory of the Johnson homomorphisms (see Johnson [7; 9] and Morita [16]), several authors considered 1–cocycles on $\widehat{\mathcal{G}}(\Sigma_{g,1})$ with coefficients in various $\mathcal{M}_{g,1}$–modules. In 2008, Morita and Penner [18] first gave such a 1–cocycle $j \in Z^1(\widehat{\mathcal{G}}(\Sigma_{g,1}); \Lambda^3 H)$, where $\Lambda^3 H$ is the third exterior power of the first homology group $H = H_1(\Sigma_{g,1}; \mathbb{Z})$. (In fact, they worked with a once-punctured surface, but their construction works for $\Sigma_{g,1}$ as well.) Being a 1–cocycle on $\widehat{\mathcal{G}}(\Sigma_{g,1})$, the cocycle j associates an element of $\Lambda^3 H$ to each flip. Fixing a trivalent

Published: 17 July 2017 DOI: 10.2140/agt.2017.17.1785

fatgraph spine of $\Sigma_{g,1}$, one obtains from j a twisted 1–cocycle on $\mathcal{M}_{g,1}$. Morita and Penner proved that its cohomology class in $H^1(\mathcal{M}_{g,1};\Lambda^3 H)$ is six times the extended first Johnson homomorphism $\widetilde{k}$ discovered by Morita [17]. Similar constructions are also considered by Bene, Kawazumi and Penner [1] for the second and higher Johnson homomorphisms, by Massuyeau [14] for Morita's refinement [16] of the higher Johnson homomorphisms, and by Kuno, Penner and Turaev [13] for the Earle class $k \in H^1(\mathcal{M}_{g,1};H)$.

We emphasize that these cocycles on $\widehat{\mathcal{G}}(\Sigma_{g,1})$ are all explicit and simple. In this way, the Johnson homomorphisms and related objects extend *canonically* to the *Ptolemy groupoid*, the combinatorial fundamental path groupoid of $\widehat{\mathcal{G}}(\Sigma_{g,1})$; see Bene, Kawazumi and Penner [1].

It is interesting that there are many ways of constructing cocycle representatives for the cohomology classes such as $\widetilde{k}$ and k, and that each construction reflects its own viewpoint for studying the mapping class group. It can happen that two cocycles constructed differently give the same cohomology class. In such a case, it is quite natural to compare these cocycles and to expect a secondary object in the background.

We will compare the Morita–Penner cocycle j and the cocycle $m \in Z^1(\widehat{\mathcal{G}}(\Sigma_{g,1});H)$ which is related to k and considered in Kuno, Penner and Turaev [13]. Contracting the coefficients by using the intersection pairing on H, one has a natural homomorphism

$$C\colon Z^1(\widehat{\mathcal{G}}(\Sigma_{g,1});\Lambda^3 H) \to Z^1(\widehat{\mathcal{G}}(\Sigma_{g,1});H).$$

Let $j' = C \circ j$. It turns out that there is an $\mathcal{M}_{g,1}$–equivariant 0–cochain $\xi \in C^0(\widehat{\mathcal{G}}(\Sigma_{g,1});H)$ such that $2j' - m = \delta\xi$ (Proposition 3.1). The 0–cochain ξ assigns an element $\xi_G \in H$ to each trivalent fatgraph spine $G \subset \Sigma_{g,1}$.

We will study the secondary object ξ_G as an H–valued invariant for trivalent fatgraph spines $G \subset \Sigma_{g,1}$. First of all, Theorem 3.4 gives an explicit formula for ξ_G. Based on this formula, we show in Theorem 5.2 that ξ_G is nontrivial. At the present moment, we do not have a full understanding of the topological meaning of the invariant ξ_G. In Theorem 6.7, we give a partial result in this direction by relating the mod 2 reduction of ξ_G to two naturally defined spin structures on $\Sigma_{g,1}$. It would be interesting to seek for or to find an obstruction to an extension of ξ_G to fatgraph spines which are not necessarily trivalent. In view of the fact that trivalent fatgraph spines correspond to maximal-dimensional simplices of the ideal simplicial decomposition of $\mathcal{T}(\Sigma_{g,1})$, this is related to finding a $\mathcal{M}_{g,1}$–equivariant function on $\mathcal{T}(\Sigma_{g,1})$ which takes values in $H \otimes_{\mathbb{Z}} \mathbb{R} = H_1(\Sigma_{g,1};\mathbb{R})$.

This paper is organized as follows. In Section 2, we first review the dual fatgraph complex and in particular describe its 2–skeleton. Then we recall the 1–cocycles j

from [18] and m from [13]. Also, we correct an error in [13] about the evaluation of m. In Section 3, we show the existence and uniqueness of ξ, and then present an explicit formula for ξ_G (Theorem 3.4). In Section 4, we show a certain gluing formula for ξ_G, and then the behavior of ξ_G under a special kind of flip. The latter result makes it possible to define ξ_G for a trivalent fatgraph spine G of a *punctured* surface. In Section 5, we discuss the nontriviality of ξ_G based on its explicit formula. In Section 6, we construct two spin structures on $\Sigma_{g,1}$ from each trivalent fatgraph spine $G \subset \Sigma_{g,1}$. Then we prove that their difference coincides with the mod 2 reduction of ξ_G. Along the way we give a combinatorial description of spin structures on $\Sigma_{g,1}$ (Theorem 6.2), which seems to be new. In the appendix, we consider another spin structure coming from a naturally defined nonsingular vector field on $\Sigma_{g,1}$.

Acknowledgements The author would like to thank Robert Penner for helpful remarks on a description of spin structures on $\Sigma_{g,1}$ in Section 6, Gwénaël Massuyeau for communicating to him the construction of the vector field $\mathcal{X}_G$ in the appendix, and Vladimir Turaev and Nariya Kawazumi for valuable comments to a draft of this paper. This work is supported by JSPS KAKENHI (no. 26800044).

2 Fatgraph complex and cocycles

We fix some notation about graphs. By a graph we mean a finite CW complex of dimension one. For a graph G, we denote by $V(G)$ the set of vertices of G, by $E(G)$ the set of edges of G, and by $E^{\mathrm{ori}}(G)$ the set of oriented edges of G. For $v \in V(G)$, we denote by $E^{\mathrm{ori}}_v(G)$ the set of oriented edges pointing toward v. The number of elements of $E^{\mathrm{ori}}_v(G)$ is called the *valency* of v. For $e \in E^{\mathrm{ori}}(G)$, we denote by $\overline{e} \in E^{\mathrm{ori}}(G)$ the edge e with reversed orientation. A *fatgraph* is a graph G endowed with a cyclic ordering to $E^{\mathrm{ori}}_v(G)$ about each $v \in V(G)$.

Let $\Sigma_{g,1}$ be a compact connected oriented C^∞–surface of genus $g > 0$ with one boundary component. We fix two distinct points p and q on the boundary $\partial\Sigma_{g,1}$.

Definition 2.1 An embedding $\iota\colon G \hookrightarrow \Sigma_{g,1}$ of a fatgraph G into $\Sigma_{g,1}$ is called a *fatgraph spine* of $\Sigma_{g,1}$ if the following conditions are satisfied:

(1) The map ι is a homotopy equivalence.

(2) For any $v \in V(G)$, the cyclic ordering given to $E^{\mathrm{ori}}_v(G)$ is compatible with the orientation of $\Sigma_{g,1}$.

(3) We have $\iota(G) \cap \partial\Sigma_{g,1} = \{p\}$ and $\iota^{-1}(p)$ is a unique univalent vertex of G. The other vertices have valencies greater than 2.

The unique edge connected to $\iota^{-1}(p)$ is called the *tail* of G. We consider fatgraph spines up to isotopies relative to $\partial\Sigma_{g,1}$. If there is no danger of confusion, we identify G with $\iota(G)$, and write G instead of $\iota\colon G \hookrightarrow \Sigma_{g,1}$. We denote by $V^{\mathrm{int}}(G)$ the set of nonunivalent vertices of G. We say that G is *trivalent* if the valency of any nonunivalent vertex of G is 3.

Fatgraph spines appear naturally in the combinatorial description of the Teichmüller space of a punctured or bordered surface. This was first shown for punctured surfaces by Harer and Mumford [5] and Thurston from the holomorphic point of view based on a work by Strebel [24], and by Penner [19] and Bowditch and Epstein [2] from the point of view of hyperbolic geometry.

In this paper, we work mainly with the once-bordered surface $\Sigma_{g,1}$. For definiteness, let us define the Teichmüller space $\mathcal{T}(\Sigma_{g,1})$ as the space of Riemannian metric on $\Sigma_{g,1}$ of constant Gaussian curvature -1 with geodesic boundary, modulo pullback of the metric by self-diffeomorphisms of $\Sigma_{g,1}$ fixing q which are isotopic to the identity relative to q. Let $\mathcal{M}_{g,1}$ be the mapping class group of $\Sigma_{g,1}$ relative to $\partial\Sigma_{g,1}$. Namely, $\mathcal{M}_{g,1}$ is the group of self-diffeomorphisms of $\Sigma_{g,1}$ fixing the boundary $\partial\Sigma_{g,1}$ pointwise, modulo isotopies fixing $\partial\Sigma_{g,1}$ pointwise. Note that $\mathcal{M}_{g,1}$ is identified with the group of connected components of the group of self-diffeomorphisms of $\Sigma_{g,1}$ fixing q. Then pullback of the metric induces an action of $\mathcal{M}_{g,1}$ on $\mathcal{T}(\Sigma_{g,1})$. This action is known to be free and properly discontinuous.

Theorem 2.2 (Penner [21]) *There is an $\mathcal{M}_{g,1}$–equivariant ideal simplicial decomposition of $\mathcal{T}(\Sigma_{g,1})$ with the following properties:*

- *Each simplex corresponds to a fatgraph spine of $\Sigma_{g,1}$.*
- *The face relation between simplices corresponds to the contraction of a nonloop edge of a fatgraph spine.*

Let $\widehat{\mathcal{G}}(\Sigma_{g,1})$ be the *dual* of this ideal simplicial decomposition. This is an honest CW complex of dimension $4g-2$. We call $\widehat{\mathcal{G}}(\Sigma_{g,1})$ the *dual fatgraph complex* of $\Sigma_{g,1}$. Note that there is a natural cellular action of the mapping class group $\mathcal{M}_{g,1}$ on $\widehat{\mathcal{G}}(\Sigma_{g,1})$. In fact, there is an $\mathcal{M}_{g,1}$–equivariant deformation retract of $\mathcal{T}(\Sigma_{g,1})$ onto $\widehat{\mathcal{G}}(\Sigma_{g,1})$; see [22].

The 2–skeleton of $\widehat{\mathcal{G}}(\Sigma_{g,1})$ is described as follows:

- Each 0–cell corresponds to a trivalent fatgraph spine of $\Sigma_{g,1}$.
- Each 1–cell corresponds to a fatgraph spine G, where G has a unique 4–valent vertex and the other nonunivalent vertices have valency 3.

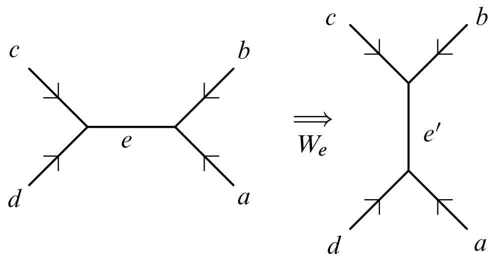

Figure 1: The flip along e applied to G (left) produces G' (right)

- Each *oriented* 1–cell corresponds to a *flip* (or a *Whitehead move*) between trivalent fatgraph spines. Here, if e is a nontail edge of a trivalent fatgraph spine, collapsing e and expanding the resulting 4–valent vertex to the unique distinct direction, one produces another trivalent fatgraph spine. We call this move a flip along e, and denote it by W_e. See Figure 1. If G' is obtained from G by a flip $W = W_e$, we write it as $G \xrightarrow{W} G'$. There is a natural bijection from $E(G)$ to $E(G')$ which restricts to an obvious identification of $E(G)\setminus\{e\}$ with $E(G')\setminus\{e'\}$. For this reason, we often use the same letter for edges of G and G' corresponding to each other via this bijection.
- Each 2–cell corresponds to a fatgraph spine G, where either G has a unique 5–valent vertex and the other nonunivalent vertices have valency 3, or G has two 4–valent vertices and the other nonunivalent vertices have valency 3.

Let G and G' be trivalent fatgraph spines. Since $\widehat{\mathcal{G}}(\Sigma_{g,1})$ is connected, there is a finite sequence of flips

$$G = G_0 \xrightarrow{W_1} G_1 \xrightarrow{W_2} G_2 \xrightarrow{W_3} \cdots \xrightarrow{W_m} G_m = G'$$

from G to G'. This sequence is not uniquely determined, but any two such sequences are related to each other by the following three types of relations among flips:

(1) **Involutivity relation** $W_{e'} \circ W_e = 1$ in the notation of Figure 1.

(2) **Commutativity relation** $W_{e_1} \circ W_{e_2} = W_{e_2} \circ W_{e_1}$ if e_1 and e_2 share no vertices.

(3) **Pentagon relation** $W_{f_4} \circ W_{g_3} \circ W_{f_2} \circ W_{g_1} \circ W_f = 1$ in the notation of Figure 2.

Here, we read composition of flips from right to left. The relations (2) and (3) come from the boundaries of 2–cells of $\widehat{\mathcal{G}}(\Sigma_{g,1})$.

There is a construction of twisted 1–cocycles on the mapping class group using the fatgraph complex appeared first in [18]. Let M be a (left) $\mathcal{M}_{g,1}$–module. By definition, a cellular 1–cochain c on $\widehat{\mathcal{G}}(\Sigma_{g,1})$ with values in M is an assignment of an element

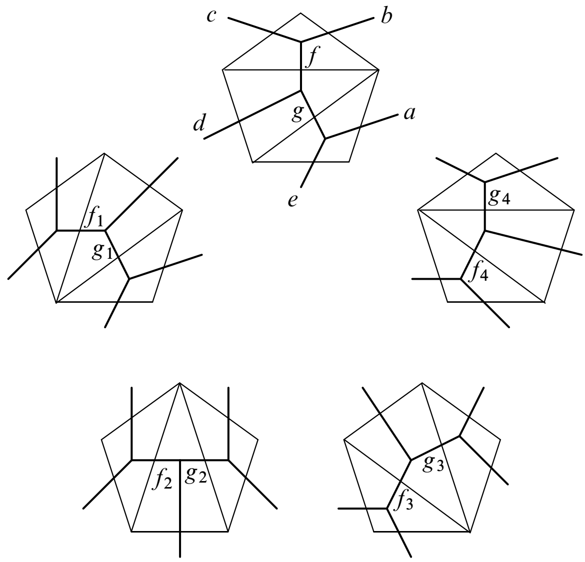

Figure 2: Pentagon relation

of M to each flip W satisfying $c(W_{e'}) = -c(W_e)$ for any pair of flips W_e and $W_{e'}$ as in Figure 1. Such a c is a 1–cocycle if it satisfies both the commutative equation

$$c(W_{e_1}) + c(W_{e_2}) = c(W_{e_2}) + c(W_{e_1}),$$

where e_1 and e_2 are any edges on a trivalent fatgraph spine sharing no vertices, and the pentagon equation

$$c(W_{f_4}) + c(W_{g_3}) + c(W_{f_2}) + c(W_{g_1}) + c(W_f) = 0$$

for any 5–tuple of flips as in Figure 2.

Now we assume that c is a 1–cocycle and is $\mathcal{M}_{g,1}$–equivariant in the sense that $\varphi \cdot c(W) = c(\varphi W)$ for any flip W and $\varphi \in \mathcal{M}_{g,1}$. Fix a trivalent fatgraph spine G. For $\varphi \in \mathcal{M}_{g,1}$, taking a sequence of flips

$$G = G_0 \xrightarrow{W_1} G_1 \xrightarrow{W_2} G_2 \xrightarrow{W_3} \cdots \xrightarrow{W_m} G_m = \varphi(G)$$

from G to $\varphi(G)$, we set

$$c_G(\varphi) := \sum_{i=1}^{m} c(W_i) \in M.$$

Since c is a 1–cocycle, this value does not depend on the choice of the sequence. The map $c_G \colon \mathcal{M}_{g,1} \to M$ is a twisted 1–cocycle. In fact, for $\varphi, \psi \in \mathcal{M}_{g,1}$, take a sequence of flips from G to $\varphi(G)$, and one from G to $\psi(G)$. Then the first sequence followed by application of φ to the second is a sequence of flips from G to $\varphi\psi(G)$. Since c is $\mathcal{M}_{g,1}$–equivariant, we obtain the cocycle condition

$$c_G(\varphi\psi) = c_G(\varphi) + \varphi \cdot c_G(\psi).$$

It is easy to see that the cohomology class $[c_G] \in H^1(\mathcal{M}_{g,1}; M)$ does not depend on the choice of G.

Here we record an elementary fact which will be used later.

Lemma 2.3 *Let M be an $\mathcal{M}_{g,1}$–module and suppose that c is an $\mathcal{M}_{g,1}$–equivariant cellular 1–cocycle on $\widehat{\mathcal{G}}(\Sigma_{g,1})$ with values in M. Then for any trivalent fatgraph spine G and any $\varphi, \psi \in \mathcal{M}_{g,1}$, we have*

$$c_G(\psi) + c_{\psi(G)}(\varphi) = c_G(\varphi) + \varphi \cdot c_G(\psi).$$

Proof Consider a sequence of flips from G to $\psi(G)$ and one from $\psi(G)$ to $\varphi\psi(G)$. The composition of these sequences is a sequence from G to $\varphi\psi(G)$, and thus we obtain $c_G(\varphi\psi) = c_G(\psi) + c_{\psi(G)}(\varphi)$. On the other hand, by the cocycle condition for c_G, we have $c_G(\varphi\psi) = c_G(\varphi) + \varphi \cdot c_G(\psi)$. □

We denote by $H = H_1(\Sigma_{g,1}; \mathbb{Z})$ the first integral homology group of $\Sigma_{g,1}$. Before giving examples of $\mathcal{M}_{g,1}$–equivariant cellular 1–cochains on $\widehat{\mathcal{G}}(\Sigma_{g,1})$, we recall from [18] homology markings for edges of fatgraph spines. Let G be a (not necessarily trivalent) fatgraph spine of $\Sigma_{g,1}$. For $e \in E^{\mathrm{ori}}(G)$, there is an oriented simple loop $\hat{e}$ on $\Sigma_{g,1}$ satisfying the following two conditions:

- The loop $\hat{e}$ intersects G once transversely at the middle point of e.
- The ordered pair of the velocity vectors of $\hat{e}$ and e at their intersection is compatible with the orientation of $\Sigma_{g,1}$.

Since the surface obtained from $\Sigma_{g,1}$ by cutting along G is a disk, the homotopy class of such an $\hat{e}$ is unique. We define $\mu(e) \in H$ to be the homology class of $\hat{e}$ and call it the *homology marking* of e. The map $\mu\colon E^{\mathrm{ori}}(G) \to H$ has the following properties:

(1) For any $e \in E^{\mathrm{ori}}(G)$, we have $\mu(\overline{e}) = -\mu(e)$.

(2) The set $\{\mu(e)\}_{e \in E^{\mathrm{ori}}(G)}$ generates H.

(3) For any $v \in V(G)$, we have

$$\sum_{e \in E_v^{\mathrm{ori}}(G)} \mu(e) = 0.$$

For example, in the notation of the left part of Figure 1, where we orient edges a, b, c, d as indicated, we have $\mu(a) + \mu(b) + \mu(c) + \mu(d) = 0$.

In what follows, we consider $\mathcal{M}_{g,1}$–modules such as H and its third exterior power $\Lambda^3 H$. There is a twisted cohomology class $\widetilde{k} \in H^1\big(\mathcal{M}_{g,1}; \frac{1}{2}\Lambda^3 H\big)$ called the *extended first Johnson homomorphism* [17]. Here, $\frac{1}{2}\Lambda^3 H = \big\{\frac{1}{2}u \in \Lambda^3(H \otimes_{\mathbb{Z}} \mathbb{Q}) \mid u \in \Lambda^3 H\big\}$, where we take the canonical embedding $H \to H \otimes_{\mathbb{Z}} \mathbb{Q}$, $x \mapsto x \otimes 1$. This cohomology class has a fundamental importance in the study of the cohomology of the mapping class group; see [11].

Theorem 2.4 (Morita and Penner [18]) *Keep the notation in Figure 1. For the flip W_e, set*

$$j(W_e) = \mu(a) \wedge \mu(b) \wedge \mu(c) \in \Lambda^3 H.$$

Then j is an $\mathcal{M}_{g,1}$–equivariant 1–cocycle on $\widehat{\mathcal{G}}(\Sigma_{g,1})$, and $[j_G] = 6\widetilde{k}$.

Using the intersection pairing $(\cdot)$ on the homology, we define an $\mathrm{Sp}(H)$–equivariant map

$$C\colon \Lambda^3 H \to H, \quad x \wedge y \wedge z \mapsto (x \cdot y)z + (y \cdot z)x + (z \cdot x)y$$

called the contraction. Morita [15] showed that if $g \geq 2$, the twisted cohomology group $H^1(\mathcal{M}_{g,1}; H)$ is infinite cyclic. As is remarked in [17], the element $k := C(2\widetilde{k})$ is a generator of this cohomology group. Since Earle [3] first gave a cocycle representative for k, we call k the *Earle class*; see [10]. The restriction of k to the Torelli subgroup of $\mathcal{M}_{g,1}$ is called the *Chillingworth homomorphism*; see [7, Section 5].

Theorem 2.5 (Kuno, Penner and Turaev [13]) *Keep the notation in Figure 1. For the flip W_e, set*

$$m(W_e) = \mu(a) + \mu(c) \in H.$$

Then m is an $\mathcal{M}_{g,1}$–equivariant 1–cocycle on $\widehat{\mathcal{G}}(\Sigma_{g,1})$, and $[m_G] = 6k$.

Here we correct an error in [13]. Let $\varphi_{\mathrm{BP}} = \varphi$ be the torus BP map in [13, Figure 3], which was first considered in [18]. In [13, Lemma 1], it was asserted that $m(\varphi_{\mathrm{BP}}) = 4a$, but this is not true. More precisely, in the proof of the lemma, we computed the contribution of the second Dehn twist (5 flips) as $-4a$, but this should be corrected to $4a$.

Lemma 2.6 (correction of [13, Lemma 1]) *Let φ_{BP} be the torus BP map as above. Then $m(\varphi_{\mathrm{BP}}) = 12\mu(a)$.*

In [13, Theorem 6], it is asserted that $[m_G] = -2k$, but this should be corrected as in Theorem 2.5 above.

3 A secondary invariant

We consider the cocycle $j' = C \circ j$. For the flip W_e in the notation of Figure 1, we have

$$j'(W_e) = (a \cdot b)\mu(c) + (b \cdot c)\mu(a) + (c \cdot a)\mu(b) \in H.$$

Here and throughout the paper, we write eg $(a \cdot b)$ instead of $(\mu(a) \cdot \mu(b))$ for simplicity. By Theorems 2.4 and 2.5, for any trivalent fatgraph spine G, we have

$$2[j'_G] = [m_G] = 6k.$$

Therefore, there exists an element $\xi_G \in H$ such that $2j'_G - m_G = \delta\xi_G$. Here the symbol δ in the right-hand side means the coboundary map in the standard cochain complex of $\mathcal{M}_{g,1}$ with coefficients in H. Explicitly, we have $(\delta\xi_G)(\varphi) = \varphi \cdot \xi_G - \xi_G$ for any $\varphi \in \mathcal{M}_{g,1}$. Such a ξ_G is unique since only 0 is $\mathcal{M}_{g,1}$–invariant in H. We regard the collection $\xi = \{\xi_G\}_G$ as a cellular 0–cochain of $\widehat{\mathcal{G}}(\Sigma_{g,1})$ with coefficients in H.

Proposition 3.1 (1) *The 0–cochain ξ is $\mathcal{M}_{g,1}$–equivariant in the sense that $\xi_{\psi(G)} = \psi \cdot \xi_G$ for any $\psi \in \mathcal{M}_{g,1}$ and any trivalent fatgraph spine G.*

(2) *We have $2j' - m = \delta\xi$. Namely, for any flip $G \xrightarrow{W} G'$, we have $\xi_{G'} - \xi_G = 2j'(W) - m(W)$.*

Moreover, these two properties characterize ξ.

Proof (1) For simplicity we write $s = 2j' - m$. Take $\varphi \in \mathcal{M}_{g,1}$. Using $s_G(\varphi) = \delta\xi_G(\varphi) = \varphi \cdot \xi_G - \xi_G$, etc, we compute from Lemma 2.3 that

$$\begin{aligned} s_{\psi(G)}(\varphi) &= s_G(\varphi) + \varphi \cdot s_G(\psi) - s_G(\psi) \\ &= \varphi \cdot \xi_G - \xi_G + \varphi \cdot (\psi \cdot \xi_G - \xi_G) - (\psi \cdot \xi_G - \xi_G) \\ &= \varphi \cdot (\psi \cdot \xi_G) - \psi \cdot \xi_G \\ &= \delta(\psi \cdot \xi_G)(\varphi). \end{aligned}$$

This proves $s_{\psi(G)} = \delta(\psi \cdot \xi_G)$. By the uniqueness of $\xi_{\psi(G)}$, it follows that $\xi_{\psi(G)} = \psi \cdot \xi_G$.

(2) This follows from $s_G(\varphi) + \varphi \cdot s(W) = s(W) + s_{G'}(\varphi)$ analogously, and so we omit the details.

Finally, suppose that ξ^0 is an $\mathcal{M}_{g,1}$–equivariant 0–cochain satisfying $2j' - m = \delta\xi^0$. Then $\xi - \xi^0$ is an $\mathcal{M}_{g,1}$–equivariant 0–cocycle. This shows that $\eta := \xi(G) - \xi^0(G) \in H$ is independent of G and $\varphi \cdot \eta = \eta$ for any $\varphi \in \mathcal{M}_{g,1}$. Therefore η must be zero and $\xi^0 = \xi$. □

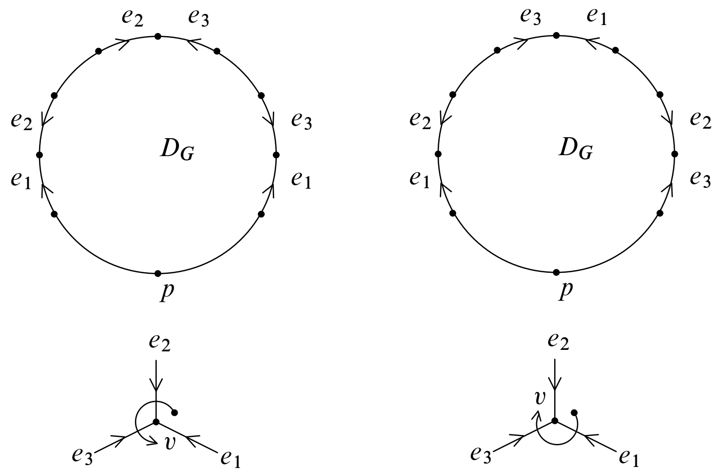

Figure 3: A vertex of type 1 (left) and a vertex of type 2 (right)

Let G be a trivalent fatgraph spine of $\Sigma_{g,1}$. We present an explicit formula for ξ_G. To begin with, we introduce a total ordering for $E^{\mathrm{ori}}(G)$. Note that if we cut $\Sigma_{g,1}$ along G, we obtain an oriented closed disk D_G.

Definition 3.2 (1) For $e, e' \in E^{\mathrm{ori}}(G)$, we say $e \prec e'$ if the edge e occurs first when we go *clockwise* along the boundary of D_G from p.

(2) Let $e \in E^{\mathrm{ori}}(G)$. We say that e has the *preferred orientation* (or e is *preferably oriented*) if $e \prec \overline{e}$.

Note that any unoriented edge of G has a unique preferred orientation.

Let $v \in V^{\mathrm{int}}(G)$. We name the three elements of $E_v^{\mathrm{ori}}(G)$ as e_1, e_2 and e_3 so that

(1) $e_1 \prec e_2$ and $e_1 \prec e_3$, and

(2) the edge e_2 is next to e_1 in the cyclic ordering given to $E_v^{\mathrm{ori}}(G)$.

There are two possibilities for the ordering of e_i and its inverse $\overline{e}_i$, namely,

$$e_1 \prec \overline{e}_2 \prec e_2 \prec \overline{e}_3 \prec e_3 \prec \overline{e}_1 \quad \text{and} \quad e_1 \prec \overline{e}_2 \prec e_3 \prec \overline{e}_1 \prec e_2 \prec \overline{e}_3.$$

The vertex v is said to be *of type* 1 if the former case happens, and is said to be *of type* 2 otherwise. Figure 3 is an illustration of the situation.

We can count the number of vertices of type 1 and the number of type 2.

Proposition 3.3 *For any trivalent fatgraph spine G of $\Sigma_{g,1}$, the number of trivalent vertices of type 1 is $2g-1$, and that of type 2 is $2g$.*

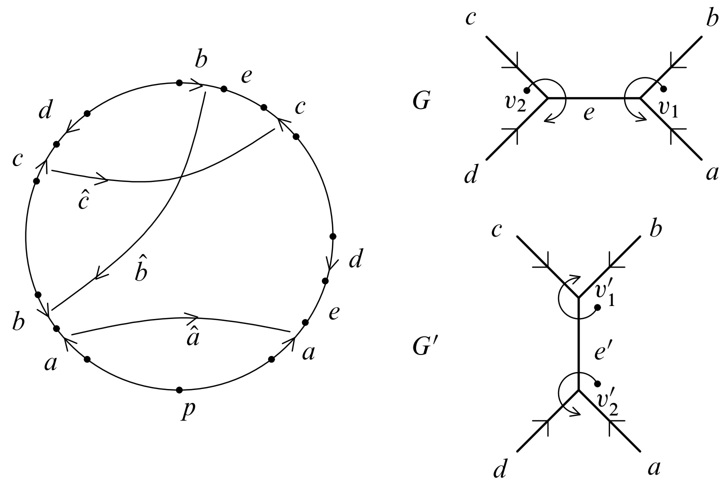

Figure 4: The case where $a \prec c \prec b \prec d$

Proof For $i = 1, 2$, let V_i be the number of trivalent vertices of type i. Since the number of trivalent vertices of G is $4g-1$, we have $V_1 + V_2 = 4g-1$. We observe that if a trivalent vertex v is of type i ($i = 1, 2$), the number of preferably oriented edges toward v is i. Thus $V_1 + 2V_2$ is equal to the number of edges of G, ie $6g-1$. Hence we obtain $V_1 = 2g-1$ and $V_2 = 2g$. □

We set

$$\begin{cases} e_v = e_2 \text{ and } f_v = e_3 & \text{if } v \text{ is of type } 1, \\ e_v = e_1 \text{ and } f_v = e_3 & \text{if } v \text{ is of type } 2. \end{cases}$$

Theorem 3.4 *We have*

$$\xi_G = \sum_v (\mu(e_v) - \mu(f_v)),$$

where the sum is taken over all trivalent vertices of G.

Proof We set $\xi^0_G = \sum_v (\mu(e_v) - \mu(f_v))$ and consider the collection $\xi^0 = \{\xi^0_G\}_G$. Clearly, ξ^0 is $\mathcal{M}_{g,1}$–equivariant. By Proposition 3.1, it is sufficient to prove that $2j' - m = \delta\xi^0$.

Take the notation as in Figure 1. For example, assume that $a \prec c \prec b \prec d$. For simplicity, we write e instead of $\mu(e)$ for $e \in E^{\mathrm{ori}}(G)$. Then we can see from the left part of Figure 4 that $(a \cdot b) = (c \cdot a) = 0$ and $(b \cdot c) = 1$, and so $j'(W_e) = a$. Thus $2j'(W_e) - m(W_e) = 2a - (a + c) = a - c$. On the other hand, we can compute from

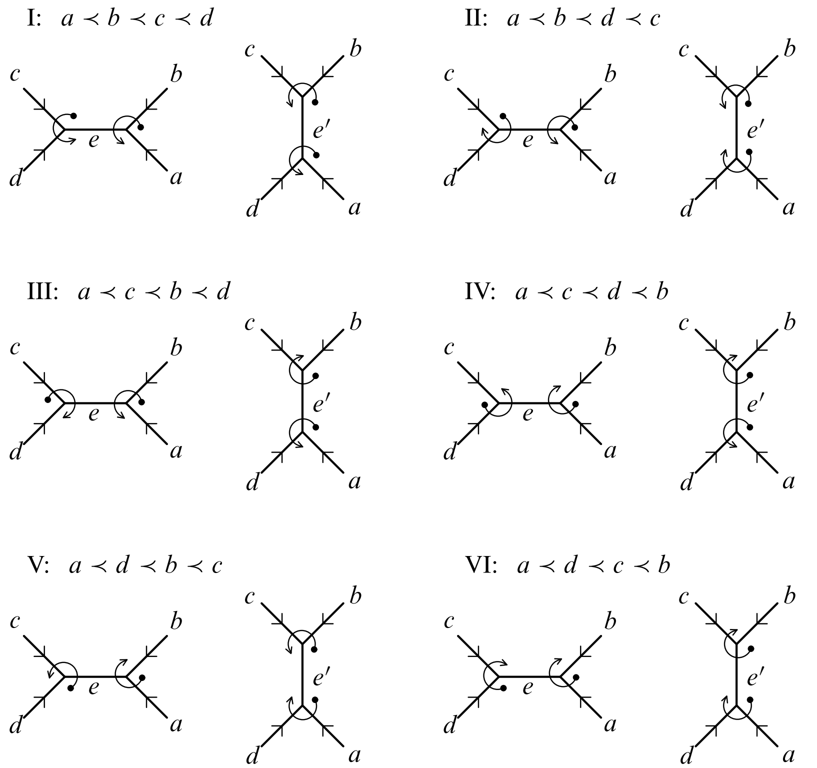

Figure 5: Situations near e

the right part of Figure 4 that

$$\begin{aligned}\xi^0_{G'} - \xi^0_G &= (e_{v'_1} - f_{v'_1}) + (e_{v'_2} - f_{v'_2}) - (e_{v_1} - f_{v_1}) - (e_{v_2} - f_{v_2}) \\ &= (a + d - c) + (b + c - d) - (b - (c + d)) - (c - (a + b)) \\ &= 2a + b + d = 2a + b + (-a - b - c) = a - c.\end{aligned}$$

We can compute similarly for other cases as well, and we obtain $2j'(W_e) - m(W_e) = \xi^0_{G'} - \xi^0_G$. (There are essentially six cases to consider; in each case in Figure 5, we may assume that G corresponds to the left picture.) Hence $2j' - m = \delta\xi^0$, as required. $\square$

Example 3.5 Let G be the fatgraph as shown in Figure 6. We name edges as in the figure and give them the preferred orientation. For $1 \leq i \leq g$ and $1 \leq j \leq 3$, let $v_i^j \in V^{\mathrm{int}}(G)$ be the start point of e_i^j. For $1 \leq i \leq g-1$, let $v_i^4 \in V^{\mathrm{int}}(G)$ be the endpoint of e_i^4.

Since v_i^1 is of type 1, its contribution is $\mu(\overline{e}_i^1) - \mu(\overline{e}_i^4) = \mu(e_i^4) - \mu(e_i^1)$. Since v_i^2 is of type 2, its contribution is $\mu(e_i^1) - \mu(e_i^3)$. Since v_i^3 is of type 2, its contribution is

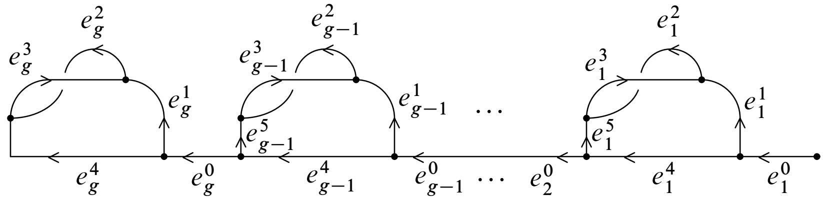

Figure 6: The fatgraph in Example 3.5

$\mu(e_i^2)-\mu(e_i^5)$. Here we understand that $e_g^5=e_g^4$. Since v_i^4 is of type 1, its contribution is $\mu(\overline{e}_i^5)-\mu(\overline{e}_{i+1}^0)=\mu(e_{i+1}^0)-\mu(e_i^5)$.

Moreover, we have $\mu(e_i^0)=0$, $\mu(e_i^1)+\mu(e_i^3)=\mu(e_i^2)$, and $\mu(e_i^4)=\mu(e_i^5)=-\mu(e_i^1)$. Using these relations, we obtain

$$\xi_G=\mu(e_g^1)+\sum_{i=1}^{g-1}2\mu(e_i^1).$$

4 Elementary properties

In this section, we record two elementary properties of ξ_G.

We first show a certain gluing formula. Let g and g' be positive integers, and suppose that we have two trivalent fatgraph spines $\iota\colon G\hookrightarrow\Sigma_{g,1}$ and $\iota'\colon G'\hookrightarrow\Sigma_{g',1}$. Fix $e\in E^{\mathrm{ori}}(G)$. Plugging the tail of G' in the right side of e, one produces a new fatgraph spine of $\Sigma_{g+g',1}$. A precise construction is as follows. Let v_e be the middle point of e.

1. (1) Take a small closed disk D_e in $\Sigma_{g,1}$ such that $\mathrm{Int}(D_e)\cap G=\varnothing$, the boundary ∂D_e intersects G once at v_e, and the center of D_e is on the right side of e with respect to the orientation of e.
2. (2) Glue $\Sigma_{g,1}\setminus\mathrm{Int}(D_e)$ with $\Sigma_{g',1}$ along the boundaries ∂D_e and $\partial\Sigma_{g',1}$ so that the univalent vertex of G' is identified with v_e.
3. (3) Let G'' be the union of the images of G and G' in the result of gluing.

The glued surface is diffeomorphic to $\Sigma_{g+g',1}$. We consider G'' as a trivalent fatgraph spine of $\Sigma_{g+g',1}$ by dividing e into two edges sharing the newly created trivalent vertex v_e. These two edges receive their orientation from e. We name them as $e_1,e_2\in E^{\mathrm{ori}}(G'')$ so that v_e is the endpoint of e_1. The edges e_1 and e_2 have the same homology marking as e.

A schematic figure of this construction is Figure 7. We call G'' the *gluing* of G and G' at e. Note that the inclusions $\Sigma_{g,1}\setminus\mathrm{Int}(D_e)\hookrightarrow\Sigma_{g+g',1}$ and $\Sigma_{g',1}\hookrightarrow\Sigma_{g+g',1}$ induce

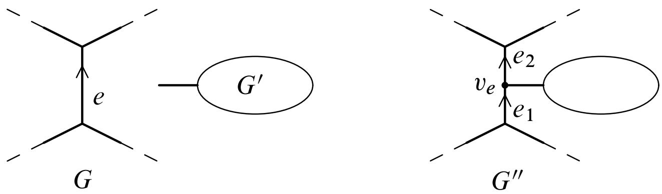

Figure 7: Gluing

a direct sum decomposition

$$H_1(\Sigma_{g+g',1};\mathbb{Z}) \cong H_1(\Sigma_{g,1};\mathbb{Z}) \oplus H_1(\Sigma_{g',1};\mathbb{Z}). \tag{4-1}$$

Proposition 4.1 (gluing formula) *Let G'' be the gluing of G and G' at e, as above. Then $\xi_{G''} = \xi_G + \mu(e) + \xi_{G'}$.*

Proof We have a natural identification $V^{\mathrm{int}}(G'') \cong V^{\mathrm{int}}(G) \sqcup \{v_e\} \sqcup V^{\mathrm{int}}(G')$. Observe that this identification respects the type of vertices. With the direct sum decomposition (4-1) in mind, we see that $V^{\mathrm{int}}(G)$ and $V^{\mathrm{int}}(G')$ contribute to $\xi_{G''}$ as ξ_G and $\xi_{G'}$, respectively.

We compute the contribution from v_e. Let $t' \in E^{\mathrm{ori}}_{v_e}(G'')$ be an edge coming from the tail of G'. The homology marking of t' is trivial. Then the contribution from v_e is $\mu(t') - \mu(\overline{e}_2) = \mu(e)$ if e has the preferred orientation, and is $\mu(e_1) - \mu(t') = \mu(e)$ otherwise. This completes the proof. □

We next show a formula describing how ξ_G changes under a special kind of flip. For a trivalent fatgraph spine $G \subset \Sigma_{g,1}$, we use the following notation:

- We denote by t the tail of G, and give it the preferred orientation.
- $e_1 \in E^{\mathrm{ori}}(G)$ is the oriented edge next to t in the total ordering given to $E^{\mathrm{ori}}(G)$.
- v_1 and v_2 are the start and end points of e_1, respectively.
- $b, c \in E^{\mathrm{ori}}_{v_2}(G)$ are the edges such that e_1, b and c are in this order in the cyclic ordering given to $E^{\mathrm{ori}}_{v_2}(G)$.

The situation is illustrated in Figure 8. We call the flip along (the unoriented edge underlying) e_1 the *tail slide* to G.

Proposition 4.2 (tail slide formula) *Let G' be the result of the tail slide to G. Then $\xi_{G'} = \xi_G + \mu(c)$.*

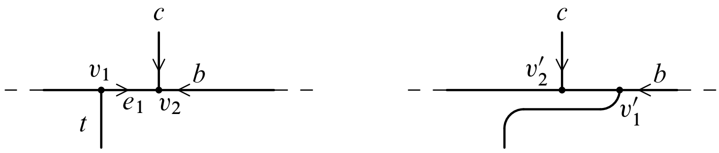

Figure 8: The tail slide applied to G (left) gives G' (right)

Proof We work with Figure 8. Suppose $b \prec c$ in $E^{\mathrm{ori}}(G)$. For simplicity, we write e instead of $\mu(e)$ for $e \in E^{\mathrm{ori}}(G)$. Then we compute

$$\begin{aligned}\xi_{G'} - \xi_G &= (e_{v_1'} - f_{v_1'}) + (e_{v_2'} - f_{v_2'}) - (e_{v_1} - f_{v_1}) - (e_{v_2} - f_{v_2}) \\ &= (b - (-b)) + (c - (-b - c)) - (b + c - (-b - c)) - (b - c) \\ &= c.\end{aligned}$$

The case where $c \prec b$ can be computed similarly. □

As an application of Proposition 4.2, we can extend the definition of our invariant to trivalent fatgraph spines of a *once-punctured* surface. Let Σ_g^1 be a surface obtained from $\Sigma_{g,1}$ by gluing a once-punctured disk along the boundaries. We regard $\Sigma_{g,1}$ as a subset of Σ_g^1. By definition, a fatgraph spine of Σ_g^1 is an embedding $\iota\colon G \hookrightarrow \Sigma_g^1$ of a fatgraph G into Σ_g^1 satisfying the first two conditions in Definition 2.1 (with $\Sigma_{g,1}$ replaced by Σ_g^1), and the condition that all vertices have valency greater than 2.

Let G be a trivalent fatgraph spine of Σ_g^1. By a suitable isotopy, we arrange that $G \subset \Sigma_{g,1}$. Let $e \in E^{\mathrm{ori}}(G)$. Take a simple arc ℓ on $\Sigma_{g,1}$ starting from p, reaching v_e from the right, and disjoint from $G \setminus \{v_e\}$. We say that such an arc ℓ is *admissible* for e. Regarding v_e as a newly created trivalent vertex, we can consider the union $\widetilde{G}(e,\ell) = G \cup \ell$ as a trivalent fatgraph spine of $\Sigma_{g,1}$. The arc ℓ becomes the tail of $\widetilde{G}(e,\ell)$.

Corollary 4.3 *Keep the notation as above. Then the element $\xi_{\widetilde{G}(e,\ell)} - \mu(e)$ does not depend on the choice of e and ℓ. In particular, for a trivalent fatgraph spine $G \subset \Sigma_g^1$, we can define $\xi_G \in H = H_1(\Sigma_{g,1};\mathbb{Z}) \cong H_1(\Sigma_g^1;\mathbb{Z})$ as*

$$\xi_G := \xi_{\widetilde{G}(e,\ell)} - \mu(e).$$

Proof Let ℓ^0 be another admissible arc for e. Then ℓ^0 is isotopic to the concatenation of some power of a simple based loop parallel to $\partial\Sigma_{g,1}$ and ℓ. This implies that $\widetilde{G}(e,\ell^0)$ is obtained from $\widetilde{G}(e,\ell)$ by application of some power of the Dehn twist along $\partial\Sigma_{g,1}$. Since the Dehn twist along $\partial\Sigma_{g,1}$ acts on H trivially, we have $\xi_{\widetilde{G}(e,\ell)} = \xi_{\widetilde{G}(e,\ell^0)}$. Hence $\xi_{\widetilde{G}(e,\ell)} - \mu(e)$ does not depend on the choice of ℓ.

Now, we can give a *cyclic* ordering to the set $E^{\mathrm{ori}}(G)$ in a way similar to that in the case where $G \subset \Sigma_{g,1}$ as in Definition 3.2. Suppose that $e, e' \in E^{\mathrm{ori}}(G)$ are consecutive in this cyclic ordering. Fix an admissible arc ℓ for e. Let v_0 be the vertex of G shared by e and e', and let $c \in E^{\mathrm{ori}}_{v_0}(G)$ be an edge other than e and $\overline{e}'$. We denote by e_0 an unoriented edge of $\widetilde{G}(e,\ell)$ with endpoints v_e and v_0.

Let $\widetilde{G}'$ be the result of flip along e_0. Then $\widetilde{G}'$ can be identified with $\widetilde{G}(e',\ell')$, where ℓ' corresponds to the tail of $\widetilde{G}'$. By Proposition 4.2, we have $\xi_{\widetilde{G}(e',\ell')} = \xi_{\widetilde{G}(e,\ell)} + \mu(c)$. Since $\mu(c)+\mu(e) = \mu(e')$, we obtain $\xi_{\widetilde{G}(e',\ell')} - \mu(e') = \xi_{\widetilde{G}(e,\ell)} - \mu(e)$. Hence $\xi_{\widetilde{G}(e,\ell)} - \mu(e)$ does not depend on the choice of e either. □

5 Nontriviality and primitivity

Let us consider the mod 2 reduction of ξ_G:

$$\xi^2_G := \xi_G \otimes (1 \bmod 2) \in H \otimes \mathbb{Z}_2 \cong H_1(\Sigma_{g,1};\mathbb{Z}_2).$$

Hereafter, $\equiv$ stands for an equality in $H\otimes\mathbb{Z}_2$. Since $\mu(\overline{e}) = -\mu(e) \equiv \mu(e) \in H\otimes\mathbb{Z}_2$ for any $e \in E^{\mathrm{ori}}(G)$, the homology marking μ induces a well-defined map $\mu^2\colon E(G) \to H\otimes\mathbb{Z}_2$. We call μ^2 the *mod* 2 *homology marking*.

Proposition 5.1 *Let G be a trivalent fatgraph spine of $\Sigma_{g,1}$. Then we have*

$$\xi^2_G = \sum_{e\in E(G)} \mu^2(e).$$

Proof Let $v \in V^{\mathrm{int}}(G)$. We work with Figure 3 and count preferably oriented edges toward v. By abuse of notation, we use the same letter for an oriented edge and its underlying unoriented edge. If v is of type 1, only e_1 has the preferred orientation. Since $\mu(e_1)+\mu(e_2)+\mu(e_3) = 0$, we have

$$\mu(e_v) - \mu(f_v) = \mu(e_2) - \mu(e_3) \equiv \mu(e_1).$$

If v is of type 2, e_1 and e_3 have the preferred orientation and e_2 does not. Then we have

$$\mu(e_v) - \mu(f_v) = \mu(e_1) - \mu(e_3) \equiv \mu(e_1) + \mu(e_3).$$

Therefore, we have

$$\xi^2_G = \sum_{v\in V^{\mathrm{int}}(G)} \begin{pmatrix} \text{sum of the mod 2 homology markings} \\ \text{of preferably oriented edges toward } v \end{pmatrix} = \sum_{e\in E(G)} \mu^2(e).$$

The last equality holds since any preferably oriented edge of G points to some trivalent vertex of G. □

Theorem 5.2 *Let G be a trivalent fatgraph spine of $\Sigma_{g,1}$. Then the mod 2 reduction ξ^2_G is nontrivial. In particular, we have $\xi_G \neq 0$.*

To prove this theorem, we need the following lemma.

Lemma 5.3 *Let G be a trivalent fatgraph spine of $\Sigma_{g,1}$. Then G contains an edge cycle of odd length.*

Proof We introduce some terminology: a pair of consecutive oriented edges of G is called a *corner* of G. There are $3\,\#V^{\mathrm{int}}(G) = 3(4g-1)$ corners. We number them as $c_1, \dots, c_{3(4g-1)}$, so that c_1 contains the preferably oriented tail of G, and for each i, c_i and c_{i+1} share an oriented edge in common. There are $n_o := 6g-1$ odd-numbered corners, and $n_e := 6g-2$ even-numbered corners.

Since n_o and n_e are not divisible by 3, there exist distinct indices i and j with $1 \le i < j \le 3(4g-1)$ such that the corners c_i and c_j are around the same vertex and $i - j \equiv 1 \bmod 2$. We can write c_i and c_j as $c_i = (e_i, e_i')$ and $c_j = (e_j, e_j')$ with $e_i \prec e_i'$ and $e_j \prec e_j'$. Consider the edge cycle following consecutive oriented edges of G from e_i' to e_j. Since i and j have different parity, the length of this edge cycle must be odd. □

Proof of Theorem 5.2 By Lemma 5.3, G contains an edge cycle γ of odd length. By Proposition 5.1, the mod 2 intersection pairing of ξ_G^2 and γ is computed as

$$(\xi_G^2 \cdot \gamma) = \left(\sum_{e \in E(G)} \mu^2(e) \cdot \gamma \right) = (\text{length of } \gamma) = 1.$$

Therefore, $\xi_G^2 \neq 0$. □

Remark 5.4 As far as we observed, ξ_G seems to be a primitive element of H for any trivalent fatgraph spine $G \subset \Sigma_{g,1}$. Here, an element $x \in H$ is called *primitive* if there do not exist $m \in \mathbb{Z}$ and $y \in H$ such that $|m| \ge 2$ and $x = my$. This primitivity of ξ_G holds for $g \le 2$. In fact, there is only one combinatorial isomorphism class of trivalent fatgraph spines for $g = 1$, and there are 105 classes for $g = 2$. By a direct computation, we can show the primitivity of ξ_G for these cases. The case $g \ge 3$ remains open.

In the case of trivalent fatgraph spines of a once-punctured surface Σ_g^1, it can happen that $\xi_G = 0$. Two examples for $g = 2$ are given in Figure 9.

Let G be a trivalent fatgraph spine of Σ_g^1. A corner of G is a pair of consecutive oriented edges of G in the cyclic ordering given to $E^{\mathrm{ori}}(G)$ (see the proof of Corollary 4.3). Now we give labels α or β to each corner of G so that any pair of consecutive corners of G have distinct labels. Since the number of corners of G is even, this labeling is always possible and is determined once we choose the label of a fixed corner.

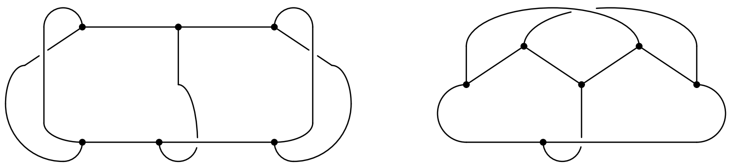

Figure 9: Trivalent fatgraph spines with $\xi_G = 0$

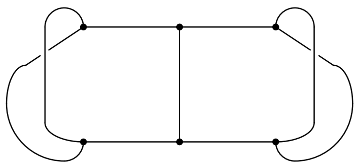

Figure 10: A balanced trivalent fatgraph spine with $\xi_G \neq 0$

We say that G is *balanced* if for any vertex of G, the three corners around the vertex have the same label. For example, trivalent fatgraph spines in Figure 9 and Figure 10 are balanced.

Theorem 5.5 *Let G be a trivalent fatgraph spine of Σ_g^1. Then the mod 2 reduction $\xi_G^2 = \xi_G \otimes (1 \bmod 2)$ is trivial if and only if G is balanced.*

Proof Pick a corner c of G and write it as $c = (e, e')$, where e' is next to e in the cyclic ordering given to $E^{\mathrm{ori}}(G)$. We give the label α to c and extend this labeling to all other corners as above. Take an admissible arc ℓ for e and set $\widetilde{G} = \widetilde{G}(e, \ell)$. The oriented edge e is split at the middle point v_e into two oriented edges. We name them as $e_1, e_2 \in E^{\mathrm{ori}}(\widetilde{G})$ so that v_e is the endpoint of e_1. We extend the labeling of corners of G to that of corners of $\widetilde{G}$ by giving α to $(e_1, \overline{\ell})$ and $(\overline{e}_2, \overline{e}_1)$, and β to (ℓ, e_2).

In view of Corollary 4.3, the condition $\xi_G^2 = 0$ is equivalent to $\xi_{\widetilde{G}}^2 = \mu^2(e_2)$. Furthermore, since the mod 2 homology markings $\{\mu^2(f)\}_{f \in E(\widetilde{G})}$ generate the mod 2 homology $H_1(\Sigma_{g,1}; \mathbb{Z}_2)$, this condition is equivalent to the condition that $(\xi_{\widetilde{G}}^2 \cdot \mu^2(f)) = (\mu^2(e_2) \cdot \mu^2(f))$ for any $f \in E(\widetilde{G})$.

Assume that G is balanced. For any vertex of $\widetilde{G}$ other than v_e, the three corners about it are labeled by the same symbol. Let $f \in E(\widetilde{G})$. Let $\gamma(f)$ be the edge cycle following consecutive oriented edges of $\widetilde{G}$ from f to $\overline{f}$, where we give the preferred orientation to f. The mod 2 homology class $\mu^2(f)$ is represented by $\gamma(f)$. By the property of the labeling, the length of this edge cycle is odd if $f \prec \overline{e}_2 \prec \overline{f}$ (this also implies $f \neq e_2$), and is even otherwise. Note that the condition $f \prec \overline{e}_2 \prec \overline{f}$ is equivalent to $(\mu^2(e_2) \cdot \mu^2(f)) = 1$. Hence $(\xi_{\widetilde{G}}^2 \cdot \mu^2(f)) = (\text{the length of } \gamma(f)) = 1$ if and only if $(\mu^2(e_2) \cdot \mu^2(f)) = 1$. Therefore, $\xi_G^2 = 0$.

On the other hand, assume that $\xi_G^2 = 0$. Then for $f \in E(\widetilde{G})$, the length of $\gamma(f)$ is odd if and only if $f \prec \overline{e}_2 \prec \overline{f}$. Now we remove the tail from $\widetilde{G}$ and go back to G. Then $\gamma(f)$ is reduced to an edge cycle of G. Its length is 1 less than the length of $\gamma(f)$ if $f \prec \overline{e}_2 \prec \overline{f}$, and is the same as the length of $\gamma(f)$ otherwise. This implies that the reduced edge cycle of G has even length. Since f can be arbitrary, this shows that G is balanced. □

6 Mod 2 reduction and spin structures

In this section, we give a topological interpretation of the mod 2 reduction ξ_G^2. We start with the following description of the mod 2 homology of $\Sigma_{g,1}$.

Lemma 6.1 *Let G be a fatgraph spine of $\Sigma_{g,1}$. For $v \in V^{\mathrm{int}}(G)$, let $\{e_i^v\}_i$ be the set of unoriented edges of G having v as an endpoint. If there is an edge loop based at v, we count it twice. Then the mod 2 homology marking induces an isomorphism*

$$H_1(\Sigma_{g,1};\mathbb{Z}_2) \cong \bigoplus_{e \in E(G)} \mathbb{Z}_2 e \Big/ \sum_{v \in V^{\mathrm{int}}(G)} \mathbb{Z}_2 \Big(\sum_i e_i^v\Big).$$

Proof Recall from Section 2 that we associate an oriented simple loop $\widehat{e}$ to each (oriented) edge e. In the proof of this lemma we forget the orientation of e and $\widehat{e}$. We can arrange that the simple loops $\{\widehat{e}\}_{e \in E(G)}$ share only one point $q \in \partial\Sigma_{g,1}$, and that if t is the tail of G then $\widehat{t} = \partial\Sigma_{g,1}$ with basepoint q. Then we obtain a cell decomposition of $\Sigma_{g,1}$ whose 1–cells coincide with $\{\widehat{e}\}_{e \in E(G)}$. Now the right-hand side of the assertion can be identified with the first mod 2 cellular homology group of this cell decomposition. □

Recall that a *spin structure* on $\Sigma_{g,1}$ is an element $w \in H^1(UT\Sigma_{g,1};\mathbb{Z}_2)$, where $UT\Sigma_{g,1}$ is the unit tangent bundle of $\Sigma_{g,1}$ (with respect to some Riemannian metric), such that the restriction of w to a fiber of the projection $UT\Sigma_{g,1} \to \Sigma_{g,1}$ is nontrivial. As Johnson [8] showed, the set of spin structures on $\Sigma_{g,1}$ is naturally identified with the set of quadratic forms on $H_1(\Sigma_{g,1};\mathbb{Z}_2)$. Here, a map $q\colon H_1(\Sigma_{g,1};\mathbb{Z}_2) \to \mathbb{Z}_2$ is called a *quadratic form* on $H_1(\Sigma_{g,1};\mathbb{Z}_2)$ if it satisfies

$$q(x+y) = q(x) + q(y) + (x \cdot y)$$

for any $x, y \in H_1(\Sigma_{g,1};\mathbb{Z}_2)$. The set of spin structures on $\Sigma_{g,1}$ is a torsor under the action of $H^1(\Sigma_{g,1};\mathbb{Z}_2)$. In other words, the difference between two quadratic forms on $H_1(\Sigma_{g,1};\mathbb{Z}_2)$ can be written as a uniquely determined element of

$\mathrm{Hom}(H_1(\Sigma_{g,1};\mathbb{Z}_2),\mathbb{Z}_2) \cong H^1(\Sigma_{g,1};\mathbb{Z}_2)$. Note that, using the mod 2 intersection pairing, we have a natural isomorphism

$$H_1(\Sigma_{g,1};\mathbb{Z}_2) \cong \mathrm{Hom}(H_1(\Sigma_{g,1};\mathbb{Z}_2),\mathbb{Z}_2), \quad x \mapsto [y \mapsto (x\cdot y)]. \tag{6-1}$$

In what follows, G is a trivalent fatgraph spine of $\Sigma_{g,1}$. The following result gives an identification of certain $\mathbb{Z}_2$–valued functions on $E(G)$ with the set of quadratic forms on $H_1(\Sigma_{g,1};\mathbb{Z}_2)$, thus with the set of spin structures on $\Sigma_{g,1}$ via Johnson's result stated above.

Theorem 6.2 *Let G be a trivalent fatgraph spine of $\Sigma_{g,1}$. Let $Q(G)$ be the set of maps $q\colon E(G)\to\mathbb{Z}_2$ such that, for any $v\in V^{\mathrm{int}}(G)$, the sum of values of q at the three edges having v as an endpoint is 0 if v is of type 1, and is 1 if v is of type 2. Then there is a natural bijection from $Q(G)$ to the set of quadratic forms on $H_1(\Sigma_{g,1};\mathbb{Z}_2)$.*

Proof Given a map $q\colon E(G)\to\mathbb{Z}_2$, we extend q to a map from the free $\mathbb{Z}_2$–module generated by $E(G)$ by

$$q\Bigl(\sum_{e\in E(G)} m_e e\Bigr) := \sum_{e\in E(G)} m_e q(e) + \sum_{e\prec e'} m_e m_{e'}(\mu^2(e)\cdot\mu^2(e')), \tag{6-2}$$

for $m_e\in\mathbb{Z}_2$, $e\in E(G)$. Here $(\cdot)$ is the mod 2 intersection pairing and we give the preferred orientation to each element of $E(G)$. By a direct computation, we can check that, for any $x,y\in\bigoplus_{e\in E(G)}\mathbb{Z}_2 e$,

$$q(x+y) = q(x)+q(y)+(x\cdot y). \tag{6-3}$$

Here $(x\cdot y)$ is the mod 2 intersection pairing of the homology class determined by x and y through the isomorphism in Lemma 6.1.

We claim that if $q\in Q(G)$, then for any $v\in V^{\mathrm{int}}(G)$,

$$q(e_1^v+e_2^v+e_3^v)=0.$$

By (6-2), this condition is equivalent to the equality

$$\sum_{i=1}^{3} q(e_i^v) + (\mu^2(e_1^v)\cdot\mu^2(e_2^v)) + (\mu^2(e_1^v)\cdot\mu^2(e_3^v)) + (\mu^2(e_2^v)\cdot\mu^2(e_3^v)) = 0. \tag{6-4}$$

If v is of type 1, then $(\mu^2(e_i^v)\cdot\mu^2(e_j^v))=0$ for any $1\le i,j\le 3$. If v is of type 2, then $(\mu^2(e_i^v)\cdot\mu^2(e_j^v))=1$ for any $1\le i,j\le 3$ with $i\neq j$. See Figure 3. Therefore, the condition (6-4) is exactly equivalent to the condition for q being an element of $Q(G)$. This proves the claim.

By the claim, Lemma 6.1 and (6-3), it follows that the map q induces a quadratic form on $H_1(\Sigma_{g,1};\mathbb{Z}_2)$. The above construction gives a map from $Q(G)$ to the set of quadratic forms on $H_1(\Sigma_{g,1};\mathbb{Z}_2)$, and the inverse of this map is given by composing any quadratic form on $H_1(\Sigma_{g,1};\mathbb{Z}_2)$ with the mod 2 homology marking $\mu^2\colon E(G)\to H_1(\Sigma_{g,1};\mathbb{Z}_2)$. □

We record how the set $Q(G)$ changes under a flip.

Proposition 6.3 *Let $W = W_e$ be a flip from G to G'. Then the bijection in Theorem 6.2 induces a bijection from $Q(G)$ to $Q(G')$, which maps a given $q \in Q(G)$ to the element $q' \in Q(G')$ defined as follows:*

- *For any edge f in $E(G')\setminus\{e'\}\cong E(G)\setminus\{e\}$, we have $q'(f)=q(f)$.*
- *We adopt the notation in Figure 5, and assume that in each case G and G' correspond to the left and right pictures, respectively. Then the value $q'(e')$ is given by the following formula:*

$$\begin{aligned}
\text{I:}\quad & q'(e') = q(b)+q(c) = q(a)+q(d),\\
\text{II:}\quad & q'(e') = q(b)+q(c) = q(a)+q(d)+1,\\
\text{III:}\quad & q'(e') = q(b)+q(c)+1 = q(a)+q(d),\\
\text{IV:}\quad & q'(e') = q(b)+q(c)+1 = q(a)+q(d),\\
\text{V:}\quad & q'(e') = q(b)+q(c) = q(a)+q(d)+1,\\
\text{VI:}\quad & q'(e') = q(b)+q(c)+1 = q(a)+q(d)+1.
\end{aligned}$$

By a suitable replacement of labels of edges, one can similarly obtain a formula for q' in terms of q for the case where G and G' correspond to the right and left pictures, respectively, in each case in Figure 5.

Proof To prove the first condition, note that the mod 2 homology marking of f as an edge of $E(G)$ is the same as that of f as an edge of $E(G')$. The second condition follows from the first condition and the defining relation for elements of $Q(G')$. For example, in case VI, two endpoints of e' are of type 2, and hence we have $q'(b)+q'(c)+q'(e') = q'(a)+q'(d)+q'(e') = 1$. □

Remark 6.4 The description of spin structures on $\Sigma_{g,1}$ given in Theorem 6.2 and how it changes under a flip as in Proposition 6.3 was pointed out by Robert Penner. Recently, Penner and Zeitlin [23] gave another natural description of spin structures on a punctured surface in terms of orientations on a trivalent fatgraph spine of the surface, and they also showed how it changes under a flip. In other words, Penner and Zeitlin gave a lift of the action of the mapping class group on the set of quadratic forms to the action of the Ptolemy groupoid, and the present construction gives another lift. It should be remarked that while their description works for any surfaces with multiple punctures, our description here is for a once (punctured/bordered) surface. It is an interesting question whether our description generalizes to any (punctured/bordered) surface.

In what follows, we denote by $E^+(G)$ the set of preferably oriented edges of G (see Definition 3.2), and let $E^-(G) := E^{\mathrm{ori}}(G) \setminus E^+(G)$.

Let $e \in E(G)$. We give e the preferred orientation and use the same letter e for the resulting element in $E^+(G)$. We define elements $q_G(e), \overline{q}_G(e) \in \mathbb{Z}_2$ by

$$q_G(e) := \#\{f \in E^+(G) \mid e \prec f \prec \overline{e}\} \bmod 2,$$
$$\overline{q}_G(e) := \#\{f \in E^-(G) \mid e \prec f \prec \overline{e}\} \bmod 2.$$

Here # means the number of elements of a set.

Proposition 6.5 *The maps q_G and $\overline{q}_G$ are elements of $Q(G)$.*

Proof We consider the case of q_G only.

We work with Figure 3. Suppose that v is of type 1. Then e_1, $\overline{e}_2$ and $\overline{e}_3$ have the preferred orientation, and we have a disjoint union decomposition

$$\{f \in E^+(G) \mid e_1 \prec f \prec \overline{e}_1\}$$
$$= \{\overline{e}_2, \overline{e}_3\} \sqcup \{f \in E^+(G) \mid \overline{e}_2 \prec f \prec e_2\} \sqcup \{f \in E^+(G) \mid \overline{e}_3 \prec f \prec e_3\}.$$

This implies that $q_G(e_1) = q_G(e_2) + q_G(e_3)$.

Suppose that v is of type 2. Then e_1, $\overline{e}_2$, and e_3 have the preferred orientation, and we have a disjoint union decomposition

$$\{f \in E^+(G) \mid \overline{e}_2 \prec f \prec e_2\}$$
$$= \big(\{f \in E^+(G) \mid e_1 \prec f \prec \overline{e}_1\} \setminus \{\overline{e}_2\}\big) \sqcup \{f \in E^+(G) \mid e_3 \prec f \prec \overline{e}_3\}.$$

This implies that $q_G(e_2) = q_G(e_1) + q_G(e_3) + 1$. Therefore, $q_G \in Q(G)$. $\square$

By Theorem 6.2, q_G and $\overline{q}_G$ induce quadratic forms on $H_1(\Sigma_{g,1}; \mathbb{Z}_2)$. For simplicity, we use the same letter q_G and $\overline{q}_G$ for these quadratic forms. This construction of quadratic forms is $\mathcal{M}_{g,1}$–equivariant in the following sense.

Proposition 6.6 *Let G be a trivalent fatgraph spine of $\Sigma_{g,1}$, and let $\varphi \in \mathcal{M}_{g,1}$. Then we have $q_{\varphi(G)} \circ \varphi_* = q_G$ and $\overline{q}_{\varphi(G)} \circ \varphi_* = \overline{q}_G$, where φ_* is the automorphism of $H_1(\Sigma_{g,1}; \mathbb{Z}_2)$ induced by φ.*

Proof We consider the case of q_G only. Consider a homomorphism

$$\Phi\colon \bigoplus_{e\in E(G)} \mathbb{Z}_2 e \to \bigoplus_{e'\in E(\varphi(G))} \mathbb{Z}_2 e', \quad \Phi(e)=\varphi(e).$$

Since φ gives a combinatorial isomorphism from G to $\varphi(G)$, we have $q_{\varphi(G)}\circ\Phi = q_G$. Now Φ induces the map φ_* on the level of homology, and we conclude $q_{\varphi(G)}\circ\varphi_* = q_G$. □

Finally, we compute the difference between q_G and $\overline{q}_G$.

Theorem 6.7 *Under the isomorphism* (6-1), *we have*

$$q_G - \overline{q}_G = \xi_G^2.$$

Moreover, we have $q_G \neq \overline{q}_G$.

Proof For $e\in E(G)$, we have

$$\begin{aligned} q_G(e)-\overline{q}_G(e) &= q_G(e)+\overline{q}_G(e)\\ &= \#\{f\in E^{\mathrm{ori}}(G)\mid e\prec f\prec \overline{e}\} \bmod 2\\ &= \Bigl(\sum_{f\in E(G)} \mu^2(f)\cdot\mu^2(e)\Bigr) = (\xi_G^2\cdot\mu^2(e)), \end{aligned}$$

where the last equality follows from Proposition 5.1. Since $\{\mu^2(e)\}_{e\in E(G)}$ generates $H_1(\Sigma_{g,1};\mathbb{Z}_2)$, we obtain $q_G-\overline{q}_G=\xi_G^2$. The second statement follows from Theorem 5.2. □

Appendix: A nonsingular vector field associated to a once-bordered trivalent fatgraph spine

Let G be a trivalent fatgraph spine of $\Sigma_{g,1}$. In this appendix, we define a nonsingular vector field $\mathcal{X}_G$ on $\Sigma_{g,1}$, and then consider the induced quadratic form on $H_1(\Sigma_{g,1};\mathbb{Z}_2)$. In particular, we discuss a relationship among this quadratic form, q_G and $\overline{q}_G$.

The following construction of $\mathcal{X}_G$ was communicated to the author by Gwénaël Massuyeau.

Let $\mathrm{Vect}(\Sigma_{g,1})$ be the homotopy set of nonsingular vector fields on $\Sigma_{g,1}$. In other words, $\mathrm{Vect}(\Sigma_{g,1})$ is the homotopy set of sections of the projection $\pi\colon UT\Sigma_{g,1}\to\Sigma_{g,1}$.

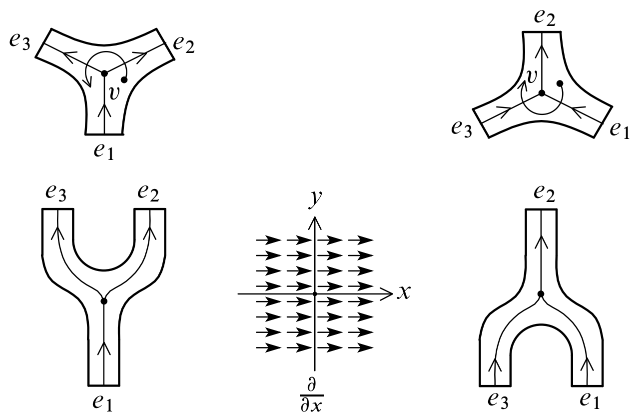

Figure 11: $\mathcal{X}_G$ on N_v. Left: a vertex of type 1. Right: a vertex of type 2.

For $\mathcal{X} \in \mathrm{Vect}(\Sigma_{g,1})$, the winding number

$$\mathrm{wind}_{\mathcal{X}}\colon \pi_1(UT\Sigma_{g,1}) \to \mathbb{Z}$$

is defined as follows. Let $\tilde{\gamma}\colon S^1 \to UT\Sigma_{g,1}$ be a (based) loop. For any $t \in S^1$, there exists a unique element $\Phi_t = \Phi(\mathcal{X}, \tilde{\gamma}, t) \in S^1 = U(1)$ such that $\mathcal{X}(\pi \circ \tilde{\gamma}(t))\Phi_t = \tilde{\gamma}(t)$. Then $\mathrm{wind}_{\mathcal{X}}(\tilde{\gamma})$ is defined to be the mapping degree of the map $S^1 \to S^1$, $t \mapsto \Phi_t$. The map $\mathrm{wind}_{\mathcal{X}}$ is a group homomorphism, and its mod 2 reduction

$$w_{\mathcal{X}} \in \mathrm{Hom}(\pi_1(UT\Sigma_{g,1}), \mathbb{Z}_2) \cong H^1(UT\Sigma_{g,1}; \mathbb{Z}_2)$$

is a spin structure on $\Sigma_{g,1}$.

Now we give the preferred orientation to any unoriented edge of G. Let $v \in V^{\mathrm{int}}(G)$. According to the type of v, we realize a small neighborhood N_v of v in the xy–plane as in Figure 11, and then restrict the horizontal vector field $\partial/\partial x$ to N_v. We extend the vector field on $\bigsqcup_v N_v$ thus obtained to a globally defined nonsingular vector field $\mathcal{X}_G$, so that outside $\bigsqcup_v N_v$, each trajectory of $\mathcal{X}_G$ is perpendicular to G.

Let $q_{\mathcal{X}_G}$ be the quadratic form on $H_1(\Sigma_{g,1}; \mathbb{Z}_2)$ corresponding to $w_{\mathcal{X}}$. Following Johnson [8], one can compute it as follows. Let γ be an oriented simple closed curve and consider its lift $\tilde{\gamma} = (\gamma, \dot{\gamma})$ to a loop in $UT\Sigma_{g,1}$ (here $\dot{\gamma}$ is the velocity vector of γ normalized to have unit length). Then

(A-1) $$q_{\mathcal{X}_G}([\gamma]) = \mathrm{wind}_{\mathcal{X}_G}(\tilde{\gamma}) + 1 \bmod 2.$$

We apply this formula to $\gamma = \hat{e}$, where $e \in E^{\mathrm{ori}}(G)$. Assume that e has the preferred orientation. Let $L(e)$ be the set of corners (f, f') of G (see the proof of Lemma 5.3) such that

(1) $e \preceq f \prec f' \preceq \overline{e}$, and

(2) exactly one of f and f' have the preferred orientation.

Here, $e \preceq f$ means $e \prec f$ or $e = f$. For example, if v is a vertex of type 1 as in the left part of Figure 11, only $(\overline{e}_2, e_3)$ is an element of $L(e)$ among the three corners around v. Set $\lambda(e) = \#L(e)$.

Lemma A.1 *We have* $\mathrm{wind}_{\mathcal{X}_G}(\widetilde{\widehat{e}}) = (1-\lambda(e))/2$.

Proof Take a small regular neighborhood $N(G)$ of G; we may arrange that $\widehat{e}$ stays inside $N(G)$ throughout. Every time when $\widehat{e}$ goes through a common vertex of a member of $L(e)$, the velocity vector of $\widehat{e}$ rotates by an angle $-\pi$ with respect to $\mathcal{X}_G$. Also, when $\widehat{e}$ goes through the middle point of e, the velocity vector of $\widehat{e}$ rotates by an angle π with respect to $\mathcal{X}_G$. This proves the lemma. □

In particular, using the fact that $\lambda(e)$ is odd (since e has the preferred orientation and $\overline{e}$ does not), we have from (A-1) that

$$q_{\mathcal{X}_G}(e) = q_{\mathcal{X}_G}([\widehat{e}]) = \tfrac{1}{2}(1-\lambda(e)) + 1 \bmod 2 = \tfrac{1}{2}(1+\lambda(e)) \bmod 2.$$

Proposition A.2 *Let G be a trivalent fatgraph spine of $\Sigma_{g,1}$. Then the quadratic forms $q_{\mathcal{X}_G}$, q_G and $\overline{q}_G$ are distinct from each other.*

Proof By Theorem 6.7, it is sufficient to prove $q_{\mathcal{X}_G} \neq q_G$ and $q_{\mathcal{X}_G} \neq \overline{q}_G$.

Let $e_1 \in E^{\mathrm{ori}}(G)$ be the "last" preferably oriented edge. Namely, e_1 is the unique element such that e_1 has the preferred orientation and if $e_1 \prec f$ then f does not have the preferred orientation. We have $\lambda(e_1) = 1$ and $q_{\mathcal{X}_G}(e_1) = (1+1)/2 = 1$. On the other hand, since there are no preferably oriented edges f with $e_1 \prec f \prec \overline{e}_1$, we have $q_G(e_1) = 0$. Hence $q_{\mathcal{X}_G} \neq q_G$.

Let $e_2 \in E^{\mathrm{ori}}(G)$ be the unique element such that e_2 has the preferred orientation and if $f \prec \overline{e}_2$ then f has the preferred orientation. We have $\lambda(e_2) = 1$ and $q_{\mathcal{X}_G}(e_2) = 1$. On the other hand, since any edge $f \in E^{\mathrm{ori}}(G)$ with $e_2 \prec f \prec \overline{e}_2$ has the preferred orientation, $\overline{q}_G(e_2) = 0$. Hence $q_{\mathcal{X}_G} \neq \overline{q}_G$. □

References

[1] **A J Bene**, **N Kawazumi**, **R C Penner**, *Canonical extensions of the Johnson homomorphisms to the Torelli groupoid*, Adv. Math. 221 (2009) 627–659 MR

[2] **B H Bowditch**, **D B A Epstein**, *Natural triangulations associated to a surface*, Topology 27 (1988) 91–117 MR

[3] **C J Earle**, *Families of Riemann surfaces and Jacobi varieties*, Ann. Math. 107 (1978) 255–286 MR

[4] **J L Harer**, *Stability of the homology of the mapping class groups of orientable surfaces*, Ann. of Math. 121 (1985) 215–249 MR

[5] **J L Harer**, *The virtual cohomological dimension of the mapping class group of an orientable surface*, Invent. Math. 84 (1986) 157–176 MR

[6] **J Harer**, **D Zagier**, *The Euler characteristic of the moduli space of curves*, Invent. Math. 85 (1986) 457–485 MR

[7] **D Johnson**, *An abelian quotient of the mapping class group* $\mathcal{I}_g$, Math. Ann. 249 (1980) 225–242 MR

[8] **D Johnson**, *Spin structures and quadratic forms on surfaces*, J. London Math. Soc. 22 (1980) 365–373 MR

[9] **D Johnson**, *A survey of the Torelli group*, from "Low-dimensional topology" (S J Lomonaco, Jr, editor), Contemp. Math. 20, Amer. Math. Soc., Providence, RI (1983) 165–179 MR

[10] **N Kawazumi**, *Canonical 2–forms on the moduli space of Riemann surfaces*, from "Handbook of Teichmüller theory, Volume II" (A Papadopoulos, editor), IRMA Lect. Math. Theor. Phys. 13, Eur. Math. Soc., Zürich (2009) 217–237 MR

[11] **N Kawazumi**, **S Morita**, *The primary approximation to the cohomology of the moduli space of curves and cocycles for the stable characteristic classes*, Math. Res. Lett. 3 (1996) 629–641 MR

[12] **M Kontsevich**, *Intersection theory on the moduli space of curves and the matrix Airy function*, Comm. Math. Phys. 147 (1992) 1–23 MR

[13] **Y Kuno**, **R C Penner**, **V Turaev**, *Marked fatgraph complexes and surface automorphisms*, Geom. Dedicata 167 (2013) 151–166 MR

[14] **G Massuyeau**, *Canonical extensions of Morita homomorphisms to the Ptolemy groupoid*, Geom. Dedicata 158 (2012) 365–395 MR

[15] **S Morita**, *Families of Jacobian manifolds and characteristic classes of surface bundles, I*, Ann. Inst. Fourier (Grenoble) 39 (1989) 777–810 MR

[16] **S Morita**, *Abelian quotients of subgroups of the mapping class group of surfaces*, Duke Math. J. 70 (1993) 699–726 MR

[17] **S Morita**, *The extension of Johnson's homomorphism from the Torelli group to the mapping class group*, Invent. Math. 111 (1993) 197–224 MR

[18] **S Morita**, **R C Penner**, *Torelli groups, extended Johnson homomorphisms, and new cycles on the moduli space of curves*, Math. Proc. Cambridge Philos. Soc. 144 (2008) 651–671 MR

[19] **R C Penner**, *The decorated Teichmüller space of punctured surfaces*, Comm. Math. Phys. 113 (1987) 299–339 MR

[20] **R C Penner**, *Perturbative series and the moduli space of Riemann surfaces*, J. Differential Geom. 27 (1988) 35–53 MR

[21] **R C Penner**, *Decorated Teichmüller theory of bordered surfaces*, Comm. Anal. Geom. 12 (2004) 793–820 MR

[22] **R C Penner**, *Decorated Teichmüller theory*, Eur. Math. Soc., Zürich (2012) MR

[23] **R C Penner**, **A M Zeitlin**, *Decorated super-Teichmüller space*, preprint (2015) arXiv

[24] **K Strebel**, *Quadratic differentials*, Ergeb. Math. Grenzgeb. 5, Springer, Berlin (1984) MR

Department of Mathematics, Tsuda College
2-1-1 Tsuda-Machi, Kodaira-shi, Tokyo 187-8577, Japan

kunotti@tsuda.ac.jp

Received: 25 May 2016 Revised: 20 October 2016

Algebraic & Geometric Topology 17 (2017) 1813–1870

The augmentation category map induced by exact Lagrangian cobordisms

YU PAN

To a Legendrian knot, one can associate an A_∞ category, the augmentation category. An exact Lagrangian cobordism between two Legendrian knots gives a functor of the augmentation categories of the two knots. We study this functor and establish a long exact sequence relating the corresponding cohomology of morphisms of the two ends. As applications, we prove that the functor between augmentation categories is injective on the level of equivalence classes of objects and find new obstructions to the existence of exact Lagrangian cobordisms in terms of linearized contact homology and ruling polynomials.

53D42, 57R17; 53D12, 57M50

1 Introduction

Let $\Lambda_\pm$ be Legendrian submanifolds in the standard contact manifold $(\mathbb{R}^3,\ \xi=\ker\alpha)$, where $\alpha = dz - y\,dx$. An exact Lagrangian cobordism Σ from Λ_- to Λ_+ is a 2–dimensional surface in the symplectization of $\mathbb{R}^3$ that has cylindrical ends over Λ_+ and Λ_- with certain properties. See Figure 1 for a schematic picture and Definition 2.1 for a detailed description.

Lagrangian cobordism is a natural relation between Legendrian submanifolds and is crucial in the definition of the functorial property of the Legendrian contact homology differential graded algebra (DGA). For a Legendrian knot Λ in $(\mathbb{R}^3,\ \xi=\ker\alpha)$, the Legendrian contact homology DGA is a powerful invariant of Λ that was introduced by Eliashberg [21] and Chekanov [7] in the spirit of symplectic field theory; see Eliashberg, Givental and Hofer [22]. The underlying algebra $\mathcal{A}(\Lambda;\mathbb{F}[H_1(\Lambda)])$ is a unital graded algebra freely generated by Reeb chords of Λ and a basis of $H_1(\Lambda)$ over a field $\mathbb{F}$, where $H_1(\Lambda)$ is the singular homology of Λ with $\mathbb{Z}$ coefficients. The differential ∂ is defined by a count of rigid holomorphic disks in $\mathbb{R}\times\mathbb{R}^3$ with boundary on the Lagrangian submanifold $\mathbb{R}\times\Lambda$. The DGA $(\mathcal{A}(\Lambda;\mathbb{F}[H_1(\Lambda)]),\partial)$ is invariant up to stable tame isomorphism under Legendrian isotopy of Λ. Ekholm, Honda and

Published: 17 July 2017 DOI: 10.2140/agt.2017.17.1813

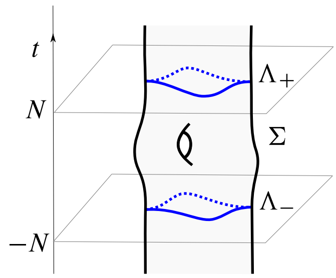

Figure 1: A Lagrangian cobordism Σ from Λ_- to Λ_+ lies in the symplectization of $\mathbb{R}^3$, which is $(\mathbb{R}_t \times \mathbb{R}^3, d(e^t\alpha))$. The vertical direction is the t direction and each horizontal plane is $\mathbb{R}^3$. Two Legendrian submanifolds Λ_+ and Λ_- sit inside different copies of $\mathbb{R}^3$ with t coordinates N and $-N$, respectively.

Kálmán [19] showed that an exact Lagrangian cobordism Σ from Λ_- to Λ_+ gives a DGA map

$$\phi_\Sigma\colon (\mathcal{A}(\Lambda_+;\mathbb{F}[H_1(\Sigma)]),\partial) \to (\mathcal{A}(\Lambda_-;\mathbb{F}[H_1(\Sigma)]),\partial),$$

which is defined by a count of rigid holomorphic disks in $\mathbb{R}\times\mathbb{R}^3$ as well, but with boundary on Σ. Here $\mathbb{F}$ can be any field if the cobordism Σ is spin. If the condition is not satisfied, the field $\mathbb{F}$ is assumed to be $\mathbb{Z}_2$.

Remark When Σ is spin, the boundary Legendrian knots Λ_+ and Λ_- get induced spin structure from the spin structure of Σ. This condition makes the moduli spaces of the holomorphic disks used in the DGA differentials and the DGA map equipped with a coherent orientation (following Ekholm, Etnyre and Sullivan [18]). In particular, when the dimension of a moduli space is 0, one can associate each rigid holomorphic disk in the moduli space with a sign. Therefore, we can count the disks with sign and get coefficients in any field $\mathbb{F}$. Otherwise, it is only reasonable to count the disks mod 2, which means ignoring the orientation. For the rest of the paper, we focus on the case where Σ is spin. If one is working on a nonspin cobordism, one can omit our description of orientation and get the corresponding statements for $\mathbb{F}=\mathbb{Z}_2$.

A fundamental question about Lagrangian cobordisms is: given two Legendrian knots Λ_+ and Λ_-, does there exist an exact Lagrangian cobordism Σ between them? In order to answer this question, we need to investigate the properties of Lagrangian cobordisms and obtain a relationship between Legendrian knots Λ_+ and Λ_-. If the two given Legendrian knots do not satisfy the desired relationship, there does not exist a cobordism between them. In this way, we can find obstructions to the existence of exact Lagrangian cobordisms. Chantraine [4] first gave a relationship between the Thurston–Bennequin numbers of two Legendrian knots:

$$\mathrm{tb}(\Lambda_+) - \mathrm{tb}(\Lambda_-) = -\chi(\Sigma). \tag{1}$$

This question was explored further in many works, including Bourgeois, Sabloff and Traynor [3], Sabloff and Traynor [34], Baldwin and Sivek [1], Cornwell, Ng and Sivek [11] and Chantraine, Dimitroglou Rizell, Ghiggini and Golovko [6] using generating families, normal rulings and Floer theory.

We approach the question through studying the relationship between the augmentation category of the Legendrian knots that are connected by an exact Lagrangian cobordism. Analogous to the derived Fukaya category of exact Lagrangian compact submanifolds introduced in Nadler and Zaslow [29], the augmentation category is an A_∞ category of Legendrian knots in $(\mathbb{R}^3, \ker\alpha)$. Bourgeois and Chantraine first introduced a nonunital A_∞ category in [2] and then Ng, Rutherford, Sivek, Shende and Zaslow introduced a unital version in [33]. We will focus on the latter one.

For a fixed DGA $(\mathcal{A}(\Lambda), \partial)$ of a Legendrian knot Λ, the augmentation category $\mathcal{A}\mathrm{ug}_+(\Lambda)$ consists of objects, morphisms and A_∞ operations. The objects in the category are augmentations ϵ of the Legendrian contact homology DGA, ie DGA maps $\epsilon\colon (\mathcal{A}(\Lambda), \partial) \to (\mathbb{F}, 0)$. For any two objects ϵ_1 and ϵ_2, the morphism space $\mathrm{Hom}_+(\epsilon_1, \epsilon_2)$ is a vector space over the field $\mathbb{F}$ generated by Reeb chords from Λ to Λ', where Λ' is a positive Morse perturbation of Λ. The A_∞ operations are composition maps $\{m_n \mid n \geq 1\}$ that satisfy certain relations. These relations allow us to take cohomology of the $\mathrm{Hom}_+(\epsilon_1, \epsilon_2)$ space with respect to m_1, denoted by $H^* \mathrm{Hom}_+(\epsilon_1, \epsilon_2)$. From [33], we know that up to A_∞ equivalence, the augmentation category $\mathcal{A}\mathrm{ug}_+(\Lambda)$ is an invariant of Legendrian knots under Legendrian isotopy.

We will show that an exact Lagrangian cobordism Σ from a Legendrian knot Λ_- to a Legendrian knot Λ_+ gives a DGA map ϕ_Σ from the DGA $\mathcal{A}(\Lambda_+; \mathbb{F}[H_1(\Lambda_+)])$ to the DGA $\mathcal{A}(\Lambda_-; \mathbb{F}[H_1(\Lambda_-)])$. By [33], this DGA map induces an A_∞–category map $f\colon \mathcal{A}\mathrm{ug}_+(\Lambda_-) \to \mathcal{A}\mathrm{ug}_+(\Lambda_+)$. As a result, the augmentation category $\mathcal{A}\mathrm{ug}_+$ acts functorially under Lagrangian cobordisms as well. For each augmentation ϵ_- of $\mathcal{A}(\Lambda_-)$, the cobordism Σ induces an augmentation ϵ_+ of $\mathcal{A}(\Lambda_+)$ by composing with the DGA map ϕ_Σ, ie

$$\epsilon_+ = \epsilon_- \circ \phi_\Sigma.$$

The augmentation category map f sends an object ϵ_- of $\mathcal{A}\mathrm{ug}_+(\Lambda_-)$ to the object ϵ_+ of $\mathcal{A}\mathrm{ug}_+(\Lambda_+)$. For any two objects ϵ_-^1 and ϵ_-^2 in $\mathcal{A}\mathrm{ug}_+(\Lambda_-)$, the category map f sends the morphism $\mathrm{Hom}_+(\epsilon_-^1, \epsilon_-^2)$ to the morphism $\mathrm{Hom}_+(\epsilon_+^1, \epsilon_+^2)$, where ϵ_+^1 and ϵ_+^2 are the augmentations induced by Σ.

We investigate properties of this A_∞–category map through the Floer theory of a pair of exact Lagrangian cobordisms (see Chantraine, Dimitroglou Rizell, Ghiggini and Golovko [6]), which is an analog of the construction of Ekholm [15] for a pair of Lagrangian fillings in the spirit of symplectic field theory (see Eliashberg, Givental and Hofer [22]). Let Σ be an exact Lagrangian cobordism from Λ_- to Λ_+. Perturb Σ using a positive Morse function F and get a new exact Lagrangian cobordism Σ'. In [6], Chantraine, Dimitroglou Rizell, Ghiggini and Golovko constructed a chain complex for this pair of exact Lagrangian cobordisms $\Sigma \cup \Sigma'$, called the Cthulhu chain complex. The generators of this chain complex are the union of double points of $\Sigma \cup \Sigma'$ and Reeb chords on the cylindrical ends from Σ to Σ'. Indeed, the second part agrees with the union of Hom_+ spaces in the augmentation category of the Legendrian submanifolds on two ends. The differential of this chain complex is defined by a count of rigid holomorphic disks with boundary on $\Sigma \cup \Sigma'$ as well. From [6], the Cthulhu chain complex is acyclic, which implies the following long exact sequence:

Theorem 1.1 (see Corollary 5.2) *Let Σ be an exact Lagrangian cobordism with Maslov number 0 from Λ_- to Λ_+. If ϵ_-^i, for $i = 1, 2$, is an augmentation of $\mathcal{A}(\Lambda_-)$ and ϵ_+^i is the augmentation of $\mathcal{A}(\Lambda_+)$ induced by Σ, then we have the following long exact sequence:*

$$\cdots \to H^k(\Sigma, \Lambda_-) \to H^k \operatorname{Hom}_+(\epsilon_+^1, \epsilon_+^2) \to H^k \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2) \\ \to H^{k+1}(\Sigma, \Lambda_-) \to \cdots.$$

If $\epsilon_-^1 = \epsilon_-^2 = \epsilon_-$, we can identify $H^k \operatorname{Hom}_+(\epsilon, \epsilon)$ with the linearized contact homology $\mathrm{LCH}^{\epsilon}_{1-k}(\Lambda)$ by Ng, Rutherford, Shende, Sivek and Zaslow [33, Section 5.2]. The long exact sequence above can be rewritten as

$$\cdots \to H^k(\Sigma, \Lambda_-) \to \mathrm{LCH}^{\epsilon_+}_{1-k}(\Lambda_+) \to \mathrm{LCH}^{\epsilon_-}_{1-k}(\Lambda_-) \to H^{k+1}(\Sigma, \Lambda_-) \to \cdots.$$

Computing the Euler characteristics of the exact triangle, we have

$$\mathrm{tb}(\Lambda_+) - \mathrm{tb}(\Lambda_-) = -\chi(\Sigma),$$

where $\chi(\Sigma)$ is the Euler characteristic of the surface Σ. This result was previously shown by Chantraine in [4].

Combine Theorem 1.1 with the augmentation category map induced by exact Lagrangian cobordisms and we have the following theorem.

Theorem 1.2 (see Theorem 5.4) *Let Σ be an exact Lagrangian cobordism with Maslov number 0 from a Legendrian knot Λ_- to a Legendrian knot Λ_+. Assume* *neither Λ_+ nor Λ_- are empty. For $i = 1, 2$, if ϵ^i_- is an augmentation of $\mathcal{A}(\Lambda_-)$ and ϵ^i_+ is the augmentation of $\mathcal{A}(\Lambda_+)$ induced by Σ, the map*

$$i^0\colon H^0 \operatorname{Hom}_+(\epsilon^1_+, \epsilon^2_+) \to H^0 \operatorname{Hom}_+(\epsilon^1_-, \epsilon^2_-)$$

in the long exact sequence in Theorem 1.1 is an isomorphism. Moreover, we have that

$$H^* \operatorname{Hom}_+(\epsilon^1_+, \epsilon^2_+) \cong H^* \operatorname{Hom}_+(\epsilon^1_-, \epsilon^2_-) \oplus \mathbb{F}^{-\chi(\Sigma)}[1],$$

where $\mathbb{F}^{-\chi(\Sigma)}[1]$ denotes the vector space $\mathbb{F}^{-\chi(\Sigma)}$ in degree 1 and $\chi(\Sigma)$ is the Euler characteristic of the surface Σ.

This relation was shown for positive braid closures in Menke [28]. Theorem 1.2 shows that this is true for general Legendrian knots.

When $\epsilon^1_- = \epsilon^2_-$, we restate Theorem 1.2 in terms of linearized contact homology as follows:

Corollary 1.3 (see Corollary 5.5) *Let Σ be an exact Lagrangian cobordism with Maslov number 0 from a Legendrian knot Λ_- to a Legendrian knot Λ_+. Assume neither Λ_+ nor Λ_- is empty. If ϵ_- is an augmentation of $\mathcal{A}(\Lambda_-)$ and ϵ_+ is the augmentation of $\mathcal{A}(\Lambda_+)$ induced by Σ, then*

$$\mathrm{LCH}^{\epsilon_+}_*(\Lambda_+) \cong \mathrm{LCH}^{\epsilon_-}_*(\Lambda_-) \oplus \mathbb{F}^{-\chi(\Sigma)}[0],$$

where $\mathbb{F}^{-\chi(\Sigma)}[0]$ denotes the vector space $\mathbb{F}^{-\chi(\Sigma)}$ in degree 0.

Therefore, if there exists an exact Lagrangian cobordism Σ from Λ_- to Λ_+, the Poincaré polynomials of the linearized contact homology of Λ_+ and Λ_- agree in all degrees except 0. In degree 0 their coefficients differ by $-\chi(\Sigma)$. This is a stronger obstruction to the existence of the exact Lagrangian cobordism than the Thurston–Bennequin number relation (1).

For instance, Figure 2 shows two Legendrian knots Λ_1 and Λ_2 of smooth knot types 4_1 and 6_1, respectively. There is a topological cobordism between 4_1 and 6_1 with genus 1. The Thurston–Bennequin numbers of Λ_1 and Λ_2 are -3 and -5, respectively, and thus satisfy the Thurston–Bennequin number relation (1). Therefore, there possibly exists an exact Lagrangian cobordism from Λ_2 to Λ_1 with genus 1. However, the Poincaré polynomials of the linearized contact homology for Λ_1 and Λ_2 are $t^{-1} + 2t$ and $2t^{-1} + 3t$, respectively. Thus, we have the following proposition.

Proposition 1.4 (see Proposition 5.6) *There does not exist an exact Lagrangian cobordism with Maslov number 0 from Λ_2 to Λ_1, where Λ_1 and Λ_2 are as shown in Figure 2.*

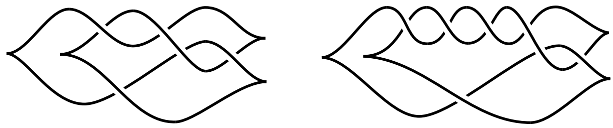

Figure 2: Legendrian knot Λ_1 of knot type 4_1 (left) and Legendrian knot Λ_2 of knot type 6_1 (right)

Various long exact sequences similar to that in Theorem 1.1 have been explored. Sabloff and Traynor [34] gave a long exact sequence using generating families. Chantraine, Dimitroglou Rizell, Ghiggini and Golovko gave three long exact sequences in [6] in the same spirit as this paper but use different Morse functions to perturb the cobordism. The way we construct the pair of cobordisms allows us to have more control over the behavior of the Morse function. This turns out to be a key point toward proving the following surprising theorem.

Theorem 1.5 (see Theorems 5.14 and 5.15) *Let Σ be an exact Lagrangian cobordism with Maslov number 0 from a Legendrian knot Λ_- to a Legendrian knot Λ_+. Then the A_∞–category map $f\colon \mathcal{A}\mathrm{ug}_+(\Lambda_-) \to \mathcal{A}\mathrm{ug}_+(\Lambda_+)$ induced by the exact Lagrangian cobordism Σ is injective on the level of equivalence classes of objects. In addition, the corresponding cohomology category map $\widetilde{f}\colon H^*\mathcal{A}\mathrm{ug}_+(\Lambda_-) \to H^*\mathcal{A}\mathrm{ug}_+(\Lambda_+)$ is faithful. In particular, when $\chi(\Sigma)=0$, the functor $\widetilde{f}$ is fully faithful.*

By Ng, Rutherford, Shende, Sivek and Zaslow [33], for a Legendrian knot with a single basepoint, two augmentations are equivalent if and only if they are isomorphic as DGA maps. This theorem tells us that the number of augmentations of Λ_- is smaller than or equal to the number of augmentations of Λ_+ up to equivalence. However, in general, it is hard to count the number of augmentations of Λ up to equivalence. Ng, Rutherford, Shende and Sivek [32] introduced a new way to count the augmentations, called the homotopy cardinality, which is related to the ruling polynomial. Recall that the ruling polynomial is defined by $R_\Lambda(z)=\sum_R z^{-\chi(R)}$, where the sum is over all normal rulings R of Λ (see Chekanov [8] for the detailed definition). This invariant is much easier to compute than the augmentation equivalence class. Using Theorem 1.5, we have the following corollary.

Corollary 1.6 (see Corollary 5.17) *Suppose there exists a spin exact Lagrangian cobordism with Maslov number 0 from a Legendrian knot Λ_- to a Legendrian knot Λ_+. Then the ruling polynomials R_{Λ_-} and R_{Λ_+} satisfy*

$$R_{\Lambda_-}(q^{1/2}-q^{-1/2}) \le q^{-\chi(\Sigma)/2} R_{\Lambda_+}(q^{1/2}-q^{-1/2})$$

for any q that is a power of a prime number.

This corollary gives a new obstruction to the existence of exact Lagrangian cobordisms. In particular, we have a new and simpler proof to the fact given by Chantraine [5] that there does not exist an exact Lagrangian cobordism from the Legendrian $m(9_{46})$ knot shown in Figure 18 to the Legendrian unknot. This fact is crucial to prove that Lagrangian concordance is not a symmetric relation.

Another important step toward proving the injectivity and faithfulness in Theorem 1.5 is to understand the differential map of the Cthulhu chain complex better. Analogous to a result for Legendrian submanifolds in Ekholm, Etnyre and Sabloff [16], we give a bijective correspondence between rigid holomorphic disks with boundary on a 2–copy of Σ and rigid holomorphic disks with boundary on Σ together with Morse flow lines. With this in hand, we can decompose the Cthulhu chain complex in various ways and recover the three long exact sequences in Chantraine, Dimitroglou Rizell, Ghiggini and Golovko [6].

Outline In Section 2, we review the Chekanov–Eliashberg DGA of a Legendrian submanifold and the DGA map induced by a Lagrangian cobordism. In Section 3, we introduce the augmentation category for a Legendrian submanifold and describe the A_∞–category map induced by an exact Lagrangian cobordism. In Section 4, we review the Floer theory of Lagrangian cobordisms. Finally, using the techniques in Section 4, we prove the main result Theorem 1.1 in Section 5 and discuss its applications.

Acknowledgements The author would like to thank Lenhard Ng for introducing the problem and many enlightening discussions. The author also thanks Baptiste Chantraine, John Etnyre and Michael Abel for helpful conversations, the referee for pointing out Theorem 5.12, Corollary 5.13 and Theorem 5.15, and Caitlin Leverson for comments on an earlier draft. This work was partially supported by NSF grants DMS-0846346 and DMS-1406371.

2 Legendrian contact homology DGA and exact Lagrangian cobordisms

2.1 The Legendrian contact homology DGA

In this section, we review the Legendrian contact homology DGA from the geometric perspective of [19] and the combinatorial perspective of [33, Section 2.2.1]. We refer readers to [7; 23; 30] for a more detailed introduction.

Let Λ be a Legendrian submanifold in the standard contact space $(\mathbb{R}^3, \xi = \ker\alpha)$, where $\alpha = dz - y\,dx$. For simplicity when defining the degree, we assume throughout the paper that Λ has rotation number 0.

Let $(\mathcal{A}(\Lambda;\mathbb{F}[H_1(\Lambda)]),\partial)$ denote the Legendrian contact homology DGA of Λ, which is also called the Chekanov–Eliashberg DGA. The underlying algebra $\mathcal{A}(\Lambda;\mathbb{F}[H_1(\Lambda)])$ is a noncommutative unital graded algebra over a field $\mathbb{F}$ generated by

$$\{c_1,\dots,c_m,t_1,t_1^{-1},\dots,t_M,t_M^{-1}\}$$

with relations $t_i t_i^{-1}=1$ for $i=1,\dots,M$. Here $c_1,\dots,c_m$ are Reeb chords of Λ and $\{t_1,\dots,t_M\}$ is a basis of the singular homology $H_1(\Lambda)$. The grading of a Reeb chord c is defined as

$$|c|=\mathrm{CZ}(\gamma_c)-1,$$

where γ_c is a capping path for c and CZ is the Conley–Zehnder index introduced in [17]. See [13, Section 4.1] for the way to choose a capping path γ_c for a Reeb chord of a Legendrian link. The grading of a Reeb chord depends on the choice of capping paths, but the difference between two Reeb chords' gradings is independent of the choice of capping paths. Furthermore, set the grading of t_i to be zero for $i=1,\dots,M$, and then extend the definition of degree to $\mathcal{A}(\Lambda;\mathbb{F}[H_1(\Lambda)])$ through the relation $|ab|=|a|+|b|$.

To define the differential ∂, we need a *cylindrical almost complex structure* J on $(\mathbb{R}\times\mathbb{R}^3,d(e^t\alpha))$, ie

- J is compatible with the symplectic form $d(e^t\alpha)$;
- J is invariant under the action of $\mathbb{R}_t$;
- $J(\partial_t)=\partial_z$ and $J(\xi)=\xi$.

For a generic choice of cylindrical almost complex structure J, the differential ∂ is defined by counting rigid J–holomorphic disks in $(\mathbb{R}_t\times\mathbb{R}^3,d(e^t\alpha))$ with boundary on $\mathbb{R}\times\Lambda$. See Figure 3 for an example. For Reeb chords $a,b_1,\dots,b_m$ of Λ, let $\mathcal{M}(a;b_1,\dots,b_m)$ denote the moduli space of J–holomorphic disks

$$u\colon (D_{m+1},\partial D_{m+1})\to(\mathbb{R}\times\mathbb{R}^3,\mathbb{R}\times\Lambda)$$

such that

- D_{m+1} is a 2–dimensional unit disk with $m+1$ boundary points $p,q_1,\dots,q_m$ removed and the points $p,q_1,\dots,q_m$ are labeled in a counterclockwise order;
- u is asymptotic to $[0,\infty)\times a$ at p;
- u is asymptotic to $(-\infty,0]\times b_i$ at q_i.

Let $\widetilde{\mathcal{M}}(a;b_1,\dots,b_m)$ denote the quotient of $\mathcal{M}(a;b_1,\dots,b_m)$ by vertical translation of $\mathbb{R}_t$. When $\dim\widetilde{\mathcal{M}}(a;b_1,\dots,b_m)=0$, the disk $u\in\mathcal{M}(a;b_1,\dots,b_m)$ is called *rigid*. The gradings of corresponding Reeb chords satisfy

$$|a|-|b_1|-\cdots-|b_m|=1.$$

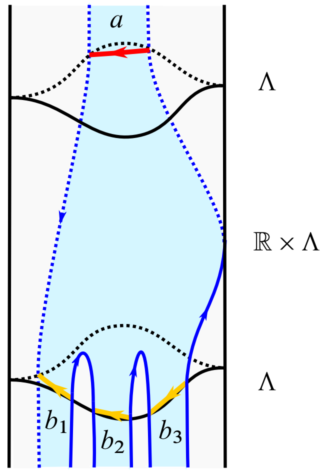

Figure 3: An example of a J–holomorphic disk with boundary on $\mathbb{R}\times\Lambda$. The arrows on the Reeb chords indicate the orientations of the Reeb chords, while the arrows on the disk boundary indicate the orientation inherited from the unit disk boundary with counterclockwise orientation through u.

For the image of the boundary segment from q_i to q_{i+1} under u, one can close it up on $\mathbb{R}\times\Lambda$ in a particular way as described in [19, Section 3.2] and take the homology class of this curve in $H_1(\Lambda)$, denoted by τ_i. Here we use $q_0=q_{m+1}=p$. Moreover, if Λ is spin, all the relevant moduli spaces of J–holomorphic disks admit a coherent orientation. Hence, one can associate a sign $s(u)$ to each rigid J–holomorphic disk u. In this way, associate the rigid J–holomorphic disk u with a monomial

$$w(u)=s(u)\tau_0 b_1\tau_1\cdots b_m\tau_m.$$

We call the homology classes τ_i, for $i=1,\dots,m$, the *coefficients* of $w(u)$. The *differential* on Reeb chords is defined by counting rigid J–holomorphic disks:

$$\partial a=\sum_{\dim\widetilde{\mathcal{M}}(a;b_1,\dots,b_m)=0}\ \sum_{u\in\mathcal{M}(a;b_1,\dots,b_m)} w(u).$$

Let $\partial t_i=\partial t_i^{-1}=0$ for $i=1,\dots,M$ and extend the differential to $\mathcal{A}(\Lambda;\mathbb{F}[H_1(\Lambda)])$ through the Leibniz rule

$$\partial(xy)=(\partial x)y+(-1)^{|x|}x(\partial y).$$

An implicit condition for J–holomorphic disks is the positive energy constraint. For a Reeb chord c, define the action of c by

$$\mathfrak{a}(c)=\int_c\alpha,$$

which is the length of the Reeb chord c. The energy $E(u)$ of a J–holomorphic disk $u \in \mathcal{M}(a; b_1, \dots, b_m)$ satisfies

$$E(u) = \mathfrak{a}(a) - \mathfrak{a}(b_1) - \cdots - \mathfrak{a}(b_m).$$

Therefore, to make each J–holomorphic disk have positive energy, we must have

$$\mathfrak{a}(b_1) + \cdots + \mathfrak{a}(b_m) < \mathfrak{a}(a).$$

There is an equivalent definition from the combinatorial perspective. Project Λ onto the xy–plane to get the Lagrangian projection $\pi_{xy}(\Lambda)$ of Λ. After possibly perturbing Λ, we can assume that there is a one-to-one correspondence between the double points of $\pi_{xy}(\Lambda)$ and the Reeb chords of Λ. Suppose Λ is an M–component Legendrian link. Decorate the diagram with an orientation and a set of *minimum basepoints* $\{*_1, \dots, *_M\}$, ie

- there is exactly one point in $\{*_1, \dots, *_M\}$ on each component of Λ, and
- the set $\{*_1, \dots, *_M\}$ does not include any end points of Reeb chords of Λ.

The graded algebra $\mathcal{A}(\Lambda, *_1, \dots, *_M)$ is a noncommutative unital graded algebra over a field $\mathbb{F}$ generated by

$$\{c_1, \dots, c_m, t_1, t_1^{-1}, \dots, t_M, t_M^{-1}\} \quad \text{with relations} \quad \{t_i t_i^{-1} = 1 \mid i = 1, \dots, M\},$$

where $c_1, \dots, c_m$ are double points of $\pi_{xy}(\Lambda)$ and $t_1, \dots, t_M$ correspond to the basepoints $*_1, \dots, *_M$. The grading is defined the same as above. For the unit disk D_{m+1} as defined above, consider $\Delta(a; b_1, \dots, b_m)$, the space of orientation-preserving smooth immersions up to parametrization

$$u\colon (D_{m+1}, \partial D_{m+1}) \to (\mathbb{R}^2, \pi_{xy}(\Lambda))$$

with the following properties:

- u can be extended to the unit disk $\overline{D_{m+1}}$ continuously.
- $u(p) = a$ and the neighborhood of a in the image of u is a single positive quadrant (see Figure 4).
- $u(q_i) = b_i$ and the neighborhood of b_i in the image of u is a single negative quadrant for $1 \le i \le m$ (see Figure 4).

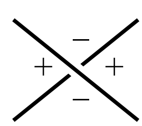

Figure 4: At each crossing, the quadrants labeled with a $+$ sign are called *positive quadrants* while the other two are called *negative quadrants*.

If, when traversing $\overline{\partial D_{m+1}}$ counterclockwise from a, one encounters Reeb chords and basepoints in a sequence $s_1, \dots, s_l$, then we associate u with a monomial $w(u) = s(u)w(s_1)\cdots w(s_l)$, where:

- $s(u)$ is the sign associated to the disk u induced from the moduli space coherent orientation.
- If s_i is a Reeb chord b_j, then $w(s_i) = b_j$.
- If s_i is a basepoint $*_j$ for some $j = 1, \dots, M$, then $w(s_i) = t_j$ if the orientation of the boundary agrees with the orientation of the link and $w(s_i) = t_j^{-1}$ otherwise.

Define the *differential* on generators as follows:

$$\partial a = \sum_{|a|-\sum|b_i|=1} \sum_{u\in\Delta(a;b_1,\dots,b_m)} w(u) \qquad \text{and} \qquad \partial t_j = \partial t_j^{-1} = 0 \quad \text{for } j = 1, \dots, M.$$

This can be extended to the whole DGA through the Leibniz rule.

For all the definitions of DGAs $(\mathcal{A}, \partial)$ above, the differential ∂ has degree -1 and satisfies $\partial^2 = 0$ [7; 23]. Up to stable tame isomorphism, the Legendrian contact homology DGA is an invariant of Λ under Legendrian isotopy. In this sense of equivalence, the combinatorial definition does not depend on the choice of basepoints [31].

However, the homology of the DGA is hard to compute in general. Let us introduce augmentations of a DGA and use that to deduce linearized contact homology, which is much easier to compute. Let $(\mathcal{A}, \partial)$ be a DGA over a field $\mathbb{F}$ of a Legendrian link Λ with basepoints. A *graded augmentation* of $\mathcal{A}$ is a DGA map

$$\epsilon\colon (\mathcal{A}, \partial) \to (\mathbb{F}, 0),$$

where $(\mathbb{F}, 0)$ is a chain complex that is $\mathbb{F}$ in degree 0 and is 0 in other degrees. In other words, a graded augmentation is an algebra map $\epsilon\colon \mathcal{A} \to \mathbb{F}$ such that $\epsilon(1) = 1$, $\epsilon \circ \partial = 0$ and $\epsilon(a) = 0$ if $|a| \neq 0$.

Given a graded augmentation ϵ, define $\mathcal{A}^\epsilon := \mathcal{A} \otimes \mathbb{F}/(t_i = \epsilon(t_i))$. Notice that the differential ∂ descends to $\mathcal{A}^\epsilon$ since $\partial(t_i) = 0$. Elements in $\mathcal{A}^\epsilon$ are summands of words of Reeb chords. Let C be a free $\mathbb{F}$–module generated by Reeb chords. We can decompose $\mathcal{A}^\epsilon$ in terms of word length as $\mathcal{A}^\epsilon = \bigoplus_{n\geq 0} C^{\otimes n}$. Let $\mathcal{A}^\epsilon_+$ be the part of $\mathcal{A}^\epsilon$ containing the words with length at least 1, ie $\mathcal{A}^\epsilon_+ = \bigoplus_{n\geq 1} C^{\otimes n}$. Consider a new differential $\partial^\epsilon\colon \mathcal{A}^\epsilon \to \mathcal{A}^\epsilon$ given by

$$\partial^\epsilon := \phi_\epsilon \circ \partial \circ \phi_\epsilon^{-1},$$

where $\phi_\epsilon\colon \mathcal{A}^\epsilon \to \mathcal{A}^\epsilon$ is an automorphism defined by $\phi_\epsilon(a) = a + \epsilon(a)$. Observe that ∂^ϵ preserves $\mathcal{A}^\epsilon_+$ and does not decrease the minimal length of a word. Thus, it descends to a differential on $\mathcal{A}^\epsilon_+/(\mathcal{A}^\epsilon_+)^2 \cong C$. The homology of (C, ∂^ϵ) is called the *linearized contact homology* of Λ with respect to ϵ, denoted by $\mathrm{LCH}^{\epsilon}_{*}(\Lambda)$. The chain complex (C, ∂^{ϵ}) is called the *linearized contact homology chain complex*.

2.2 Exact Lagrangian cobordisms

We now review the DGA map induced by an exact Lagrangian cobordism [19] with coefficients in general fields (following the orientation convention of [18]). In other words, an exact Lagrangian cobordism Σ from Λ_- to Λ_+ gives a DGA map

$$\phi_\Sigma\colon \mathcal{A}(\Lambda_+;\mathbb{F}[H_1(\Sigma)]) \to \mathcal{A}(\Lambda_-;\mathbb{F}[H_1(\Sigma)]).$$

As required in Section 3, we restrict to the case where Λ_+ and Λ_- are Legendrian knots with a single basepoint, denoted by $*_+$ and $*_-$, respectively. We modify the DGA map such that the coefficients only depend on the basepoints but not depend on the cobordism, ie we get a DGA map

$$\phi_\Sigma\colon \mathcal{A}(\Lambda_+, *_+) \to \mathcal{A}(\Lambda_-, *_-).$$

Definition 2.1 Suppose $\Lambda_\pm$ are Legendrian submanifolds in $(\mathbb{R}^3, \ker\alpha)$, where $\alpha = dz - y\,dx$. An *exact Lagrangian cobordism* Σ from Λ_- to Λ_+ is a 2–dimensional surface in $(\mathbb{R}\times\mathbb{R}^3, \omega = d(e^t\alpha))$ (see Figure 1) such that for some big number $N > 0$,

- $\Sigma \cap ((N,\infty)\times\mathbb{R}^3) = (N,\infty)\times\Lambda_+$,
- $\Sigma \cap ((-\infty,-N)\times\mathbb{R}^3) = (-\infty,-N)\times\Lambda_-$, and
- $\Sigma \cap ([-N,N]\times\mathbb{R}^3)$ is compact.

Moreover, there exists a smooth function $g\colon \Sigma \to \mathbb{R}$ such that

$$e^t\alpha\,|_{T\Sigma} = dg$$

and g is constant when $t \le -N$ and $t \ge N$. The function g is called a *primitive* of Σ.

For a spin exact Lagrangian cobordism Σ from Λ_- to Λ_+, the Legendrian submanifolds $\Lambda_\pm$ inherit induced spin structures. Hence $\Lambda_\pm$ have $\mathbb{F}[H_1(\Lambda_\pm)]$–coefficients DGAs $(\mathcal{A}(\Lambda_\pm;\mathbb{F}[H_1(\Lambda_\pm)]), \partial)$, respectively, as described in Section 2.1. Ekholm, Honda and Kálmán in [19] showed that an exact Lagrangian cobordism Σ induces a DGA map from $\mathcal{A}(\Lambda_+)$ to $\mathcal{A}(\Lambda_-)$ with $\mathbb{F}[H_1(\Sigma)]$ coefficients. In order to see that, first, we need to view the DGAs of $\Lambda_\pm$ as DGAs with $\mathbb{F}[H_1(\Sigma)]$ coefficients. Notice that the inclusion $H_1(\Lambda_\pm) \hookrightarrow H_1(\Sigma)$ induces a canonical inclusion map $\mathbb{F}[H_1(\Lambda_\pm)] \hookrightarrow \mathbb{F}[H_1(\Sigma)]$ of the group ring coefficients, which makes it natural to consider the DGAs of $\Lambda_\pm$ with $\mathbb{F}[H_1(\Sigma)]$ coefficients. Specifically, the new DGA $\mathcal{A}(\Lambda_\pm;\mathbb{F}[H_1(\Sigma)])$ is generated by Reeb chords of $\Lambda_\pm$ and elements in $H_1(\Sigma)$ over $\mathbb{F}$. The differential is defined by the original differential in $\mathcal{A}(\Lambda_\pm;\mathbb{F}[H_1(\Lambda_\pm)])$ composed with the inclusion map $H_1(\Lambda_\pm) \hookrightarrow H_1(\Sigma)$.

Second, construct a DGA map with $\mathbb{F}[H_1(\Sigma)]$ coefficients. Consider an almost complex structure J that is compatible with the symplectic form ω and is cylindrical on both ends. In other words, J matches the cylindrical almost complex structures on both cylindrical ends. Fix a generic choice of such an almost complex structure J. For Reeb chords a of Λ_+ and $b_1, \dots, b_m$ of Λ_-, define $\mathcal{M}(a; b_1, \dots, b_m)$ to be the moduli space of the J–holomorphic disks

$$u\colon (D_{m+1}, \partial D_{m+1}) \to (\mathbb{R} \times \mathbb{R}^3, \Sigma)$$

such that

- D_{m+1} is a 2–dimensional unit disk with $m+1$ boundary points $p, q_1, \dots, q_m$ removed and the points $p, q_1, \dots, q_m$ are arranged in a counterclockwise order;
- u is asymptotic to $[N, \infty) \times a$ at p;
- u is asymptotic to $(-\infty, -N] \times b_i$ at q_i.

When $\dim \mathcal{M}(a; b_1, \dots, b_m) = 0$, the disk $u \in \mathcal{M}(a; b_1, \dots, b_m)$ is called *rigid*. The gradings of corresponding Reeb chords satisfy

$$|a| - |b_1| - \cdots - |b_m| = 0.$$

For the image of the boundary segment from q_i to q_{i+1}, one can close up in a similar way as the one in the definition of the DGA differential and take the homology class in $H_1(\Sigma)$, denoted by τ_i. If Σ is spin, all the relevant moduli spaces of J–holomorphic disks admit a coherent orientation. In particular, each rigid J–holomorphic disk obtains a sign, denoted by $s(u)$. Associate a monomial $w(u)$ to the J–holomorphic disk u as

$$w(u) = s(u)\tau_0 b_1 \tau_1 \cdots b_m \tau_m.$$

The homology classes τ_i, for $i = 1, \dots, m$, are called the *coefficients* of $w(u)$. The DGA map is defined by counting rigid J–holomorphic disks with boundary on Σ:

$$\phi(a) = \sum_{\dim \mathcal{M}(a;b_1,\dots,b_m)=0} \; \sum_{u \in \mathcal{M}(a;b_1,\dots,b_m)} w(u).$$

We can extend the morphism to $\mathcal{A}(\Lambda_+; \mathbb{F}[H_1(\Sigma)])$ by setting $\phi(t) = t$ for any generator t in $H_1(\Sigma)$ and applying the Leibniz rule.

In order to modify the coefficients of the DGA map ϕ, let us consider $H_1(\Sigma)$ more precisely. To simplify the description, we restrict Σ to $[-N, N] \times \mathbb{R}^3$ and denote it by Σ as well. According to Poincaré duality, $H^1(\Sigma) \cong H_1(\Sigma, \Lambda_+ \cup \Lambda_-)$. In particular, for any loop α in Σ with ends on $\Lambda_+ \cup \Lambda_-$ which is an element in $H_1(\Sigma, \Lambda_+ \cup \Lambda_-)$ there is an element θ_α in $H^1(\Sigma)$ such that for any oriented loop γ on Σ, the intersection number of α and γ is $\theta_\alpha(\gamma)$. Thus, in order to know the homology class of a curve γ in $H_1(\Sigma)$, we only need to count the intersection number of each generator curve of $H_1(\Sigma, \Lambda_+ \cup \Lambda_-)$ with γ.

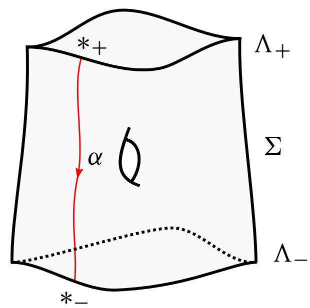

Figure 5: Curve α on a cobordism

Consider a connected exact Lagrangian cobordism Σ from a Legendrian knot Λ_- to a Legendrian knot Λ_+ (see Remark 5.3 for the reason that we assume that Σ is connected). Choose basepoints $*_+$ and $*_-$ for Λ_+ and Λ_-, respectively. There exists a curve α on Σ from $*_+$ to $*_-$ with exactly one intersection with Λ_+ and Λ_-, respectively. An example is shown in Figure 5. Let V^* denote the subgroup of $H^1(\Sigma)$ that is generated by the Poincaré dual of curve α. The dual space V in $H_1(\Sigma)$ is isomorphic to $\mathbb{Z}$.

Now we can modify the DGA map ϕ described above to be a map from $\mathcal{A}(\Lambda_+; \mathbb{F}[V])$ to $\mathcal{A}(\Lambda_-; \mathbb{F}[V])$. First, restrict the generators of $\mathcal{A}(\Lambda_\pm)$ to Reeb chords of $\Lambda_\pm$ and a basis of V. Second, project the coefficients τ_i of the monomial $w(u)$ from $H_1(\Sigma)$ to V. Therefore, the DGA map works in $\mathbb{F}[V]$ coefficients. Indeed, the definitions of $\mathcal{A}(\Lambda_\pm; \mathbb{F}[V])$ match the definition of $\mathcal{A}(\Lambda_\pm, *_\pm)$, respectively. Hence a connected exact Lagrangian cobordism Σ induces a DGA map with $\mathbb{F}[V]$ coefficients from the DGA of Λ_+ with a single basepoint to the DGA of Λ_- with a single basepoint:

$$\phi\colon (\mathcal{A}(\Lambda_+, *_+), \partial) \to (\mathcal{A}(\Lambda_-, *_-), \partial).$$

This DGA map does depend on the choice of the curve α connecting the two basepoints.

3 The augmentation category

3.1 A_∞ categories

In this section, we give a lightning review of A_∞ algebras and A_∞ categories, following [33]. See [26; 25] for a more detailed introduction.

Definition 3.1 [26, Section 3.1] An A_∞ *algebra* over a field $\mathbb{F}$ is a $\mathbb{Z}$–graded vector space A endowed with degree-$(2-n)$ maps $m_n\colon A^{\otimes n} \to A$ such that

$$\sum_{r+s+t=n} (-1)^{r+st} m_{r+1+t}(1^{\otimes r} \otimes m_s \otimes 1^{\otimes t}) = 0.$$

The most important things we need among these complicated relations are:

- m_1 is a differential on A (ie $m_1^2 = 0$).
- m_2 is associative after passing to the homology with respect to m_1.

An A_∞ algebra can be achieved nicely through the following construction. Let $\overline{T}(C) = \bigoplus_{n\geq 1} C^{\otimes n}$ be a graded vector space over $\mathbb{F}$ equipped with a *codifferential* b, ie

- b has degree 1,
- $b^2 = 0$,
- $b = \bigoplus b_n$, where b_n is a map $C^{\otimes n} \to C$, and
- b satisfies the co-Leibniz rule
$$\Delta b = (1 \otimes b + b \otimes 1)\Delta,$$
where $\Delta(a_1 \otimes \cdots \otimes a_n) = \sum_{i=1}^n (a_1 \otimes \cdots \otimes a_i) \otimes (a_{i+1} \otimes \cdots \otimes a_n)$.

Let $C^\vee := C[-1]$ and let $s\colon C \to C^\vee$ be the canonical degree-1 identification map $a \mapsto a$. Define maps $m_n\colon (C^\vee)^{\otimes n} \to C^\vee$ such that the following diagram commutes for all n:

$$\begin{tikzcd} C^{\otimes n} \arrow[r, "b_n"] \arrow[d, "s^{\otimes n}"'] & C \arrow[d, "s"] \\ (C^\vee)^{\otimes n} \arrow[r, "m_n"'] & C^\vee \end{tikzcd}$$

Then $C^\vee$ is an A_∞ algebra with the m_n as A_∞ operations [36; 37]. One can check that the degree of m_n is $2-n$.

Example 3.2 If a Legendrian contact homology DGA $(\mathcal{A}(\Lambda), \partial)$ has an augmentation ϵ, the conjugated differential ∂^ϵ is a differential of $\mathcal{A}_+^\epsilon = \bigoplus_{n\geq 1} C^{\otimes n} = \overline{T}(C)$, where C is the vector space over a field $\mathbb{F}$ generated by Reeb chords of Λ. We define δ^ϵ to be the adjoint of ∂^ϵ on $\overline{T}(C^*) = \bigoplus_{n\geq 1} (C^*)^{\otimes n}$, where C^* is the dual of C. More specifically,

$$\delta^\epsilon(b_m^* \otimes \cdots \otimes b_1^*) = \sum_a \mathrm{Coeff}_{b_1 b_2 \cdots b_m}(\partial^\epsilon(a)).$$

It is not hard to check that δ^ϵ is a codifferential of $\overline{T}(C^*)$. Hence one can use the construction above to construct an A_∞ algebra $(C^*)^\vee$.

Definition 3.3 [10] An A_∞ *category* over a field $\mathbb{F}$ is a category where, for any two objects ϵ_1 and ϵ_2, the morphism is a graded vector space $\operatorname{Hom}(\epsilon_1,\epsilon_2)$. Moreover, for any objects $\epsilon_1,\epsilon_2,\dots,\epsilon_{n+1}$, there exists a degree-$(2-n)$ map

$$m_n\colon \operatorname{Hom}(\epsilon_n,\epsilon_{n+1})\otimes\cdots\otimes\operatorname{Hom}(\epsilon_1,\epsilon_2)\to\operatorname{Hom}(\epsilon_1,\epsilon_{n+1})$$

satisfying

$$\sum_{r+s+t=n}(-1)^{r+st}m_{r+1+t}(1^{\otimes r}\otimes m_s\otimes 1^{\otimes t})=0.$$

As noted before, the first A_∞ operation m_1 is a differential for $\operatorname{Hom}(\epsilon_1,\epsilon_2)$ with degree 1. Denote its cohomology by $H^*\operatorname{Hom}(\epsilon_1,\epsilon_2)$. Moreover, we have that m_2 descends to an associative map on the cohomology level:

$$m_2\colon H^*\operatorname{Hom}(\epsilon_2,\epsilon_3)\otimes H^*\operatorname{Hom}(\epsilon_1,\epsilon_2)\to H^*\operatorname{Hom}(\epsilon_1,\epsilon_3)$$

for any objects ϵ_1, ϵ_2, ϵ_3.

An A_∞ *morphism* between two A_∞ categories $f\colon\mathcal{A}\to\mathcal{B}$ maps the object ϵ of $\mathcal{A}$ to $f(\epsilon)$ of $\mathcal{B}$ and for any objects $\epsilon_1,\epsilon_2,\dots,\epsilon_{n+1}$ of $\mathcal{A}$, there exists a map

$$f_n\colon \operatorname{Hom}(\epsilon_n,\epsilon_{n+1})\otimes\cdots\otimes\operatorname{Hom}(\epsilon_1,\epsilon_2)\to\operatorname{Hom}(f(\epsilon_1),f(\epsilon_{n+1}))$$

satisfying the A_∞ relations [26]. In particular, the first map f_1, called *the category map on the level of morphisms*, maps the morphism $\operatorname{Hom}(\epsilon_1,\epsilon_2)$ of $\mathcal{A}$ to the morphism $\operatorname{Hom}(f(\epsilon_1),f(\epsilon_2))$ of $\mathcal{B}$. From the A_∞ relations, we know that:

- The functor f_1, the category map on the level of morphisms, commutes with m_1 and thus f_1 descends to a map on cohomology
$$f^*\colon H^*\operatorname{Hom}(\epsilon_1,\epsilon_2)\to H^*\operatorname{Hom}(f(\epsilon_1),f(\epsilon_2)).$$
- For any $a\in\operatorname{Hom}(\epsilon_2,\epsilon_3)$ and $b\in\operatorname{Hom}(\epsilon_1,\epsilon_2)$, we have
$$f^*(m_2([a],[b]))=m_2(f^*[a],f^*[b]),$$
ie, the composition map m_2 commutes with f^* when passing to the cohomology level.

An A_∞ morphism between two A_∞ categories $f\colon\mathcal{A}\to\mathcal{B}$ induces a functor on the cohomology categories, $\widetilde{f}\colon H^*\mathcal{A}\to H^*\mathcal{B}$. It behaves the same as f on the object level. On the level of morphisms $\widetilde{f}=f^*$. The functor $\widetilde{f}$ is *faithful* if f^* is injective and is *fully faithful* if f^* is an isomorphism for any morphism in $H^*\mathcal{A}$.

3.2 The augmentation category

In this section, we briefly review the augmentation category $\mathcal{A}\mathrm{ug}_+(\Lambda)$, following [33].

Let Λ be an oriented Legendrian knot in $(\mathbb{R}^3, \ker\alpha)$ endowed with a single basepoint $*$. Denote its Legendrian contact homology DGA by $(\mathcal{A}, \partial)$. Given a field $\mathbb{F}$, the objects of the augmentation category $\mathcal{A}\mathrm{ug}_+(\Lambda)$ are augmentations of $(\mathcal{A}, \partial)$ to $\mathbb{F}$,

$$\epsilon\colon \mathcal{A} \to \mathbb{F}.$$

In order to describe the morphism $\mathrm{Hom}_+(\epsilon_1, \epsilon_2)$ for any two objects ϵ_1 and ϵ_2, we need to study the DGA of a 2–copy of Λ, denoted by $\Lambda^{(2)}$.

By the Weinstein tubular neighborhood theorem, we can identify a neighborhood of Λ with a neighborhood of the zero section in the 1–jet space $J^1(\Lambda) = T^*(\Lambda) \times \mathbb{R}$ through a contactomorphism. The contact form in $J^1(\Lambda)$ is $\alpha = dz - p\,dq$, where q is the coordinate on Λ and p is the coordinate in the cotangent direction. For any C^1 small function $f\colon \Lambda \to \mathbb{R}$, the 1–jet $j^1 f = \{(q, f'(q), f(q)) \mid q \in \Lambda\}$ is a Legendrian knot in $J^1(\Lambda)$ and thus is a Legendrian knot in $\mathbb{R}^3$. Now choose a particular Morse function $f\colon \Lambda \to (0, \delta)$ such that

- δ is smaller than the minimum length of Reeb chords of Λ,
- the Morse function f has exactly 1 local maximum point at x and 1 local minimum point at y, and
- around the basepoint $*$, the three points $*$, x, y show up in order when traveling along the link (see Figure 6).

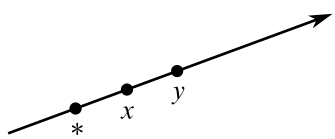

Figure 6: A neighborhood of the basepoint $*$ on Λ. The arrow indicates the orientation of Λ.

Decorate $j^1 f$ with a basepoint in the same location and with the same orientation as Λ. Now $\Lambda \cup j^1 f$ is a 2–copy of Λ, denoted by $\Lambda^{(2)}$. Label $\Lambda^{(2)}$ from top (higher z coordinate) to bottom (lower z coordinate) by Λ^1 and Λ^2. An example of the 2–copy of the trefoil with a single basepoint is shown in Figure 7.

The Legendrian contact homology DGA $(\mathcal{A}(\Lambda^{(2)}), \partial^{(2)})$ of $\Lambda^{(2)}$ can be recovered from the data carried by the DGA $(\mathcal{A}(\Lambda), \partial)$ of Λ. Recall that $\mathcal{A}(\Lambda)$ is generated by the set $\mathcal{R}$ of Reeb chords $\{a_1, \dots, a_m\}$ and the set $\mathcal{T} = \{t, t^{-1}\}$ that corresponds to the basepoint as stated in Section 2.1. Similarly, divide the set of generators of $\mathcal{A}(\Lambda^{(2)})$ into two parts $\mathcal{R}^{(2)}$ and $\mathcal{T}^{(2)}$. It is obvious that $\Lambda^{(2)}$ has two basepoints, and thus we write $\mathcal{T}^{(2)}$ as $\{(t^1)^{\pm 1}, (t^2)^{\pm 1}\}$. As for the set of Reeb chords $\mathcal{R}^{(2)}$, we divide it into four parts: $\mathcal{R}^{(2)} = \bigcup_{i,j=1,2} \mathcal{R}^{ij}$, where $\mathcal{R}^{ij}$ is the set of Reeb chords to Λ_i from Λ_j.

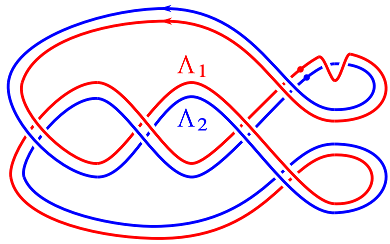

Figure 7: The Lagrangian projection of a 2–copy of the trefoil with a single basepoint

Observe that Reeb chords of $\Lambda^{(2)}$ come from two sources:

- Each critical point x or y of the Morse function f gives one Reeb chord in $\mathcal{R}^{12}$, denoted by x^{12} and y^{12}, respectively. We call these Reeb chords *Morse Reeb chords*.
- Each Reeb chord a_l of Λ gives four Reeb chords of $\Lambda^{(2)}$, denoted by $a_l^{ij} \in \mathcal{R}^{ij}$, where $i, j = 1, 2$ and $l = 1, \dots, m$. We call these Reeb chords *non-Morse Reeb chords*.

It is obvious that a^{ii} and t^i, for $i = 1, 2$, inherit the grading from a and t in $\mathcal{A}(\Lambda)$, respectively. We can choose a family of capping paths such that $|a^{ij}| = |a|$ for any Reeb chord a of Λ. Under this choice of capping paths γ, one can show that $\mathrm{CZ}(\gamma_{x^{12}}) = \mathrm{Ind}_f(x)$ for any Morse Reeb chord x^{12} through a computation similar to that in [16]. Hence we have $|x^{12}| = 0$ and $|y^{12}| = -1$.

In order to describe the differential $\partial^{(2)}$, we encode the generators in matrices. Let A_l, for $1 \le l \le m$, and X, Y, Δ be 2×2 matrices given by

$$A_l = \begin{pmatrix} a_l^{11} & a_l^{12} \\ a_l^{21} & a_l^{22} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x^{12} \\ 0 & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & y^{12} \\ 0 & 0 \end{pmatrix}, \quad \Delta = \begin{pmatrix} t^1 & 0 \\ 0 & t^2 \end{pmatrix}.$$

The differential $\partial^{(2)}$ is defined on generators as follows by applying it entry-by-entry to these matrices:

$$\begin{aligned} \partial^{(2)} A_l &= \Phi(\partial a_l) + YA_l - (-1)^{|a_l|} A_l Y, \\ \partial^{(2)} X &= \Delta^{-1} Y \Delta X - XY, \\ \partial^{(2)} Y &= Y^2, \\ \partial^{(2)} \Delta &= 0, \end{aligned}$$

where $\Phi\colon \mathcal{A} \to \mathrm{Mat}(2, \mathcal{A}(\Lambda^{(2)}))$ is a ring homomorphism given by $\Phi(a_l) = A_l$ and $\Phi(t) = \Delta X$.

Given two augmentations ϵ^1 and ϵ^2 of $(\mathcal{A}, \partial)$, we get an augmentation ϵ of $(\Lambda^{(2)}, \partial^{(2)})$ by sending $a_l^{ii} \mapsto \epsilon^i(a_l)$ and $t^{ii} \mapsto \epsilon^i(t)$ and sending everything else to 0. Therefore

$\partial_\epsilon^{(2)} = \phi_\epsilon \circ \partial^{(2)} \circ \phi_\epsilon^{-1}$ is a differential of $\mathcal{A}^{(2)} = \mathcal{A}(\Lambda^{(2)})/(t^{ii} = \epsilon(t^{ii}))$. Both the morphism $\mathrm{Hom}_+(\epsilon_1, \epsilon_2)$ and the first A_∞ operation m_1 are defined from $(\mathcal{A}^{(2)}, \partial_\epsilon^{(2)})$ through the construction stated in Section 3.1. For $i, j = 1, 2$, let C^{ij} denote the free graded $\mathbb{F}$ algebra generated by $\mathcal{R}^{ij}$, which is a subalgebra of $\mathcal{A}^{(2)}$. Notice that C^{12} and C^{21} are closed under $\partial_\epsilon^{(2)}$ since ϵ vanishes on the components in C^{11} and C^{22} of the image of $\partial_\epsilon^{(2)}$. Hence C^{12} and C^{21} are subcomplexes of $(\mathcal{A}^{(2)}, \partial_\epsilon^{(2)})$. Define the morphism $\mathrm{Hom}_+(\epsilon_1, \epsilon_2)$ between objects ϵ_1 and ϵ_2 to be $(C^{12})^\vee$. To simplify the notation, we write $(a_l^{12})^\vee$ as $a_l^\vee$, $(x^{12})^\vee$ as $x^\vee$ and $(y^{12})^\vee$ as $y^\vee$. Therefore, their gradings satisfy $|a_l^\vee| = |a_l| + 1$, $|x^\vee| = 1$ and $|y^\vee| = 0$. The first A_∞ operation m_1 is defined by the adjoint of $\partial_\epsilon^{(2)}$, ie for any Reeb chord $c \in \mathcal{R}$,

$$m_1(c^\vee) = \sum_{a \in \mathcal{R}} \mathrm{Coeff}_c(\partial_\epsilon^{(2)} a)\, a^\vee.$$

As noted before, m_1 is a differential for $\mathrm{Hom}_+(\epsilon_1, \epsilon_2)$. The corresponding cohomology is denoted by $H^* \mathrm{Hom}_+(\epsilon_1, \epsilon_2)$. Similarly, define $\mathrm{Hom}_-(\epsilon_2, \epsilon_1)$ to be $(C^{21})^\vee$. Take the cohomology of $\mathrm{Hom}_-(\epsilon_2, \epsilon_1)$ with respect to m_1, denoted by $H^* \mathrm{Hom}_-(\epsilon_2, \epsilon_1)$.

Remark One may find the notational convention of $\mathrm{Hom}_-(\epsilon_2, \epsilon_1)$ unnatural. However, the notations are consistent in the sense that both $\mathrm{Hom}_+(\epsilon, \epsilon')$ and $\mathrm{Hom}_-(\epsilon, \epsilon')$ are generated by Reeb chords from the component with the augmentation ϵ' to the component with the augmentation ϵ.

The $\mathrm{Hom}_+(\epsilon_1, \epsilon_2)$ space and the $\mathrm{Hom}_-(\epsilon_1, \epsilon_2)$ space are closely related. Recall that the generators of $\mathrm{Hom}_+(\epsilon_1, \epsilon_2)$ naturally correspond to the Reeb chords in $\mathcal{R}^{12}$, which consist of non-Morse Reeb chords and Morse Reeb chords. Note that the lengths of Morse Reeb chords are smaller than the lengths of non-Morse Reeb chords. Due to the positive energy constraint, there does not exist any holomorphic disk that has a positive puncture at a Morse Reeb chord and a negative puncture at a non-Morse Reeb chord. Therefore, the graded vector subspace of $\mathrm{Hom}_+(\epsilon_1, \epsilon_2)$ generated by non-Morse Reeb chords is closed under m_1, and thus is a subcomplex. Indeed, this subcomplex agrees with $(\mathrm{Hom}_-(\epsilon_1, \epsilon_2), m_1)$. From [27], for a Legendrian knot Λ with a single basepoint, any two augmentations ϵ_1 and ϵ_2 agree on the generator t that corresponds to the basepoint. As a result, by [33, Proposition 5.2], the quotient chain complex that is generated by $\{x^\vee, y^\vee\}$ is the Morse cochain complex induced by the Morse function f. Therefore we have the following long exact sequence:

$$\text{(2)} \quad \cdots \to H^{i-1}(\Lambda) \to H^i \mathrm{Hom}_-(\epsilon_1, \epsilon_2) \to H^i \mathrm{Hom}_+(\epsilon_1, \epsilon_2) \to H^i(\Lambda) \to \cdots.$$

Furthermore, given that both $\mathrm{Hom}_+(\epsilon_1, \epsilon_2)$ and $\mathrm{Hom}_-(\epsilon_1, \epsilon_2)$ are vector spaces over the field $\mathbb{F}$, combining the universal coefficient theorem with the Sabloff duality in

[33, Section 5.1.2], we have

$$H^k \operatorname{Hom}_-(\epsilon_1,\epsilon_2) \cong H^{-k}(\operatorname{Hom}_-(\epsilon_1,\epsilon_2)^\dagger) \cong H^{2-k} \operatorname{Hom}_+(\epsilon_2,\epsilon_1). \tag{3}$$

For a chain complex C, the chain complex $C^\dagger$ is obtained by dualizing the underlying vector space and differential of C and then negating the gradings.

For the other A_∞ operators m_n, one needs to consider an n–copy of Λ, denoted by $\Lambda^{(n)}$. Construct a DGA $(\mathcal{A}^n, \partial_\epsilon^{(n)})$ of $\Lambda^{(n)}$ that is analogous to $(\mathcal{A}^2, \partial_\epsilon^{(2)})$. Define m_n to be the adjoint of ∂_ϵ^n as in Example 3.2. See [33] for more details.

By [33], the augmentation category described above does not depend on the choice of the Morse function f. Moreover, up to A_∞ category equivalence, the augmentation category is invariant of Legendrian knot under Legendrian isotopy.

A key property of $\mathcal{A}\mathrm{ug}_+(\Lambda)$ is that $\mathcal{A}\mathrm{ug}_+(\Lambda)$ is a *strictly unital* A_∞ category, with the *units* given by

$$e_\epsilon = -y^\vee \in \operatorname{Hom}_+(\epsilon,\epsilon),$$

ie

- $m_1(e_\epsilon) = 0$;
- for any ϵ_1, ϵ_2 and any $c \in \operatorname{Hom}_+(\epsilon_1,\epsilon_2)$, we have $m_2(c, e_{\epsilon_1}) = m_2(e_{\epsilon_2}, c) = c$;
- any higher composition involving e_ϵ is 0.

As a result, the corresponding cohomology category $H^*\mathcal{A}\mathrm{ug}_+(\Lambda)$ is a unital category, which makes it natural to talk about the equivalence relation of objects in $\mathcal{A}\mathrm{ug}_+(\Lambda)$.

Definition 3.4 Two objects ϵ_1 and ϵ_2 are *equivalent* in $\mathcal{A}\mathrm{ug}_+(\Lambda)$ if they are isomorphic in the cohomology category $H^*\mathcal{A}\mathrm{ug}_+(\Lambda)$, ie if there exist $[\alpha] \in H^0 \operatorname{Hom}_+(\epsilon_1,\epsilon_2)$ and $[\beta] \in H^0 \operatorname{Hom}_+(\epsilon_2,\epsilon_1)$ such that $m_2([\alpha],[\beta]) = [e_{\epsilon_2}] \in H^0 \operatorname{Hom}_+(\epsilon_2,\epsilon_2)$ and $m_2([\beta],[\alpha]) = [e_{\epsilon_1}] \in H^0 \operatorname{Hom}_+(\epsilon_1,\epsilon_1)$:

$$\begin{tikzcd} \epsilon_1 \arrow[loop left, "e_{\epsilon_1}"] \arrow[r, bend right, "\beta"'] & \epsilon_2 \arrow[l, bend right, "\alpha"'] \arrow[loop right, "e_{\epsilon_2}"] \end{tikzcd}$$

By [33], for a Legendrian knot with a single basepoint, two augmentations are equivalent if and only if they are isomorphic as DGA maps.

Suppose Σ is a connected exact Lagrangian cobordism from a Legendrian knot Λ_- to a Legendrian knot Λ_+. It induces a DGA map ϕ from the DGA $(\mathcal{A}(\Lambda_+),\partial)$ with a single basepoint to a DGA $(\mathcal{A}(\Lambda_-),\partial)$ with a single basepoint. By [33, Proposition 3.29], this DGA map ϕ induces a unital A_∞ category morphism f from $\mathcal{A}\mathrm{ug}_+(\Lambda_-)$ to $\mathcal{A}\mathrm{ug}_+(\Lambda_+)$. The category map sends an augmentation ϵ_- of Λ_-

to $\epsilon_+ = \epsilon_- \circ \phi$, which is an augmentation of Λ_+. The family of maps $\{f_n\}$ is constructed through a family of DGA morphisms of n–copies:

$$\begin{aligned} f^{(n)}\colon (\mathcal{A}^{(n)}(\Lambda_+), \partial^{(n)}) \mapsto (\mathcal{A}^{(n)}(\Lambda_-), \partial^{(n)}),& \\ \Delta \mapsto \Delta,& \\ Y \mapsto Y,& \\ X \mapsto \Delta^{-1} \cdot \Phi_- \circ f(t),& \\ \Phi_+(a) \mapsto \Phi_- \circ f(a) \quad \text{for } a \in \mathcal{A}(\Lambda_+).& \end{aligned}$$

Let ϵ_- be the augmentation of $(\mathcal{A}^{(n)}(\Lambda_+), \partial^{(n)})$ that sends $a_l^{ii} \mapsto \epsilon_-^i(a_l)$, $t^{ii} \mapsto \epsilon_-^i(t)$ and sends everything else to 0. Define the map f_{n-1} to be the adjoint of $f_{\epsilon_-}^{(n)}$, where

$$f_{\epsilon_-}^{(n)} = \phi_{\epsilon_-} \circ f^{(n)} \circ \phi_{\epsilon_-}^{-1}.$$

In particular, f_1 can be written as

$$\begin{aligned} f_1\colon \mathrm{Hom}_+(\epsilon_-^1, \epsilon_-^2) &\to \mathrm{Hom}_+(\epsilon_+^1, \epsilon_+^2), \\ y_-^\vee &\mapsto y_+^\vee, \\ c^\vee &\mapsto \sum_{a \in \mathcal{A}(\Lambda_+)} \mathrm{Coeff}_c(f_{\epsilon_-}^{(2)}(a)) a^\vee \quad \text{for } c \in \mathcal{A}(\Lambda_-), \\ x_-^\vee &\mapsto x_+^\vee + \sum_{a \in \mathcal{A}(\Lambda_+)} \mathrm{Coeff}_t(f_{\epsilon_-}^{(2)}(a)) a^\vee. \end{aligned} \tag{4}$$

When computing $\mathrm{Coeff}_b(f_{\epsilon_-}^{(2)}(a))$, where b is either a Reeb chord $c \in \mathcal{A}(\Lambda_-)$ or $t \in \mathcal{T}$, one considers all the terms of $f(a)$ including b. If a term of $f(a)$ including b can be written as $\boldsymbol{p} b \boldsymbol{q}$, where $\boldsymbol{p}$ and $\boldsymbol{q}$ are words of pure Reeb chords of Λ_-, this term contributes $\mathrm{Coeff}_{\boldsymbol{p}b\boldsymbol{q}}(f(a))\epsilon_-^1(\boldsymbol{p})\epsilon_-^2(\boldsymbol{q})$ to $\mathrm{Coeff}_b(f_{\epsilon_-}^{(2)}(a))$. Therefore we have

$$\mathrm{Coeff}_b(f_{\epsilon_-}^{(2)}(a)) = \sum_{\boldsymbol{p}\,\boldsymbol{q}} \mathrm{Coeff}_{\boldsymbol{p}b\boldsymbol{q}}(f(a))\epsilon_-^1(\boldsymbol{p})\epsilon_-^2(\boldsymbol{q}).$$

Remark According to [33, Proposition 3.29], the condition for a DGA map to induce a unital A_∞ category morphism is that the DGA map is compatible with the weak link gradings in the sense of [33, Definition 3.19]. In our case, where both Λ_+ and Λ_- are single component Legendrian knots with a single basepoint, this condition is trivially satisfied.

4 Floer theory for Lagrangian cobordisms

In this section, we give a brief introduction to the Floer theory of a pair of exact Lagrangian cobordisms, following [6]. Let Σ^i, for $i = 1, 2$, be exact Lagrangian cobordisms from Λ_-^i to Λ_+^i in $(\mathbb{R}\times\mathbb{R}^3, d(e^t\alpha))$, where $\alpha = dz - y\,dx$. A schematic picture is shown in Figure 8. The union of the cobordisms $\Sigma^1\cup\Sigma^2$ is cylindrical over $\Lambda_+^1\cup\Lambda_+^2$ (resp. $\Lambda_-^1\cup\Lambda_-^2$) on the positive end (resp. negative end). Viewing $\Sigma^1\cup\Sigma^2$ as a Lagrangian cobordism from the Legendrian link $\Lambda_-^1\cup\Lambda_-^2$ to the Legendrian $\Lambda_+^1\cup\Lambda_+^2$, we obtain a chain complex generated by Reeb chords of $\Lambda_-^1\cup\Lambda_-^2$ and $\Lambda_+^1\cup\Lambda_+^2$. On the other hand, if we lift the exact Lagrangian cobordism $\Sigma^1\cup\Sigma^2$ to a Legendrian manifold in $\mathbb{R}\times\mathbb{R}^3\times\mathbb{R}$, we have its Legendrian contact homology DGA, which is generated by double points of $\Sigma^1\cup\Sigma^2$. One can construct the Cthulhu chain complex $\mathrm{Cth}(\Sigma^1,\Sigma^2)$ as a mix of the two chain complexes above. It is generated by some Reeb chords on the cylindrical ends and intersection points of Σ^1 and Σ^2. Moreover, this chain complex has trivial cohomology, ie $H^*\,\mathrm{Cth}(\Sigma^1,\Sigma^2)=0$.

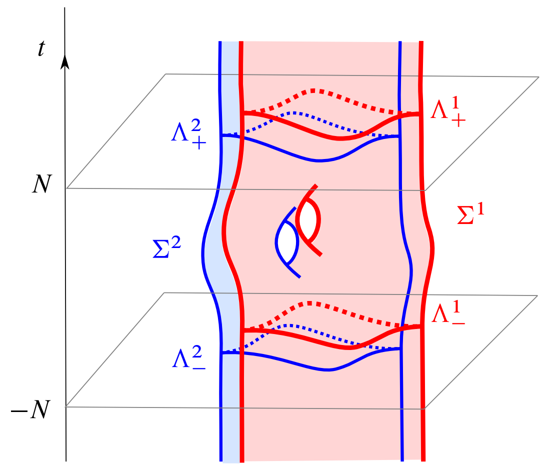

Figure 8: Pair of Lagrangian cobordisms in $(\mathbb{R}\times\mathbb{R}^3, d(e^t\alpha))$

For simplicity in defining gradings, we assume that Σ^i, for $i=1,2$, has trivial Maslov number throughout this paper.

4.1 The graded vector space

Assume that both Σ^1 and Σ^2 are cylindrical outside $[-N,N]\times\mathbb{R}^3$, where N is a positive number. The underlying vector space is a direct sum of three parts:

$$\mathrm{Cth}(\Sigma^1,\Sigma^2) = C(\Lambda_+^1,\Lambda_+^2)\oplus \mathrm{CF}(\Sigma^1,\Sigma^2)\oplus C(\Lambda_-^1,\Lambda_-^2).$$

The top level $C(\Lambda_+^1,\Lambda_+^2)$ (resp. bottom level $C(\Lambda_-^1,\Lambda_-^2)$) is an $\mathbb{F}$–module generated by Reeb chords to Λ_+^1 (resp. Λ_-^1) from Λ_+^2 (resp. Λ_-^2) that are lying on the slice of $t=N$ (resp. $t=-N$). The middle level $\mathrm{CF}(\Sigma^1,\Sigma^2)$ is an $\mathbb{F}$–module generated by intersection points of Σ^1 and Σ^2, which are all contained in $(-N,N)\times\mathbb{R}^3$.

Grading To define the degree, first fix a capping path γ_c for each generator c. For a Reeb chord a in $C(\Lambda_+^1, \Lambda_+^2)$ or $C(\Lambda_-^1, \Lambda_-^2)$, define the degree $|a|$ by

$$|a| = \mathrm{CZ}(\gamma_a) - 1,$$

which matches the definition of degree when viewing a as a generator in the Legendrian contact homology DGA of $\Lambda_+^1 \cup \Lambda_+^2$ or $\Lambda_-^1 \cup \Lambda_-^2$. For an intersection point $x \in \mathrm{CF}(\Sigma^1, \Sigma^2)$, define the degree $|x|$ by

$$|x| = \mathrm{CZ}(\gamma_x),$$

following [35]. One can also see [6, Section 4.2] for details. Note that for a Reeb chord in $C(\Lambda_+^1, \Lambda_+^2)$, its degree in $\mathrm{Cth}(\Sigma^1, \Sigma^2)$ will not necessarily coincide with its degree in $C(\Lambda_+^1, \Lambda_+^2)$. It is shifted as we will see later.

Action For $i = 1, 2$, suppose g_i is a primitive of the exact Lagrangian cobordism Σ^i, and hence g_i is constant when $t < -N$ or $t > N$. Note that primitive functions are well defined up to a overall shift by a constant. Thus we may assume that the primitives g_i are both zero on $\Sigma^i \cup ((-\infty, -N) \times \mathbb{R}^3)$ for $i = 1, 2$. The action of generators is defined under this choice of primitives.

For Reeb chords $a^+ \in C(\Lambda_+^1, \Lambda_+^2)$ and $a^- \in C(\Lambda_-^1, \Lambda_-^2)$, define the *action* $\mathfrak{a}$ by

$$\mathfrak{a}(a^+) = g_2(a^+) - g_1(a^+) + \int_{a^+} e^N \alpha$$

and

$$\mathfrak{a}(a^-) = g_2(a^-) - g_1(a^-) + \int_{a^-} e^{-N} \alpha = \int_{a^-} e^{-N} \alpha.$$

The last part is due to the special choice of primitives. For double points x of $\Sigma^1 \cup \Sigma^2$, the action $\mathfrak{a}(x)$ is defined by $\mathfrak{a}(x) = g_2(x) - g_1(x)$.

4.2 The differential

Remark 4.1 Throughout this paper, we restrict ourselves to the case where all intersection generators have positive actions since that is the case for the special pair of cobordisms constructed in Section 5.1. In general, the differential could include one more map from $\mathrm{CF}(\Sigma^1, \Sigma^2)$ to $C(\Lambda_-^1, \Lambda_-^2)$, which is called the Nessie map. However, by [6, Proposition 9.1], the positive energy condition of the holomorphic disks counted by the Nessie map requires the corresponding intersections in $\mathrm{CF}(\Sigma^1, \Sigma^2)$ to have negative actions. Therefore, in our special case, we can exclude the Nessie map and get the differential as an upper triangle as below.

With the assumption in Remark 4.1, we define the differential under the decomposition

$$\mathrm{Cth}(\Sigma^1,\Sigma^2)=C(\Lambda^1_+,\Lambda^2_+)\oplus \mathrm{CF}(\Sigma^1,\Sigma^2)\oplus C(\Lambda^1_-,\Lambda^2_-)$$

by a degree-1 map of the form

$$d=\begin{pmatrix} d_{++} & d_{+0} & d_{+-}\\ 0 & d_{00} & d_{0-}\\ 0 & 0 & d_{--}\end{pmatrix}.$$

To describe the differential explicitly, we need to study the holomorphic disks with boundary on $\Sigma^1\cup\Sigma^2$. Fix a generic domain-dependent almost complex structure J that is compatible with the symplectic form on $\mathbb{R}\times\mathbb{R}^3$ and the cylindrical ends in the sense of [6, Section 3.1.5]. Suppose that the induced cylindrical almost complex structure on the positive end $(\Sigma^1\cup\Sigma^2)\cap([N,\infty)\times\mathbb{R}^3)$ is J_+ and on the negative end $(\Sigma^1\cup\Sigma^2)\cap((-\infty,-N]\times\mathbb{R}^3)$ is J_-. The differential $d_{\pm\pm}$ of $C(\Lambda^1_\pm,\Lambda^2_\pm)$ counts rigid $J_\pm$–holomorphic disks with boundary on $\mathbb{R}\times(\Lambda^1_\pm\cup\Lambda^2_\pm)$, respectively, as described in Section 2.1. The corresponding moduli space is denoted by $\mathcal{M}_{J_\pm}(a^\pm;\boldsymbol{p}^\pm,b^\pm,\boldsymbol{q}^\pm)$, where $a^\pm$ and $b^\pm$ are Reeb chords to $\Lambda^1_\pm$ from $\Lambda^2_\pm$ while $\boldsymbol{p}^\pm$ and $\boldsymbol{q}^\pm$ are words of pure Reeb chords of $\Lambda^1_\pm$ and $\Lambda^2_\pm$, respectively. We also write $\widetilde{\mathcal{M}}_{J_\pm}(a^\pm;\boldsymbol{p}^\pm,b^\pm,\boldsymbol{q}^\pm)$ to denote the moduli space $\mathcal{M}_{J_\pm}(a^\pm;\boldsymbol{p}^\pm,b^\pm,\boldsymbol{q}^\pm)$ modulo the action of $\mathbb{R}$ in the t direction.

For the remaining maps in the differential, we need to describe J–holomorphic disks with boundary on $\Sigma^1\cup\Sigma^2$, where J is the chosen domain-dependent almost complex structure. The punctures of these J–holomorphic disks can be either Reeb chords or intersection points. For generators $c_0,c_1,\dots,c_m$ in $\mathrm{Cth}(\Sigma^1,\Sigma^2)$, let $\mathcal{M}_J(c_0;c_1,\dots,c_m)$ denote the moduli space of the J–holomorphic disks

$$u\colon (D_{m+1},\partial D_{m+1})\to(\mathbb{R}\times\mathbb{R}^3,\Sigma^1\cup\Sigma^2)$$

with the following properties:

- D_{m+1} is a 2–dimensional unit disk with $m+1$ boundary points $q_0,q_1,\dots,q_m$ removed and the points $q_0,q_1,\dots,q_m$ are arranged in a counterclockwise order.
- If c_0 is a Reeb chord, the image of u is asymptotic to $[N,\infty)\times c_0$ near q_0. If c_0 is an intersection point, then $\lim_{z\to q_0}u(z)=c_0$ and u maps the incoming segment (resp. outgoing segment) of the boundary to Σ^2 (resp. Σ^1).
- For $i>0$, if c_i is a Reeb chord, the image of u is asymptotic to $(-\infty,-N]\times c_i$ near q_i. If c_i is an intersection point, then $\lim_{z\to q_i}u(z)=c_i$ and u maps the incoming segment (resp. outgoing segment) of the boundary to Σ^1 (resp. Σ^2).

The four types of moduli spaces used in the differential d of the Cthulhu chain complex are shown in Figure 9. For any one of these four moduli spaces $\mathcal{M}$, we say a disk $u \in \mathcal{M}$ is *rigid* if $\dim \mathcal{M} = 0$. All the moduli spaces of holomorphic disks introduced above admit a coherent orientation since both Σ^1 and Σ^2 are spin. Therefore, one can associate each rigid holomorphic disk $u \in \mathcal{M}$ with a sign and thus can count the number of rigid holomorphic disks in $\mathcal{M}$ with sign.

Let each $a_i^{\pm}$ be a Reeb chord to $\Lambda_{\pm}^1$ from $\Lambda_{\pm}^2$ and each x_i a double point of $\Sigma^1 \cup \Sigma^2$. The bold letters $\boldsymbol{p}^{\pm}$, $\boldsymbol{q}^{\pm}$ are words of pure Reeb chords of $\Lambda_{\pm}^1$ and $\Lambda_{\pm}^2$, respectively. Here we assume, for $i = 1, 2$, that ϵ_-^i is an augmentation of $\mathcal{A}(\Lambda_-^i)$ and ϵ_+^i is the augmentation of $\mathcal{A}(\Lambda_+^i)$ induced by Σ_i. The differential is defined as follows:

$$d_{++}(a_i^+) = \sum_{\dim \widetilde{\mathcal{M}}_{J_+}(a_j^+;\boldsymbol{p}^+,a_i^+,\boldsymbol{q}^+)=0} |\widetilde{\mathcal{M}}_{J_+}(a_j^+;\boldsymbol{p}^+,a_i^+,\boldsymbol{q}^+)|\epsilon_+^1(\boldsymbol{p}^+)\epsilon_+^2(\boldsymbol{q}^+)a_j^+,$$

$$d_{--}(a_i^-) = \sum_{\dim \widetilde{\mathcal{M}}_{J_-}(a_j^-;\boldsymbol{p}^-,a_i^-,\boldsymbol{q}^-)=0} |\widetilde{\mathcal{M}}_{J_-}(a_j^-;\boldsymbol{p}^-,a_i^-,\boldsymbol{q}^-)|\epsilon_-^1(\boldsymbol{p}^-)\epsilon_-^2(\boldsymbol{q}^-)a_j^-,$$

$$d_{00}(x_i) = \sum_{\dim \mathcal{M}_J(x_j;\boldsymbol{p}^-,x_i,\boldsymbol{q}^-)=0} |\mathcal{M}_J(x_j;\boldsymbol{p}^-,x_i,\boldsymbol{q}^-)|\epsilon_-^1(\boldsymbol{p}^-)\epsilon_-^2(\boldsymbol{q}^-)x_j,$$

$$d_{0-}(a_i^-) = \sum_{\dim \mathcal{M}_J(x_j;\boldsymbol{p}^-,a_i^-,\boldsymbol{q}^-)=0} |\mathcal{M}_J(x_j;\boldsymbol{p}^-,a_i^-,\boldsymbol{q}^-)|\epsilon_-^1(\boldsymbol{p}^-)\epsilon_-^2(\boldsymbol{q}^-)x_j,$$

$$d_{+0}(x_i) = \sum_{\dim \mathcal{M}_J(a_j^+;\boldsymbol{p}^-,x_i,\boldsymbol{q}^-)=0} |\mathcal{M}_J(a_j^+;\boldsymbol{p}^-,x_i,\boldsymbol{q}^-)|\epsilon_-^1(\boldsymbol{p}^-)\epsilon_-^2(\boldsymbol{q}^-)a_j^+,$$

$$d_{+-}(a_i^-) = \sum_{\dim \mathcal{M}_J(a_j^+;\boldsymbol{p}^-,a_i^-,\boldsymbol{q}^-)=0} |\mathcal{M}_J(a_j^+;\boldsymbol{p}^-,a_i^-,\boldsymbol{q}^-)|\epsilon_-^1(\boldsymbol{p}^-)\epsilon_-^2(\boldsymbol{q}^-)a_j^+,$$

where $|\mathcal{M}|$ denotes the number of rigid holomorphic disks in the moduli space $\mathcal{M}$ counted with sign. Note that the definition of differential depends on the choice of augmentations ϵ_1^- and ϵ_2^-, whose existence are essential to the Floer theory.

A holomorphic disk counted by the differential must satisfy the rigidity condition and the positive energy condition. We will describe these conditions in detail.

The rigidity condition Let us interpret the condition $\dim \mathcal{M}_J(c_1;\boldsymbol{p},c_2,\boldsymbol{q}) = 0$ in terms of $|c_i|$ for $i = 1, 2$, where c_i can be either a Reeb chord or an intersection point while $\boldsymbol{p}$ and $\boldsymbol{q}$ are words of pure Reeb chords in degree 0. Instead of deriving a formula for the dimension of a moduli space, we use the idea of the wrapped Floer homology to find the relation between $|c_1|$ and $|c_2|$. Recall that both Σ^1 and Σ^2 are cylindrical outside $[-N, N] \times \mathbb{R}^3$. Consider a nondecreasing function $\sigma(t)\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such

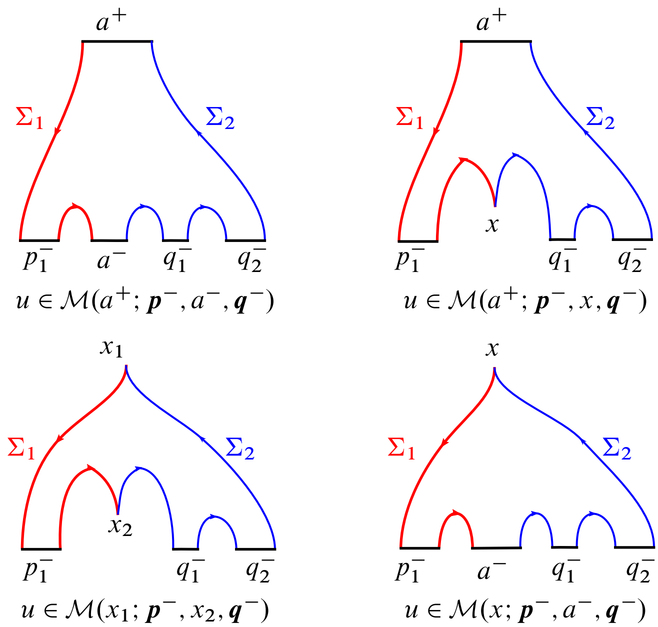

Figure 9: A sketch of the J–holomorphic disks in the differential d. Here $a^\pm$ are Reeb chords to $\Lambda^1_\pm$ from $\Lambda^2_\pm$, respectively, and x, x_1, x_2 are double points of $\Sigma^1 \cup \Sigma^2$. In these examples, $\boldsymbol{p}^-$ is a word of one pure Reeb chord p_1^- of Λ^1_- while $\boldsymbol{q}^-$ is a word $q_1^- q_2^-$ of two pure Reeb chords of Λ^2_-. The arrows denote the orientation inherited from the boundary of the unit disk.

that $\sigma'(t) = 0$ when $t \le N$ and $\sigma'(t) = 1$ when $t \ge N'$, where N' is a number bigger than N. Note that $X_H = -\sigma(|t|)\partial_z$ is a Hamiltonian vector field with its time-s flow denoted by Φ^s_H. Flow Σ^1 through X_H and get a new cobordism $\Phi^s_H(\Sigma^1)$, which is another exact Lagrangian cobordism according to Section 5.1. Observe that $\Phi^s_H(\Sigma^1)$ wraps Σ^1 on both ends in the negative Reed chord direction. Hence for a large enough number s, each Reeb chord c to Λ^1_+ (resp. Λ^1_-) from Λ^2_+ (resp. Λ^2_-) corresponds to a transversally double point $\check{c}$ of $\Phi^s_H(\Sigma^1) \cup \Sigma^2$ in $N < t < N'$ (resp. $-N' < t < -N$). Moreover, if c is a Reeb chord in $C(\Lambda^1_+, \Lambda^2_+)$, we have

$$|\check{c}| = \mathrm{CZ}(\gamma_{\check{c}}) = \mathrm{CZ}(\gamma_c) + 1 = |c| + 2.$$

If c is Reeb chord in $C(\Lambda^1_-, \Lambda^2_-)$,

$$|\check{c}| = \mathrm{CZ}(\gamma_{\check{c}}) = \mathrm{CZ}(\gamma_c) = |c| + 1.$$

Each double point x of $\Sigma^1 \cup \Sigma^2$ naturally corresponds to a double point $\check{x}$ of $\Phi^s_H(\Sigma^1) \cup \Sigma^2$ in $-N < t < N$ with gradings satisfying $|\check{x}| = |x|$.

Remark The difference in grading correspondence between the Reeb chords in $C(\Lambda^1_+,\Lambda^2_+)$ and the Reeb chords in $C(\Lambda^1_-,\Lambda^2_-)$ can be understood better in a special case where Σ^1 is a pushoff of Σ^2 through a positive Morse function $F\colon \Sigma^2\to\mathbb{R}_{>0}$. In other words, in a Weinstein neighborhood of Σ^2, the cobordism Σ^1 is the graph of dF for some positive Morse function $F\colon \Sigma^2\to\mathbb{R}_{>0}$. In this case, the cobordism $\Phi^s_h(\Sigma^1)$ is a pushoff of Σ^2 through another Morse function $\widetilde{F}$ as well. By the canonical Floer theory [24], we can choose a family of capping paths so that $\mathrm{CZ}(\gamma_x)=\mathrm{Ind}_{\widetilde{F}}(x)$ for any intersection point x of $\Phi^s_h(\Sigma^1)$ and Σ^2. Similarly, for any Morse Reeb chord c in $\Lambda^1_+\cup\Lambda^2_+$, we can further require that $\mathrm{Ind}_{f_+}(c)=\mathrm{CZ}(\gamma_c)$, where $f_+=F|_{\Lambda^2_+}$. Notice that $\mathrm{Ind}_{\widetilde{F}}(\check{c})=\mathrm{Ind}_{f_+}(c)+1$. Therefore

$$|\check{c}|=\mathrm{CZ}(\gamma_{\check{c}})=\mathrm{Ind}_{\widetilde{F}}(\check{c})=\mathrm{Ind}_{f_+}(c)+1=\mathrm{CZ}(\gamma_c)+1=|c|+2.$$

For a Morse Reeb chord c in $\Lambda^1_-\cup\Lambda^2_-$, the indices satisfy $\mathrm{Ind}_{\widetilde{F}}(\check{c})=\mathrm{Ind}_{f_-}(c)$, where $f_-=F|_{\Lambda^2_-}$. Hence $|\check{c}|=\mathrm{CZ}(\gamma_{\check{c}})=\mathrm{Ind}_{\widetilde{F}}(\check{c})=\mathrm{Ind}_{f_-}(c)=\mathrm{CZ}(\gamma_c)=|c|+1$. A schematic figure is shown in Figure 10.

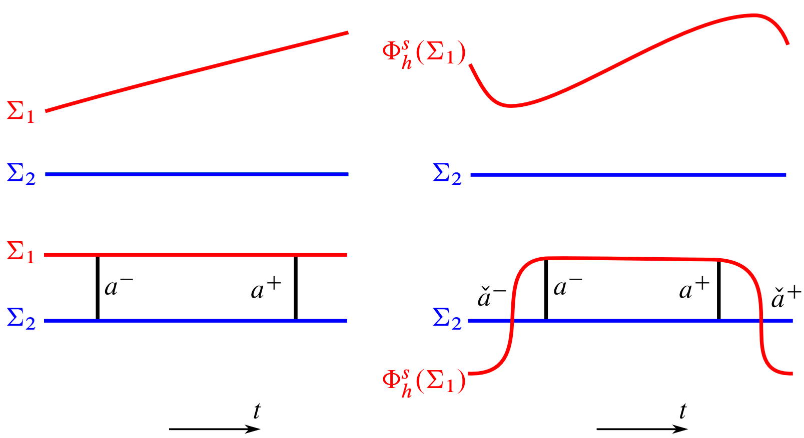

Figure 10: Comparison between Σ_1 and $\Phi^s_h(\Sigma_1)$ in terms of front projection (top) and Lagrangian projection (bottom). The indices satisfy $|\check{a}^+|=|a^+|+2$ and $|\check{a}^-|=|a^-|+1$.

So far, we have shown that generators of $\mathrm{Cth}(\Sigma^1,\Sigma^2)$ can be identified with intersection points of $\Phi^s_H(\Sigma^1)$ and Σ^2, which are generators of $\mathrm{Cth}(\Phi^s_H(\Sigma^1),\Sigma^2)$. Moreover, by [6, Proposition 8.2], the Cthulhu chain complexes $\mathrm{Cth}(\Sigma^1,\Sigma^2)$ and $\mathrm{Cth}(\Phi^s_H(\Sigma^1),\Sigma^2)$ are identified on the level of complexes as well. Note that the generators of $\mathrm{Cth}(\Phi^s_H(\Sigma^1),\Sigma^2)$ do not contain any Reeb chords and hence we have $\mathrm{Cth}(\Phi^s_H(\Sigma^1),\Sigma^2)=\big(\mathrm{CF}(\Phi^s_H(\Sigma^1),\Sigma^2),d_{00}\big)$. Lift $\Phi^s_H(\Sigma^1)\cup\Sigma^2$ to a Legendrian submanifold L in $\mathbb{R}\times\mathbb{R}^3\times\mathbb{R}$. Note that $\big(\mathrm{CF}(\Phi^s_H(\Sigma^1),\Sigma^2),d_{00}\big)$ is the dual of the linearized contact homology chain complex of L as introduced in Section 2.1. Hence if $\dim \mathcal{M}(c_1; \boldsymbol{p}, c_2, \boldsymbol{q}) = 0$, there are rigid holomorphic disks that have a positive puncture at $\check{c}_1$ and a negative puncture at $\check{c}_2$, which implies $|\check{c}_1| - |\check{c}_2| = 1$. We can get the corresponding grading relation between c_1 and c_2. In particular, let $a^{\pm}$ be Reeb chords in $C(\Lambda^1_{\pm}, \Lambda^2_{\pm})$ and x_1, x_2 be intersection points of Σ^1 and Σ^2 while $\boldsymbol{p}$ and $\boldsymbol{q}$ are words of pure Reeb chords in degree 0 of Λ^1_- and Λ^2_-, respectively. We have:

- If $\dim \mathcal{M}_J(x_1; \boldsymbol{p}, x_2, \boldsymbol{q}) = 0$, then $|c_1| - |c_2| = 1$.
- If $\dim \mathcal{M}_J(x_1; \boldsymbol{p}, a^-, \boldsymbol{q}) = 0$, then $|c_1| - |a^-| = 2$.
- If $\dim \mathcal{M}_J(a^+; \boldsymbol{p}, x_1, \boldsymbol{q}) = 0$, then $|a^+| - |x_1| = -1$.
- If $\dim \mathcal{M}_J(a^+; \boldsymbol{p}, a^-, \boldsymbol{q}) = 0$, then $|a^+| - |a^-| = 0$.

As a result, the Cthulhu chain complex can be written as

$$\mathrm{Cth}^k(\Sigma^1, \Sigma^2) = C^{k-2}(\Lambda^1_+, \Lambda^2_+) \oplus \mathrm{CF}^k(\Sigma^1, \Sigma^2) \oplus C^{k-1}(\Lambda^1_-, \Lambda^2_-).$$

Under this decomposition, the differential

$$d = \begin{pmatrix} d_{++} & d_{+0} & d_{+-} \\ 0 & d_{00} & d_{0-} \\ 0 & 0 & d_{--} \end{pmatrix}$$

has degree 1 as we expected.

The positive energy condition We interpret the positive energy condition of a holomorphic disk $u \in \mathcal{M}(c_0; c_1, \dots, c_m)$ in terms of the action of c_i, where $i = 0, \dots, m$. Following [14], we define the energy $E(u)$ of a holomorphic disk

$$u\colon (D^2, \partial D^2) \to (\mathbb{R} \times \mathbb{R}^3, \Sigma^1 \cup \Sigma^2)$$

by $E(u) = E_\omega(u) + E_\alpha(u)$, where the ω–energy is given by

$$E_\omega(u) = \int_{u^{-1}([-N,N]\times\mathbb{R}^3)} u^*(\omega) + \int_{u^{-1}((-\infty,-N)\times\mathbb{R}^3)} u^*(e^{-N} d\alpha) + \int_{u^{-1}((N,\infty)\times\mathbb{R}^3)} u^*(e^N d\alpha).$$

Write $u = (t, v)$, where $t\colon D^2 \to \mathbb{R}$ and $v\colon D^2 \to \mathbb{R}^3$. Define the α–energy $E_\alpha(u)$ by

$$\sup_{\phi_-}\left(\int_{u^{-1}((-\infty,-N)\times\mathbb{R}^3)} \phi_-(t)\, dt \wedge (v^*\alpha) \right) + \sup_{\phi_+}\left(\int_{u^{-1}((N,\infty)\times\mathbb{R}^3)} \phi_+(t)\, dt \wedge (v^*\alpha) \right),$$

where ϕ_+ and ϕ_- range over all compactly supported smooth functions such that

$$\int_{-\infty}^{-N} \phi_-(t)\, dt = e^{-N} \quad \text{and} \quad \int_N^{\infty} \phi_+(t)\, dt = e^N,$$

respectively. By Stokes' theorem, for any holomorphic disk $u \in \mathcal{M}(c_0; c_1, \dots, c_m)$, we have

$$E_\omega(u) = \mathfrak{a}(c_0) - \sum_{l=1}^{m} \mathfrak{a}(c_l).$$

A holomorphic disk has positive ω–energy, that is, $E_\omega(u) > 0$, which implies that $\mathfrak{a}(c_0) > \sum_{l=1}^{m} \mathfrak{a}(c_l)$.

Under the assumption in Remark 4.1, the differential d of the Cthulhu chain complex is of the form of an upper triangle. By [6, Section 6], we have $d^2 = 0$ and thus d is a differential map. Moreover, from [6, Section 8], the induced cohomology $H^* \operatorname{Cth}(\Sigma^1, \Sigma^2)$ is an invariant under compactly supported Hamiltonian isotopies. Push Σ^1 along the negative z direction until Σ^1 is far below Σ^2 and then there is no Reeb chord to Σ^1 from Σ^2 nor intersection point between Σ^1 and Σ^2. It is obvious that the cohomology is trivial, ie $H^* \operatorname{Cth}(\Sigma^1, \Sigma^2) = 0$.

5 Main result

In Section 5.1, we perturb an exact Lagrangian cobordism using a Morse function and obtain a pair of Lagrangian cobordisms. In Section 5.2, we apply the Floer theory to this pair of cobordisms and get the long exact sequence in Theorem 1.1. In Section 5.3, we describe the rigid holomorphic disks counted by d_{+-}, which is a part of the differential map of the Cthulhu chain complex, in terms of holomorphic disks with boundary on Σ and Morse flow lines. This is useful when identifying f_1, ie the category map on the level of morphisms, with d_{+-}. In Section 5.4, we extend the method in Section 5.3 to describe the differential d of the Cthulhu chain complex and recover the long exact sequences in [6]. Finally, we use the identification in Section 5.3 between f_1 and d_{+-} to prove Theorem 1.5 in Section 5.5.

5.1 Construction of the pair of cobordisms

First let us describe the neighborhood of a Lagrangian cobordism. Let Σ be an exact Lagrangian cobordism from Λ_- to Λ_+ in $(\mathbb{R} \times \mathbb{R}^3, d(e^t\alpha))$, where Λ_- and Λ_+ are Legendrian links. By the Weinstein Lagrangian neighborhood theorem, there is a symplectomorphism

$$\psi\colon \operatorname{nbhd}(\Sigma) \subset (\mathbb{R} \times \mathbb{R}^3, d(e^t\alpha)) \to (T^*\Sigma, d\theta),$$

where θ is the negative Liouville form $\theta = -\sum p_i\, dq_i$ of $T^*\Sigma$ with coordinates $((q_1, q_2), (p_1, p_2))$. Specifically, on the $(\pm\infty-)$boundary $\mathbb{R} \times \Lambda_\pm$, the symplectomorphism ψ is given by a composition $\psi_1 \circ \psi_0$ of two symplectomorphisms. As mentioned before, there is a contactomorphism from a tubular neighborhood of $\Lambda_\pm$ in $\mathbb{R}^3$ to a neighborhood of the zero section of $J^1(\Lambda_\pm)$. Composing with the identity map on $\mathbb{R}_t$, we get the symplectomorphism ψ_0 from the neighborhood of $\mathbb{R}\times\Lambda_\pm$ in $\mathbb{R}\times\mathbb{R}^3$ to $\mathbb{R}\times J^1(\Lambda_\pm)$. The second part ψ_1 is given by

$$\begin{aligned}\psi_1\colon \mathrm{nbhd}(\Sigma)\subset ((\mathbb{R}\times J^1(\Lambda_\pm)), d(e^t\alpha)) &\to (T^*(\mathbb{R}_{>0}\times\Lambda_\pm), d\theta),\\ (t,(q,p,z)) &\mapsto ((e^t,q),(z,e^tp)).\end{aligned}$$

For a Morse function $F\colon \Sigma\to\mathbb{R}_{\geq 0}$ such that the determinant of the Hessian matrix is small enough, the graph of dF is a Lagrangian submanifold in $T^*\Sigma$. Pull it back to $\mathbb{R}\times\mathbb{R}^3$ and denote $\psi^{-1}(\mathrm{graph}(dF))$ by Σ'.

Now we show that Σ' is an exact Lagrangian submanifold as well. Notice that $V_{(\boldsymbol{q},\boldsymbol{p})} := dF|_{\boldsymbol{q}}$ is a Hamiltonian vector field in $T^*\Sigma$ since $\iota_V d\theta = -d\widetilde{F}$, where $\widetilde{F} = F\circ\pi$ and π is the natural projection $\pi\colon T^*(\Sigma)\to\Sigma$. In order to extend $\psi_*^{-1}(V)$ to a Hamiltonian vector field in $\mathbb{R}\times\mathbb{R}^3$, we choose a smooth cutoff function $\gamma\colon T^*(\Sigma)\to\mathbb{R}$ such that $\gamma(\boldsymbol{q},\boldsymbol{p})=1$ in a tubular neighborhood of the zero section containing the graph of dF and $\gamma(\boldsymbol{q},\boldsymbol{p})=0$ outside a slightly bigger tubular neighborhood of the zero section. Pull the Hamiltonian vector field of $\gamma\cdot\widetilde{F}$ back through ψ and extend to a Hamiltonian vector field X_H in $\mathbb{R}\times\mathbb{R}^3$. For a suitable neighborhood of Σ in $\mathbb{R}\times\mathbb{R}^3$, we have

$$\iota_{X_H} d(e^t\alpha)|_{\mathrm{nbhd}(\Sigma)} = \psi^*(\iota_V d\theta) = \psi^*(-d\widetilde{F}) = d(-\widetilde{F}\circ\psi)|_{\mathrm{nbhd}(\Sigma)}.$$

Hence its Hamiltonian H is the same as $-\widetilde{F}\circ\psi$ around Σ. Denote the time-s flow of X_H by ϕ_H^s and thus $\Sigma' = \phi_H^1(\Sigma)$. We can compute the 1–form on Σ':

$$\begin{aligned}\phi_H^{1\,*} e^t\alpha &= e^t\alpha + \int_0^1 \frac{d}{ds}\phi_H^{s\,*}(e^t\alpha)\,ds \\ &= e^t\alpha + \int_0^1 \phi_H^{s\,*}(\iota_{X_H} d(e^t\alpha) + d(\iota_{X_H} e^t\alpha))\,ds \\ &= e^t\alpha + \int_0^1 \phi_H^{s\,*}(dH + d(e^t\alpha(X_H)))\,ds \\ &= e^t\alpha + d\biggl(\int_0^1 (H + e^t\alpha(X_H))\circ\phi_H^s\,ds\biggr).\end{aligned} \tag{5}$$

Thus Σ' is exact. Moreover, if Σ has a primitive g, then Σ' has a primitive

$$g + \int_0^1 (H + e^t\alpha(X_H))\circ\phi_H^s\,ds.$$

We are going to construct a particular Morse function for Σ such that the image of the Morse function has cylindrical ends as well and thus is an exact Lagrangian cobordism.

Suppose Σ is cylindrical outside $[-N+\delta, N-\delta]\times\mathbb{R}^3$, where $0<\delta<1$. Choose Morse functions $g_\pm\colon \Lambda_\pm \to \left(0,\frac{1}{2}\right)$ and $G\colon \Sigma\cap([-N,N]\times\mathbb{R}^3)\to(0,1)$ such that

$$G|_{\Sigma\cap\{t\in[-N,-N+\delta]\cup[N-\delta,N]\}} = e^t.$$

Define a smooth nondecreasing function $\rho\colon \mathbb{R}_{>0}\to[0,1]$ such that $\rho(s)=0$ for $s\le 1$ and $\rho(s)=1$ for $s\ge e^\delta$.

For $0<\eta<e^{-2}$, define a Morse function $F^\eta\colon \Sigma\to\mathbb{R}_{>0}$ to be

$$F^\eta(t,q) := \begin{cases} \eta^{2N}g_-(q)s & \text{if } t<-N,\\ \big(\rho(e^N s)(\eta^N-\eta^{2N}g_-(q))+\eta^{2N}g_-(q)\big)s & \text{if } -N\le t\le -N+\delta,\\ \eta^N G & \text{if } -N+\delta<t<N-\delta,\\ \big(\eta^N+\rho(e^{-N+\delta}s)\eta^N g_+(q)\big)s & \text{if } N-\delta\le t\le N,\\ \big(\eta^N+\eta^N g_+(q)\big)s & \text{if } t>N, \end{cases}$$

where $s=e^t$. One can check that F^η has the following properties:

- The Morse function F^η is increasing with respect to t when $t\le -N+\delta$ or $t\ge N-\delta$. This implies that the critical points of F^η and the critical points of G are in one-to-one correspondence and are all contained in $\Sigma\cap([-N,N]\times\mathbb{R}^3)$.
- The Morse function F^η is bounded by $2\eta^N e^N$ on $\Sigma\cap([-N,N]\times\mathbb{R}^3)$.
- Write $F^\eta|_{\{N\}\times\Lambda_+}$ as $f_+^\eta e^N$ and $F^\eta|_{\{-N\}\times\Lambda_-}$ as $f_-^\eta e^{-N}$, respectively. The graph of dF^η on $(-\infty,-N)\times\Lambda_-$ is the same as $(-\infty,-N)\times\operatorname{graph}(df_-^\eta)$ and the graph of dF^η on $(N,\infty)\times\Lambda_+$ is the same as $(N,\infty)\times\operatorname{graph}(df_+^\eta)$.

Push $(\Sigma,\Lambda_+,\Lambda_-)$ off through F^η and obtain a copy of $(\Sigma,\Lambda_+,\Lambda_-)$, labeled by $(\Sigma^1,\Lambda_+^1,\Lambda_-^1)$. Label the original $(\Sigma,\Lambda_+,\Lambda_-)$ by $(\Sigma^2,\Lambda_+^2,\Lambda_-^2)$. Thus Σ^1 is a pushoff of Σ^2 through F^η and Λ_+^1 (resp. Λ_-^1) is a pushoff of Λ_+^2 (resp. Λ_-^2) through f_+^η (resp. f_-^η).

5.2 The long exact sequence

Now we apply the Floer theory to the pair of cobordisms $\Sigma^1\cup\Sigma^2$ constructed in Section 5.1 and get a long exact sequence. Combining the long exact sequence with the augmentation category map induced by the exact Lagrangian cobordism, we obtain an obstruction to the existence of the exact Lagrangian cobordisms.

Recall that the grading for generators in the Cthulhu chain complex depends on the choice of capping paths. According to the canonical Floer theory [24], we can choose a family of capping paths such that the Conley–Zehnder index of any double point x of $\Sigma^1\cup\Sigma^2$ satisfies $\mathrm{CZ}(\Gamma_x)=\mathrm{Ind}_{F^\eta(x)}$. Now we apply the Floer theory to the pair of Lagrangian cobordisms $\Sigma^1\cup\Sigma^2$ and have the following theorem.

Theorem 5.1 *Let Σ^i, for $i=1,2$, be the cobordisms from Λ_-^i to Λ_+^i as constructed in Section 5.1. Suppose ϵ_-^i is an augmentation of $\mathcal{A}(\Lambda_-^i)$ and ϵ_+^i is the augmentation of $\mathcal{A}(\Lambda_+^i)$ induced by Σ^i. Fix a suitable domain-dependent almost complex structure on $\mathbb{R}\times\mathbb{R}^3$ that is compatible with the symplectic form and cylindrical ends. For η small enough, the Cthulhu chain complex is*

$$\mathrm{Cth}^k(\Sigma^1,\Sigma^2)=C^{k-2}(\Lambda_+^1,\Lambda_+^2)\oplus \mathrm{CF}^k(\Sigma^1,\Sigma^2)\oplus C^{k-1}(\Lambda_-^1,\Lambda_-^2).$$

Under this decomposition, the differential is

$$d=\begin{pmatrix} d_{++} & d_{+0} & d_{+-}\\ 0 & d_{00} & d_{0-}\\ 0 & 0 & d_{--}\end{pmatrix}.$$

Moreover:

(1) *The map d_{00} is the Morse codifferential induced by F^η, ie the chain complex $(\mathrm{CF}^k(\Sigma^1,\Sigma^2),d_{00})$ is the Morse cochain complex $(C^k_{\mathrm{Morse}}F^\eta,d_{F^\eta})$ induced by F^η.*

(2) *The chain complex $(C^{k-2}(\Lambda_+^1,\Lambda_+^2),d_{++})$ is equal to $(\mathrm{Hom}_+^{k-1}(\epsilon_+^1,\epsilon_+^2),m_1)$ while the chain complex $(C^{k-1}(\Lambda_-^1,\Lambda_-^2),d_{--})$ equals $(\mathrm{Hom}_+^{k}(\epsilon_-^1,\epsilon_-^2),m_1)$.*

Proof First, we need to show that each intersection point $x\in\mathrm{CF}^k(\Sigma^1,\Sigma^2)$ has a positive action, which is the condition for the differential to have the form above by Remark 4.1.

Let g_i be a primitive of Σ^i for $i=1,2$. According to the computation (5), we have

$$g_1=g_2+\int_0^1 (H+e^t\alpha(X_H))\circ\phi_H^s\,ds,$$

where $H=-\widetilde{F}^\eta\circ\psi$. It is not hard to check that $g_1=g_2$ on $\Sigma\cap((-\infty,-N)\times\mathbb{R}^3)$. Therefore we can assume $g_1=g_2=0$ on $\Sigma\cap((-\infty,-N)\times\mathbb{R}^3)$ and use g_1 and g_2 as primitives to define actions. The action of each intersection point x is given by

$$\mathfrak{a}(x)=g_2(x)-g_1(x)=\int_0^1 (H+e^t\alpha(X_H))\circ\phi_H^s\,ds.$$

Notice that the vector field X_H vanishes at the intersection point x. Hence

$$\mathfrak{a}(x)=-\int_0^1 H\circ\phi_H^s\,ds=-H=F^\eta\circ\psi(x)>0.$$

Next, we are going to show that for η small enough, the rigid holomorphic disks that contribute to d_{00} do not include any pure Reeb chords as negative punctures. Let x and y be two double points of $\Sigma^1\cup\Sigma^2$. For any rigid holomorphic disk

$u \in \mathcal{M}(x; \boldsymbol{p}, y, \boldsymbol{q})$, where $\boldsymbol{p}$ and $\boldsymbol{q}$ are words of pure Reeb chords of $\Lambda_-^1 \cup \Lambda_-^2$, we have the energy estimate

$$E_\omega(u) \le \mathfrak{a}(x) - \mathfrak{a}(y) - \mathfrak{a}(\boldsymbol{p}) - \mathfrak{a}(\boldsymbol{q}).$$

Since u has positive energy, we have

$$\mathfrak{a}(\boldsymbol{p}) + \mathfrak{a}(\boldsymbol{q}) \le F^\eta(\psi(x)) - F^\eta(\psi(y)) \le F^\eta(\psi(x)).$$

Therefore, for η small enough that the maximum of the Morse function F^η is smaller than the minimum action of pure Reeb chords of Λ_-^1 and Λ_-^2, the moduli space that contributes to d_{00} is of the form $\mathcal{M}(x; y)$. By [16, Lemma 6.11], the boundary of a rigid holomorphic disk with two punctures at intersection points converges to a rigid Morse flow line, which implies $d_{00} = d_{F^\eta}$. Furthermore, the gradings satisfy $|x| = \mathrm{CZ}(\gamma_x) = \mathrm{Ind}_{F^\eta(x)}$. Therefore $(\mathrm{CF}^k(\Sigma^1, \Sigma^2), d_{00}) = (C^k_{\mathrm{Morse}} F^\eta, d_{F^\eta})$.

Recall that there is a natural identity map with degree 1 from $C^{k-1}(\Lambda_\pm^1, \Lambda_\pm^2)$ to $\mathrm{Hom}_+^k(\epsilon_\pm^1, \epsilon_\pm^2)$, respectively. Moreover, the definitions of $d_{\pm\pm}$ and m_1 match as well. Hence we have $(C^{k-1}(\Lambda_-^1, \Lambda_-^2), d_{--}) = (\mathrm{Hom}_+^k(\epsilon_-^1, \epsilon_-^2), m_1)$. On the other hand, we have $(C^{k-2}(\Lambda_+^1, \Lambda_+^2), d_{++}) = (\mathrm{Hom}_+^{k-1}(\epsilon_+^1, \epsilon_+^2), m_1)$. □

For the remainder of the paper, we fix a small enough η and write F^η, f_+^η, f_-^η as F, f_+, f_-, respectively. According to the Floer theory in Section 4, we have $H^k(\mathrm{Cth}(\Sigma^1, \Sigma^2), d) = 0$, where

$$\mathrm{Cth}^k(\Sigma^1, \Sigma^2) = \mathrm{Hom}_+^{k-1}(\epsilon_+^1, \epsilon_+^2) \oplus C^k_{\mathrm{Morse}} F \oplus \mathrm{Hom}_+^k(\epsilon_-^1, \epsilon_-^2)$$

and

$$d = \begin{pmatrix} m_1 & d_{+0} & d_{+-} \\ 0 & d_F & d_{0-} \\ 0 & 0 & m_1 \end{pmatrix}.$$

Consider the chain map $\Psi = d_{+-} + d_{0-}$:

$$\Psi\colon (\mathrm{Hom}_+^k(\epsilon_-^1, \epsilon_-^2), m_1) \to (\mathrm{Hom}_+^k(\epsilon_+^1, \epsilon_+^2) \oplus C^{k+1}_{\mathrm{Morse}} F, d'), \quad \text{where } d' = \begin{pmatrix} m_1 & d_{+0} \\ 0 & d_F \end{pmatrix}.$$

Notice that the mapping cone of Ψ has trivial homology. Therefore,

$$H^k(\mathrm{Hom}_+(\epsilon_-^1, \epsilon_-^2)) \cong H^k \,\mathrm{Cone}(d_{+0}).$$

Hence we have the following long exact sequence:

$$\cdots \to H^k(C_{\mathrm{Morse}} F, d_F) \to H^k \,\mathrm{Hom}_+(\epsilon_+^1, \epsilon_+^2) \to H^k \,\mathrm{Hom}_+(\epsilon_-^1, \epsilon_-^2) \\ \to H^{k+1}(C_{\mathrm{Morse}} F, d_F) \to \cdots.$$

Moreover, notice that in the construction in Section 5.1, the gradient flows of F flow in from the bottom and out of the top. Hence we have

$$H^k(C_{\mathrm{Morse}}F, d_F) = H^k(\Sigma, \Lambda_-).$$

Corollary 5.2 *Let Σ be an exact Lagrangian cobordism with Maslov number 0 from Λ_- to Λ_+. For $i = 1, 2$, if ϵ_-^i is an augmentation of $\mathcal{A}(\Lambda_-)$ and ϵ_+^i is the augmentation of $\mathcal{A}(\Lambda_+)$ induced by Σ, then we have the following long exact sequence:*

$$\cdots \to H^k(\Sigma, \Lambda_-) \to H^k \operatorname{Hom}_+(\epsilon_+^1, \epsilon_+^2) \to H^k \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2) \to H^{k+1}(\Sigma, \Lambda_-) \to \cdots. \tag{6}$$

Remark When the Maslov number of Σ is d which is not 0 the method above works as well. The only difference is that the grading of generators in the Cthulhu chain complex is defined mod d. Thus, the long exact sequence (6) holds with gradings mod d.

If $\epsilon_-^1 = \epsilon_-^2 = \epsilon_-$, by [33, Section 5.2], we have the identification

$$H^k \operatorname{Hom}_+(\epsilon, \epsilon) \cong \mathrm{LCH}^{\epsilon}_{1-k}(\Lambda),$$

where $\mathrm{LCH}^{\epsilon}_k(\Lambda)$ is the linearized contact homology of Λ. The long exact sequence (6) can be rewritten in terms of linearized contact homology:

$$\cdots \to H^k(\Sigma, \Lambda_-) \to \mathrm{LCH}^{\epsilon_+}_{1-k}(\Lambda_+) \to \mathrm{LCH}^{\epsilon_-}_{1-k}(\Lambda_-) \to H^{k+1}(\Sigma, \Lambda_-) \to \cdots.$$

Furthermore, if Λ_- is empty, then Σ is an exact Lagrangian filling of Λ_+ and ϵ_+ is an augmentation of $\mathcal{A}(\Lambda_+)$ induced by the Lagrangian filling. The long exact sequence (6) gives

$$H^k(\Sigma) \cong H^k \operatorname{Hom}_+(\epsilon_+, \epsilon_+) \cong \mathrm{LCH}^{\epsilon_+}_{1-k}(\Lambda_+),$$

which is the Seidel isomorphism (following [33]). This theorem was conjectured by Seidel [35] and was proved by Dimitroglou Rizell [13].

If Λ_+, instead, is empty and $\mathcal{A}(\Lambda_-)$ has an augmentation ϵ_-, the long exact sequence (6) tells us that

$$H^k \operatorname{Hom}_+(\epsilon_-, \epsilon_-) \cong H^{k+1}(\Sigma, \Lambda_-) = \begin{cases} \mathbb{F} & \text{if } k = 1, \\ \mathbb{F}^{1-\chi(\Sigma)} & \text{if } k = 0, \\ 0 & \text{otherwise.} \end{cases}$$

However, by Sabloff duality (3),

$$\dim H^0 \operatorname{Hom}_+(\epsilon_-, \epsilon_-) = \dim H^2 \operatorname{Hom}_-(\epsilon_-, \epsilon_-) = \dim H^2 \operatorname{Hom}_+(\epsilon_-, \epsilon_-) = 0.$$

This is a contradiction since the unit $e_{\epsilon_-} = -y^\vee$ is always in $H^0 \operatorname{Hom}_+(\epsilon_-, \epsilon_-)$ and is not 0. Thus if Λ_+ is empty, then Λ_- does not admit any augmentation. This result was previously known [6; 12].

Remark 5.3 For the rest of the paper, we will focus on the case where Λ_+ and Λ_- are single component knots. Given the fact that there does not exist a compact Lagrangian manifold in $\mathbb{R} \times \mathbb{R}^3$ and Λ_- does not admit a cap (since Λ_- has an augmentation), we know that any cobordism Σ from Λ_- to Λ_+ must be connected.

Combining the long exact sequence (6) with the augmentation category map induced by the exact Lagrangian cobordism Σ, we have the following theorem.

Theorem 5.4 *Let Σ be an exact Lagrangian cobordism with Maslov number 0 from a Legendrian knot Λ_- to a Legendrian knot Λ_+. For $i = 1, 2$, assume ϵ_-^i is an augmentation of $\mathcal{A}(\Lambda_-)$ with a single basepoint and ϵ_+^i is the augmentation of $\mathcal{A}(\Lambda_+)$ induced by Σ. Then the map*

$$i^0\colon H^0 \operatorname{Hom}_+(\epsilon_+^1, \epsilon_+^2) \to H^0 \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2)$$

in the long exact sequence (6) *is an isomorphism. Moreover, we have that*

$$H^* \operatorname{Hom}_+(\epsilon_+^1, \epsilon_+^2) \cong H^* \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2) \oplus \mathbb{F}^{-\chi(\Sigma)}[1], \tag{7}$$

where $\mathbb{F}^{-\chi(\Sigma)}[1]$ denotes the vector space $\mathbb{F}^{-\chi(\Sigma)}$ in degree 1 and $\chi(\Sigma)$ is the Euler characteristic of the surface Σ.

Proof By Remark 5.3, we have

$$H^k(\Sigma, \Lambda_-) = \begin{cases} \mathbb{F}^{-\chi(\Sigma)} & \text{if } k = 1, \\ 0 & \text{otherwise.} \end{cases}$$

The long exact sequence (6) shows that for $k > 1$ or $k < 0$, the map

$$i^k\colon H^k \operatorname{Hom}_+(\epsilon_+^1, \epsilon_+^2) \to H^k \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2)$$

in the long exact sequence induces an isomorphism

$$H^k \operatorname{Hom}_+(\epsilon_+^1, \epsilon_+^2) \cong H^k \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2).$$

When $k = 0$ or 1, we have

$$\begin{aligned} 0 \to H^0 \operatorname{Hom}_+(\epsilon_+^1, \epsilon_+^2) \xrightarrow{i^0} H^0 \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2) \to \mathbb{F}^{-\chi(\Sigma)} \\ \to H^1 \operatorname{Hom}_+(\epsilon_+^1, \epsilon_+^2) \to H^1 \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2) \to 0. \end{aligned}$$

To finish the proof, we need to show

$$\dim(H^0 \operatorname{Hom}_+(\epsilon_+^1, \epsilon_+^2)) \ge \dim(H^0 \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2)).$$

Once this inequality holds, the fact that $i^0\colon H^0\,\mathrm{Hom}_+(\epsilon^1_+,\epsilon^2_+)\to H^0\,\mathrm{Hom}_+(\epsilon^1_-,\epsilon^2_-)$ is injective implies that it is an isomorphism. Note that the long exact sequence (6) is over the field $\mathbb{F}$. It follows that

$$H^1\,\mathrm{Hom}_+(\epsilon^1_+,\epsilon^2_+)\cong H^1\,\mathrm{Hom}_+(\epsilon^1_-,\epsilon^2_-)\oplus\mathbb{F}^{-\chi(\Sigma)}.$$

To prove the inequality, we exchange the positions of ϵ^1 and ϵ^2 in the long exact sequence (6) and get

$$\cdots\to H^k(\Sigma,\Lambda_-)\to H^k\,\mathrm{Hom}_+(\epsilon^2_+,\epsilon^1_+)\to H^k\,\mathrm{Hom}_+(\epsilon^2_-,\epsilon^1_-)\to H^{k+1}(\Sigma,\Lambda_-)\to\cdots,$$

which implies

$$H^2\,\mathrm{Hom}_+(\epsilon^2_+,\epsilon^1_+)\cong H^2\,\mathrm{Hom}_+(\epsilon^2_-,\epsilon^1_-).$$

By Sabloff duality (3), we have

$$\dim(H^0\,\mathrm{Hom}_-(\epsilon^1_\pm,\epsilon^2_\pm))=\dim(H^2\,\mathrm{Hom}_+(\epsilon^2_\pm,\epsilon^1_\pm)).$$

Thus $\dim(H^0\,\mathrm{Hom}_-(\epsilon^1_+,\epsilon^2_+))=\dim(H^0\,\mathrm{Hom}_-(\epsilon^1_-,\epsilon^2_-))$.

Since Λ_+ and Λ_- are both Legendrian knots with a single basepoint, we have the long exact sequence (2) for Λ_+ and Λ_-:

$$0\to H^0\,\mathrm{Hom}_-(\epsilon^1_\pm,\epsilon^2_\pm)\to H^0\,\mathrm{Hom}_+(\epsilon^1_\pm,\epsilon^2_\pm)\to H^0(\Lambda_\pm)\xrightarrow{\delta_\pm} H^1\,\mathrm{Hom}_-(\epsilon^1_\pm,\epsilon^2_\pm)\to\cdots.$$

From this long exact sequence, we have

$$\dim(H^0\,\mathrm{Hom}_+(\epsilon^1_\pm,\epsilon^2_\pm))=\dim(H^0\,\mathrm{Hom}_-(\epsilon^1_\pm,\epsilon^2_\pm))+\dim(\ker\delta_\pm).$$

Thus, to prove $\dim(H^0\,\mathrm{Hom}_+(\epsilon^1_+,\epsilon^2_+))\geq\dim(H^0\,\mathrm{Hom}_+(\epsilon^1_-,\epsilon^2_-))$, we only need to show $\dim(\ker\delta_+)\geq\dim(\ker\delta_-)$.

Recall that the cobordism Σ from Λ_- and Λ_+ induces an $\mathcal{A}_\infty$–category map

$$f\colon \mathcal{A}\mathrm{ug}_+(\Lambda_-)\to\mathcal{A}\mathrm{ug}_+(\Lambda_+)$$

in the way described in Section 3.2. In particular, we get the functor f_1 of augmentation categories on the level of morphisms:

$$f_1\colon \mathrm{Hom}_+(\epsilon^1_-,\epsilon^2_-)\to\mathrm{Hom}_+(\epsilon^1_+,\epsilon^2_+).$$

This map descends to the cohomology level as

$$f^*\colon H^*\,\mathrm{Hom}_+(\epsilon^1_-,\epsilon^2_-)\to H^*\,\mathrm{Hom}_+(\epsilon^1_+,\epsilon^2_+).$$

Notice that f_1 sends $\mathrm{Hom}_-(\epsilon^1_-,\epsilon^2_-)$ to $\mathrm{Hom}_-(\epsilon^1_+,\epsilon^2_+)$. Hence f_1 induces a map between the cohomology of the quotient chain complexes, denoted by f^* as well.

The following diagram commutes:

$$\begin{tikzcd}
0 \arrow[r] & H^0\mathrm{Hom}_-(\epsilon_+^1,\epsilon_+^2) \arrow[r] & H^0\mathrm{Hom}_+(\epsilon_+^1,\epsilon_+^2) \arrow[r] & H^0(\Lambda_+) \arrow[r, "\delta_+"] & H^1\mathrm{Hom}_-(\epsilon_+^1,\epsilon_+^2) \arrow[r] & \cdots \\
0 \arrow[r] & H^0\mathrm{Hom}_-(\epsilon_-^1,\epsilon_-^2) \arrow[r] \arrow[u, "f^*"] & H^0\mathrm{Hom}_+(\epsilon_-^1,\epsilon_-^2) \arrow[r] \arrow[u, "f^*"] & H^0(\Lambda_-) \arrow[r, "\delta_-"] \arrow[u, "f^*"] & H^1\mathrm{Hom}_-(\epsilon_-^1,\epsilon_-^2) \arrow[r] \arrow[u, "f^*"] & \cdots
\end{tikzcd}$$

Thus $f^*(\ker\delta_-) \subset \ker\delta_+$. Furthermore, notice that f_1 sends the generator $(y^-)^\vee \in C^0_{\mathrm{Morse}}(\Lambda_-)$ to the corresponding $(y^+)^\vee \in C^0_{\mathrm{Morse}}(\Lambda_+)$, and $C^{-1}_{\mathrm{Morse}}(\Lambda_+) = 0$. Hence f^* is injective on $H^0(\Lambda_-)$, which implies

$$\dim(\ker\delta_+) \geq \dim(\ker\delta_-). \qquad \square$$

If $\epsilon_-^1 = \epsilon_-^2 = \epsilon_-$ and ϵ_- comes from a Lagrangian filling L_-, then ϵ_+ also comes from the filling L_+, which is a concatenation of Σ and L_-. By Seidel's isomorphism (following [33]), we have $\mathrm{Hom}^k_+(\epsilon_\pm,\epsilon_\pm) \cong H^k(L_\pm)$, which implies that

$$H^k\,\mathrm{Hom}_+(\epsilon_+,\epsilon_+) \cong H^k\,\mathrm{Hom}_+(\epsilon_-,\epsilon_-) \quad \text{for } k \neq 1$$

and when $k=1$,

$$H^1\,\mathrm{Hom}_+(\epsilon_+,\epsilon_+) \cong H^1\,\mathrm{Hom}_+(\epsilon_-,\epsilon_-) \oplus \mathbb{F}^{-\chi(\Sigma)}.$$

Theorem 5.4 is a generalization of Seidel's isomorphism. Equation (7) holds even if ϵ_- does not come from a Lagrangian filling or ϵ_-^1 and ϵ_-^2 are not the same.

If the two augmentations are the same, we can identify the cohomology of Hom_+ space with the linearized contact homology by [33]:

$$\mathrm{Hom}^k_+(\epsilon,\epsilon) \cong \mathrm{LCH}^\epsilon_{1-k}(\Lambda).$$

Now we restate Theorem 5.4 in terms of linearized contact homology.

Corollary 5.5 *Let Σ be an exact Lagrangian cobordism with Maslov number 0 from a Legendrian knot Λ_- to a Legendrian knot Λ_+. Assume ϵ_- is an augmentation of $\mathcal{A}(\Lambda_-)$ and ϵ_+ is the augmentation of $\mathcal{A}(\Lambda_+)$ induced by Σ. Then*

$$\mathrm{LCH}^{\epsilon_+}_*(\Lambda_+) \cong \mathrm{LCH}^{\epsilon_-}_*(\Lambda_-) \oplus \mathbb{F}^{-\chi(\Sigma)}[0],$$

where $\mathbb{F}^{-\chi(\Sigma)}[0]$ denotes the vector space $\mathbb{F}^{-\chi(\Sigma)}$ in degree 0.

Therefore, if there exists an exact Lagrangian cobordism Σ from Λ_- to Λ_+, the Poincaré polynomial of the linearized contact homology of Λ_+ agrees with that of Λ_- in all degrees except 0. In degree 0 their coefficients differ by $-\chi(\Sigma)$. This gives a strong and computable obstruction to the existence of exact Lagrangian cobordisms. One can check the Poincaré polynomials of the linearized contact homology for any two Legendrian knots with small crossings through the atlas in [9]. If they do not

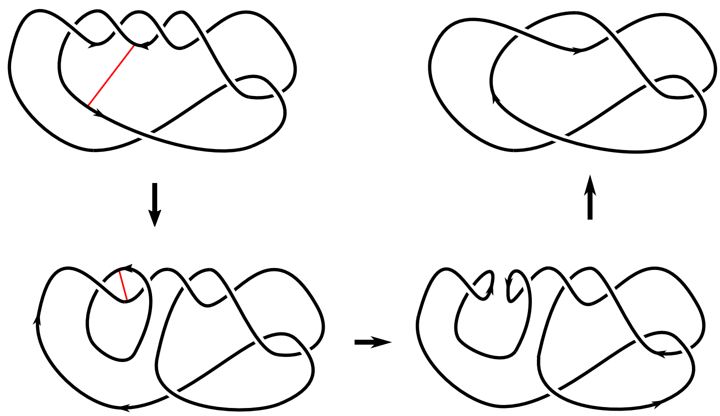

Figure 11: The topological cobordism between 6_1 and 4_1 can be achieved by two saddle moves along the red (straight) lines followed by an isotopy.

satisfy the relation given in Corollary 5.5, there does not exist an exact Lagrangian cobordism between them.

For example, let Λ_1 and Λ_2 be the Legendrian knots with maximum Thurston–Bennequin number of smooth knot types 4_1 and 6_1, respectively (as shown in Figure 2). There is a topological cobordism between 4_1 and 6_1 with genus 1 as shown in Figure 11. Moreover, the Thurston–Bennequin numbers of Λ_1 and Λ_2 are -3 and -5, respectively, which satisfy the Thurston–Bennequin number relation (1). Thus there possibly exists an exact Lagrangian cobordism from Λ_2 to Λ_1 with genus 1. However, we have the following proposition:

Proposition 5.6 *There does not exist an exact Lagrangian cobordism from Λ_2 to Λ_1 with Maslov number 0.*

Proof The Poincaré polynomials of the linearized contact homology for Λ_1 and Λ_2 are $t^{-1}+2t$ and $2t^{-1}+3t$, respectively. As a result of Corollary 5.5, there does not exist an exact Lagrangian cobordism from Λ_2 to Λ_1 with Maslov number 0. □

5.3 Geometric description of the differential map

Let Σ be an exact Lagrangian cobordism from a Legendrian knot Λ_- to a Legendrian knot Λ_+. For $i=1,2$, assume ϵ_-^i is an augmentation of $\mathcal{A}(\Lambda_-)$ and ϵ_+^i is the augmentation induced by Σ. So far we have two maps from $\mathrm{Hom}_+(\epsilon_-^1,\epsilon_-^2)$ to $\mathrm{Hom}_+(\epsilon_+^1,\epsilon_+^2)$. One is the geometric map d_{+-} in the differential of the Cthulhu chain complex $\mathrm{Cth}(\Sigma^1,\Sigma^2)$ defined by counting rigid holomorphic disks with boundary on $\Sigma^1\cup\Sigma^2$. The other map is the augmentation category map induced by Σ on the level of morphisms f_1, defined algebraically in Section 3.2. In this section, we will show that, with a choice of Morse function F on the cobordism Σ, the maps d_{+-} and f_1 are the same. To do that, we describe the two maps separately and then compare their images on each generator of $\mathrm{Hom}_+(\epsilon_-^1, \epsilon_-^2)$.

In order to describe d_{+-}, we want to interpret rigid holomorphic disks with boundary on $\Sigma^1 \cup \Sigma^2$ in terms of rigid holomorphic disks with boundary on Σ together with negative gradient flows of a Morse function. This is analogous to a result in [16], which gives a correspondence between rigid holomorphic disks with boundary on a 2–copy of a Legendrian submanifold L and rigid holomorphic disks with boundary on L together with negative gradient flows of a Morse function. Now let us describe the result in [16] in detail.

Let L be a Legendrian submanifold in the contact manifold $(P \times \mathbb{R}, \ker(dz - \theta))$, where $(P, d\theta)$ is an exact symplectic $2n$–dimensional manifold. Instead of considering holomorphic disks in the symplectization of $P \times \mathbb{R}$ with boundary on $\mathbb{R} \times L$, according to [13], we can consider holomorphic disks in P with boundary on $\pi(L)$, where π is the projection $P \times \mathbb{R} \to P$. See [16, Section 2.2.3] for the detailed definition of holomorphic disks with boundary on $\pi(L)$. As the points on $\pi(L)$ and points on L are in natural correspondence except that the double points of $\pi(L)$ correspond to the Reeb chords of L, we refer to the holomorphic disks as *J–holomorphic disks with boundary on L* as in [16], where J is a generic almost complex structure on P. Choose a Morse–Smale pair (f, g), where f is a Morse function $L \to \mathbb{R}$ and g is a Riemannian metric on L, such that (f, g, J) is *adjusted to* L in the sense of [16, Section 6.3]. Push L off through the Morse function f and get a 2–copy of L, denoted by $2L$. In order to describe rigid holomorphic disks with boundary on $2L$, we need to introduce the generalized disks determined by (f, g, J). A *generalized disk* is a pair (u, γ), where

- $u \in \mathcal{M}$ is a J–holomorphic disk with boundary on L;
- γ is a negative gradient flow of f with one end on the boundary of u and the other end at a critical point p of the Morse function f;
- the boundary of u and γ intersect transversely.

The point p is called a *negative Morse puncture* if the flow line γ flows toward p, and is called a *positive Morse puncture* if γ flows away from p. The formal dimension $\dim(u, \gamma)$ is defined by

$$\dim(u,\gamma) = \begin{cases} \dim \mathcal{M} + 1 + \mathrm{Ind}_f(p) - n & \text{if } p \text{ is a positive Morse puncture,} \\ \dim \mathcal{M} + 1 - \mathrm{Ind}_f(p) & \text{if } p \text{ is a negative Morse puncture.} \end{cases}$$

The generalized disk (u, γ) is called *rigid* if $\dim(u, \gamma) = 0$.

The rigid holomorphic disks with boundary on a 2–copy of L can be described as below in terms of whether their punctures are Morse Reeb chords or non-Morse Reeb chords (as defined in Section 3.2).

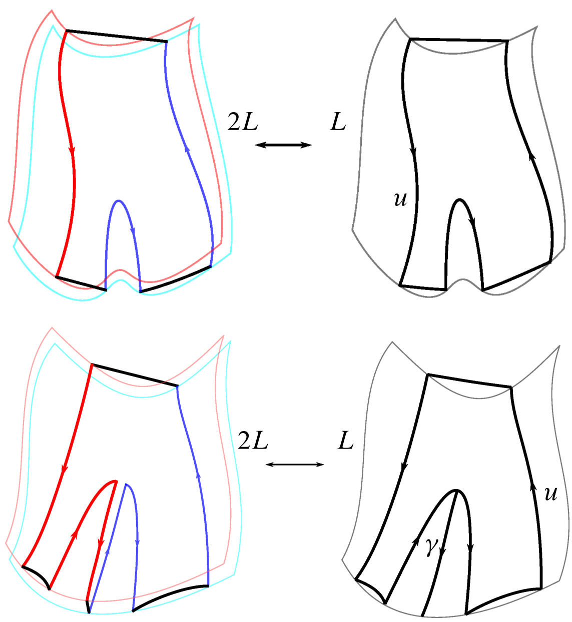

Figure 12: The correspondences in Lemma 5.7. The arrows indicate the orientations of holomorphic disks and the negative gradient flow line.

Lemma 5.7 [16, Theorem 3.6] *Let (f, g, J) be a triple as described above that is adjusted to the Legendrian submanifold L. Push L off through the Morse function f and get a 2–copy $2L$. There are the following bijective correspondences:*

- *Rigid holomorphic disks with boundary on $2L$ that have one positive puncture and one negative puncture at non-Morse mixed Reeb chords and the other punctures at pure Reeb chords are in one-to-one correspondence with rigid holomorphic disks with boundary on L as shown in Figure 12 (top).*
- *Rigid holomorphic disks with boundary on $2L$ that have exactly one puncture at a Morse Reeb chord are in one-to-one correspondence with rigid generalized disks (u, γ) determined by (f, g, J) as shown in Figure 12 (bottom).*
- *Rigid holomorphic disks with boundary on $2L$ that have two punctures at Morse Reeb chords are in one-to-one correspondence with rigid negative gradient flows of the Morse function f.*

In order to get an analogous description for a 2–copy of Σ, we need a result in [19] to relate rigid holomorphic disks with boundary on a cobordism Σ to rigid holomorphic disks with boundary on some Legendrian submanifold L_Σ.

Let us first construct the Legendrian submanifold L_Σ. Suppose Σ is an exact Lagrangian submanifold in $(\mathbb{R}\times\mathbb{R}^3, d(e^t\alpha))$ from Λ_- to Λ_+. Assume it is cylindrical outside $[-N+\delta, N-\delta]\times\mathbb{R}^3$, where δ is a small positive number. Under the symplectomorphism

$$\psi\colon (\mathbb{R}\times\mathbb{R}^3, d(e^t\alpha)) \to (T^*(\mathbb{R}_{>0}\times\mathbb{R}), d\theta), \quad (t,x,y,z)\mapsto((e^t,x),(z,e^t y)),$$

the cobordism Σ can be viewed as a cobordism in $(T^*(\mathbb{R}_{>0}\times\mathbb{R}), d\theta)$, where θ is the negative Liouville form of the cotangent bundle. Let $a_- = e^{-N}$ and $a_+ = e^N$. There is a small $\epsilon > 0$ such that Σ is cylindrical outside $T^*([a_-+\epsilon, a_+-\epsilon]\times\mathbb{R})$. Chopping off the ends of Σ, we get a cobordism in $T^*([a_-,a_+]\times\mathbb{R})$ with the canonical symplectic form. Lift it to a Legendrian submanifold $\overline{\Sigma}$ in the 1–jet space $J^1([a_-,a_+]\times\mathbb{R}) = T^*([a_-,a_+]\times\mathbb{R})\times\mathbb{R}$. The Legendrian $\overline{\Sigma}$ can be parametrized near the positive boundary $J^1((a_+-\epsilon, a_+]\times\mathbb{R})$ as

$$(s,\, x_+(q),\, z_+(q),\, sy_+(q),\, sz_+(q)+B_+) = j^1(sz_+(q)+B_+)$$

for some constant B_+, where $s = e^t$ and $(t,q)\in\Sigma\cap\big((\log(a_+-\epsilon), \log a_+]\times\mathbb{R}^3\big) = (\log(a_+-\epsilon), \log a_+]\times\Lambda_+$. Here $sz_+(q)+B_+$ may not be a function of (s,x). However, consider $\{(x_+(q), z_+(q)) \mid q\in\Lambda_+\}$, which is the front projection of Λ to the xz–plane. The cusps divide the front diagram of Λ_+ into pieces. Note that on each piece $z_+(p)$ is a perfect function of $x_+(p)$ and at each cusp, the two functions from different pieces match at the cusp. Therefore, we can write the parametrization as $j^1(sz_+(q)+B_+)$. Similarly, near the negative boundary $J^1([a_-, a_-+\epsilon)\times\mathbb{R})$, the Legendrian $\overline{\Sigma}$ can be parametrized as

$$(s,\, x_-(q),\, z_-(q),\, sy_-(q),\, sz_-(q)+B_-) = j^1(sz_-(q)+B_-),$$

where $(s,q)\in[a_-, a_-+\epsilon)\times\Lambda_-$ and B_- is a constant.

However, notice that $\overline{\Sigma}$ does not have any Reeb chords. Therefore, we consider the *Morse Legendrian* $\overline{\Sigma^{\mathrm{Mo}}}$, which is a Legendrian submanifold in $J^1([a_-,a_+]\times\mathbb{R})$ that agrees with $\overline{\Sigma}$ on $J^1((a_-+\epsilon, a_+-\epsilon)\times\mathbb{R})$. But near the $(\pm)$–boundary, the Morse Legendrian can be parametrized as $j^1\big((A_\pm \mp (s-a_\pm)^2)z_\pm(q)\big)$, ie

$$\big(s,\, x_\pm(q),\, \mp 2(s-a_\pm)z_\pm(q),\, (A_\pm\mp(s-a_\pm)^2)y_\pm(q),\, (A_\pm\mp(s-a_\pm)^2)z_\pm(q)\big), \tag{8}$$

where $A_\pm$ are positive constants. The key property of the Morse Legendrian is that the Reeb chords of $\overline{\Sigma^{\mathrm{Mo}}}$ on the $(\pm)$–boundary are in bijective correspondence with the Reeb chords of $\Lambda_\pm$, respectively.

There are isotopies from $sz_\pm(q)+B_\pm$ to $(A_\pm \mp (s-a_\pm)^2)z_\pm(q)$, respectively, which each induce a diffeomorphism from $\overline{\Sigma}$ to the Morse Legendrian $\overline{\Sigma^{\mathrm{Mo}}}$. Extend $\overline{\Sigma^{\mathrm{Mo}}}$ to a Legendrian submanifold L_Σ in $J^1(\mathbb{R}_{>0}\times\mathbb{R})$ by adding

$$j^1\big((A_+ - (s-a_+)^2)z_+(q)\big)$$

with $(s,q)\in(a_+,\infty)\times\Lambda_+$ to the positive boundary and adding

$$j^1\big((A_- + (s-a_-)^2)z_-(q)\big)$$

with $(s,q)\in(0,a_-)\times\Lambda_-$ to the negative boundary. In other words, when $s<a_-+\epsilon$ or $s>a_+-\epsilon$, we can parametrize L_Σ as (8). Note that

$$L_\Sigma\cap J^1\big((a_-+\epsilon,a_+-\epsilon)\times\mathbb{R}\big)=\overline{\Sigma^{\mathrm{Mo}}}.$$

Moreover, according to [20], there is a natural bijective correspondence between rigid holomorphic disks with boundary on L_Σ and rigid holomorphic disks with boundary on $\overline{\Sigma^{\mathrm{Mo}}}$. Combining this with a result in [19], we know that rigid holomorphic disks with boundary on an exact Lagrangian cobordism Σ are in one-to-one correspondence with rigid holomorphic disks with boundary on L_Σ that have positive (resp. negative) punctures at Reeb chords lying in the slice $s=a_+$ (resp. $s=a_-$). The proof of this result can be applied directly to the case of immersed exact Lagrangian submanifolds with cylindrical ends, where we only consider the rigid holomorphic disks with punctures on Reeb chords but no double points. Hence we have the following result for a 2–copy of Σ, denoted by $\Sigma\cup\Sigma'$.

Lemma 5.8 *Let Σ and Σ' be exact Lagrangian cobordisms from Λ_- to Λ_+ and from Λ'_- to Λ'_+, respectively. The Morse Legendrian $L_{\Sigma\cup\Sigma'}$ constructed above is a union of L_Σ and $L_{\Sigma'}$. Moreover, rigid holomorphic disks with boundary on $\Sigma\cup\Sigma'$ that have positive (resp. negative) punctures at Reeb chords of $\Lambda_+\cup\Lambda'_+$ (resp. $\Lambda_-\cup\Lambda'_-$) are in one-to-one correspondence with rigid holomorphic disks with boundary on $L_\Sigma\cup L_{\Sigma'}$ that have positive (resp. negative) punctures at the Reeb chords lying in the slice $s=a_+$ (resp. $s=a_-$).*

Note that the rigid holomorphic disks with boundary on $\Sigma\cup\Sigma'$ considered in Lemma 5.8 are not all the rigid holomorphic disks since we did not talk about holomorphic disks with punctures at double points. The disks we considered are the ones counted by d_{+-}.

In order to apply Lemma 5.7 to $L_{\Sigma\cup\Sigma'}$ and get the analogous correspondences for exact Lagrangian cobordisms, we need to view $L_{\Sigma'}$ as the 1–jet of a function

$$\widetilde{F}\colon L_\Sigma\to\mathbb{R}$$

in the neighborhood of L_Σ and show that $\widetilde{F}$ is Morse. To describe the function easily, pull it back to a function $\Sigma \to \mathbb{R}$, denoted by $\widetilde{F}$ as well. Note that $\widetilde{F} = F$ on $\Sigma \cap T^*([a_- + \epsilon, a_+ - \epsilon] \times \mathbb{R})$.

Now let us focus on the part $s \in (a_+ - \epsilon, \infty)$. Denote $L_\Sigma \cap J^1((a_+ - \epsilon, \infty) \times \mathbb{R})$ by $\partial_+(L_\Sigma)$ and denote $\Sigma \cap T^*((a_+ - \epsilon, \infty) \times \mathbb{R})$ by $\partial_+(\Sigma)$. One can check that the Reeb chords from $\partial_+(L_\Sigma)$ to $\partial_+(L_{\Sigma'})$ are in bijective correspondence with the Reeb chords from Λ_+ to Λ'_+ by a property of the Morse Legendrian. As a result, the only critical points of $\widetilde{F}$ on $\partial_+(\Sigma)$ are (s, q), where $s = a_+$ and $f'_+(q) = 0$.

Let π_1 and π_2 be the natural projections as follows:

$$\begin{tikzcd} & J^1(\mathbb{R}_{>0} \times \mathbb{R}_x) \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\ J^1(\mathbb{R}_{>0}) & & J^1(\mathbb{R}_x) \end{tikzcd}$$

First project $\partial_+(L_\Sigma)$ and $\partial_+(L_{\Sigma'})$ to $J^1(\mathbb{R}_x)$. We have

$$\pi_2(\partial_+(L_\Sigma)) = \big(x_+(q), (A_+ - (s - a_+)^2) y_+(q), (A_+ - (s - a_+)^2) z_+(q)\big)$$

and

$$\pi_2(\partial_+(L_{\Sigma'})) = \big(x'_+(q), (A_+ - (s - a_+)^2) y'_+(q), (A_+ - (s - a_+)^2) z'_+(q)\big).$$

Thus for fixed $s \in (a_+ - \epsilon, \infty)$, we have $\widetilde{F}(s, q) = (A_+ - (s - a_+)^2) f_+(q)$, where $f_+ = F|_{\{a_+\} \times \Lambda_+}$. Second, project $\partial_+(L_\Sigma)$ and $\partial_+(L_{\Sigma'})$ to $J^1(\mathbb{R}_{>0})$. We have

$$\pi_1(\partial_+(L_\Sigma)) = \big(s, -2(s - a_+) z_+(q), (A_+ - (s - a_+)^2) z_+(q)\big),$$

and

$$\pi_1(\partial_+(L_{\Sigma'})) = \big(s, -2(s - a_+) z'_+(q), (A_+ - (s - a_+)^2) z'_+(q)\big).$$

For a fixed $q \in \Lambda_+$, the only nondegenerate singularity of $\widetilde{F}(s, q)$ is a_+. In particular, it is a local maximum since $z'_+(q) > z_+(q)$, which comes from the fact that $f_+ > 0$ by the construction in Section 5.1. Therefore, we have

$$\mathrm{Ind}_{\widetilde{F}}(a_+, q) = \mathrm{Ind}_{f_+}(q) + 1.$$

Similarly, on the negative side, denote $F|_{\{a_-\} \times \Lambda_-}$ as f_-. The critical points of $\widetilde{F}$ on $\Sigma \cap T^*((-\infty, a_- + \epsilon) \times \mathbb{R})$ agree with the critical points of f_- that lie in the slice $s = a_-$. Moreover, the indices satisfy $\mathrm{Ind}_{\widetilde{F}}(a_-, q) = \mathrm{Ind}_{f_-}(q)$. Hence $\widetilde{F}$ is a Morse function.

Choose a Riemannian metric g on Σ and a generic almost complex structure J on $\mathbb{R} \times \mathbb{R}^3$ that is adjusted to cylindrical ends such that $(\widetilde{F}, g, J)$ is adjusted to L_Σ. Now we can apply Lemma 5.7 to the 2–copy $L_\Sigma \cup L_{\Sigma'}$.

Define a *generalized disk* to be a pair (u,γ) consisting of a J–holomorphic disk u with boundary on Σ as defined in Section 2.2 and a negative gradient flow line γ of $\widetilde{F}$ with one end on the boundary of u and one end at a critical point p of $\widetilde{F}$ such that the boundary of u intersects transversely with the negative gradient flow γ. The point p is called a *negative Morse puncture* if the flow line γ flows toward p, and is called a *positive Morse puncture* if γ flows away from p. The formal dimension $\dim(u,\gamma)$ is defined by

$$\dim(u,\gamma)=\begin{cases}\dim\mathcal{M}+1+\mathrm{Ind}_f(p)-2 & \text{if } p \text{ is a positive Morse puncture,}\\ \dim\mathcal{M}+1-\mathrm{Ind}_f(p) & \text{if } p \text{ is a negative Morse puncture.}\end{cases}\tag{9}$$

The generalized disk (u,γ) is called *rigid* if $\dim(u,\gamma)=0$. We have the following result, which is analogous to Lemma 5.7.

Theorem 5.9 *Let Σ be an exact Lagrangian cobordism in $(\mathbb{R}\times\mathbb{R}^3, d(e^t\alpha))$ from Λ_- to Λ_+ that is cylindrical outside $[-N+\delta, N-\delta]\times\mathbb{R}^3$. Let $F\colon \Sigma\to\mathbb{R}_{>0}$ be a positive Morse function. Push Σ off through F and get a new cobordism Σ'.*

Denote $F|_{\{N\}\times\Lambda_+}$ by f_+ and $F|_{\{-N\}\times\Lambda_-}$ by f_-. Define a new Morse function $\widetilde{F}\colon \Sigma\to\mathbb{R}$ satisfying the following properties:

- *The Morse function $\widetilde{F}$ satisfies $\widetilde{F}=F$ on $\Sigma\cap([-N+\delta,N-\delta]\times\mathbb{R}^3)$.*
- *On $\Sigma\cap((N-\delta,\infty)\times\mathbb{R}^3)$, all the critical points of the Morse function $\widetilde{F}$ lie in $\Sigma\cap(\{N\}\times\mathbb{R}^3)=\{N\}\times\Lambda_+$ and agree with the critical points of f_+. Moreover, at each critical point c, we have $\mathrm{Ind}_{\widetilde{F}}\, c=\mathrm{Ind}_{f_+} c+1$.*
- *On $\Sigma\cap((-\infty,-N+\delta)\times\mathbb{R}^3)$, all the critical points of the Morse function $\widetilde{F}$ lie in $\Sigma\cap(\{-N\}\times\mathbb{R}^3)=\{-N\}\times\Lambda_-$ and agree with the critical points of f_-. Moreover, at each critical point c, we have $\mathrm{Ind}_{\widetilde{F}}\, c=\mathrm{Ind}_{f_-} c$.*

The Riemannian metric g and almost complex structure J are chosen as above. Then we can describe the rigid holomorphic disks with boundary on $\Sigma\cup\Sigma'$ that have punctures on Reeb chords of $\Lambda_+\cup\Lambda'_+$ and $\Lambda_-\cup\Lambda'_-$ in terms of whether the Reeb chords are Morse or non-Morse as defined in Section 3.2:

1. *Rigid holomorphic disks with boundary on $\Sigma\cup\Sigma'$ that have two punctures at non-Morse mixed Reeb chords are in one-to-one correspondence with rigid holomorphic disks with boundary on Σ. See Figure 13 (top).*
2. *Rigid holomorphic disks with boundary on $\Sigma\cup\Sigma'$ that have exactly one puncture at a Morse Reeb chord are in one-to-one correspondence with rigid generalized disks (u,γ) determined by $(\widetilde{F},g,J)$. See Figure 13 (bottom).*
3. *Rigid holomorphic disks with boundary on $\Sigma\cup\Sigma'$ with two punctures at Morse Reeb chords are in one-to-one correspondence with rigid negative gradient flows of the Morse function $\widetilde{F}$ from a critical point on Λ_+ to a critical point on Λ_-.*

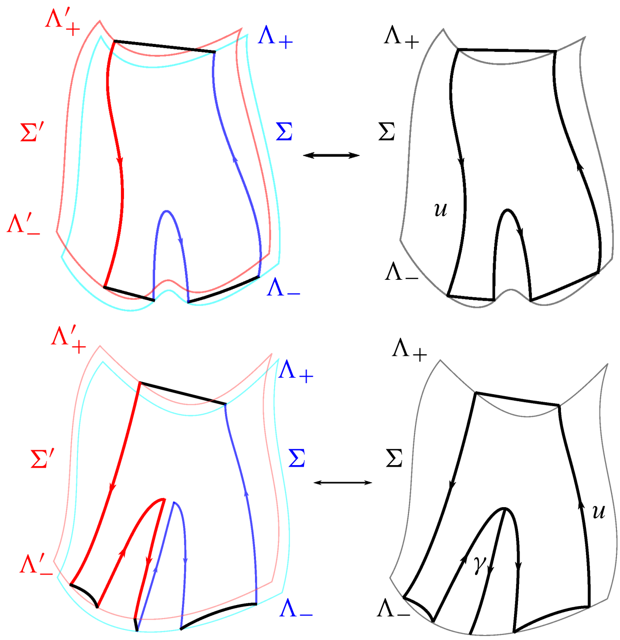

Figure 13: The correspondences in Theorem 5.9. The arrows denote the orientations of holomorphic disks and the negative gradient flow line.

Proof According to Lemma 5.8, rigid holomorphic disks with boundary on $\Sigma \cup \Sigma'$ that have two punctures at mixed Reeb chords correspond to rigid holomorphic disks with boundary on $L_\Sigma \cup L_{\Sigma'}$ that have positive (resp. negative) boundary at mixed Reeb chords lying in the slice $s = a_+$ (resp. $s = a_-$). By Lemma 5.7, these disks with boundary on $L_\Sigma \cup L_{\Sigma'}$ are in one-to-one correspondence with holomorphic disks with boundary on L_Σ that have positive (resp. negative) boundary at the Reeb chords lying in the slice $s = a_+$ (resp. $s = a_-$) together with Morse flow lines of $\widetilde{F}$. If it is a rigid Morse flow line of $\widetilde{F}$ on L_Σ, it flows from a critical point on Λ_+ to a critical point on Λ_-. Pull it back to a flow line Σ and get the correspondence (3). If it is a rigid holomorphic disk with boundary on L_Σ that has positive (resp. negative) boundary at the Reeb chords lying in the slice $s = a_+$ (resp. $s = a_-$), by [19], it corresponds to a rigid holomorphic disk with boundary on Σ, which is the correspondence (1). Otherwise, it is a rigid generalized disk (u, γ) determined by $(\widetilde{F}, g, J)$. From the construction of $\widetilde{F}$, one can note that all the critical points of $\widetilde{F}$ on $\Sigma \cap (\{N\} \times \mathbb{R}^3)$ are of index 1 or 2 while all the critical points of $\widetilde{F}$ on $\Sigma \cap (\{-N\} \times \mathbb{R}^3)$ are of index 0 or 1. By the dimension formula (9), the generalized disk (u, γ) is rigid if and only if u is a rigid holomorphic disk. Each rigid holomorphic disk u with boundary on L_Σ that has positive (resp. negative) boundary at the Reeb chords lying in the slice $s = a_+$ (resp. $s = a_-$) in turn corresponds to a rigid holomorphic disk with boundary on Σ. Pulling γ back to Σ, we get a rigid generalized disk on Σ determined by $(\widetilde{F}, g, J)$. Hence we get the correspondence (2). □

Recall f_1 is defined algebraically as follows. The exact Lagrangian cobordism Σ from a Legendrian knot Λ_- to a Legendrian knot Λ_+ induces a DGA map ϕ between the DGAs with a single basepoint by counting rigid holomorphic disks with boundary on Σ:

$$\phi\colon (\mathcal{A}(\Lambda_+), \partial) \to (\mathcal{A}(\Lambda_-), \partial),$$

as described in Section 2.2. This DGA map ϕ induces an A_∞–category map

$$f\colon \mathcal{A}\mathrm{ug}_+(\Lambda_-) \to \mathcal{A}\mathrm{ug}_+(\Lambda_+)$$

in the way described in Section 3.2. Restricting the category map on the level of morphisms, we have

$$f_1\colon \mathrm{Hom}_+(\epsilon^1_-, \epsilon^2_-) \to \mathrm{Hom}_+(\epsilon^1_+, \epsilon^2_+).$$

See calculation (4) for the explicit formula.

Theorem 5.10 *With a choice of Morse function $F\colon \Sigma \to \mathbb{R}$, we have $d_{+-} = f_1$.*

Proof We show $d_{+-} = f_1$ by checking their images on generators of $\mathrm{Hom}_+(\epsilon^1_-, \epsilon^2_-)$. Recall that $\mathrm{Hom}_+(\epsilon^1_-, \epsilon^2_-)$ is generated by the elements in $\mathrm{Hom}_-(\epsilon^1_-, \epsilon^2_-)$ that correspond to non-Morse Reeb chords and the elements in $T = \{x^\vee_-, y^\vee_-\}$ that correspond to Morse Reeb chords, respectively.

First consider the element $b^\vee$ in $\mathrm{Hom}_-(\epsilon^1_-, \epsilon^2_-)$. Notice that Morse Reeb chords are much shorter than non-Morse Reeb chords. The energy restriction ensures that $d_{+-}(b^\vee)$ does not include any element in T. Therefore d_{+-} sends $b^\vee$ to $a^\vee \in \mathrm{Hom}_-(\epsilon^1_+, \epsilon^2_+)$ by counting rigid holomorphic disks $u \in \mathcal{M}(a^{12}; \boldsymbol{p}^{11}, b^{12}, \boldsymbol{q}^{22})$ with boundary on $\Sigma^1 \cup \Sigma^2$, where $\boldsymbol{p}^{11}$ and $\boldsymbol{q}^{22}$ are words of pure Reeb chords of Λ^1_- and Λ^2_-, respectively. According to the correspondence (1) in Theorem 5.9, these disks correspond to rigid holomorphic disks $u \in \mathcal{M}(a; \boldsymbol{p}, b, \boldsymbol{q})$ with boundary on Σ (as shown in Figure 14), which are the disks counted by f_1. Notice that both d_{+-} and f_1 send $b^\vee$ to $|\mathcal{M}(a; \boldsymbol{p}, b, \boldsymbol{q})|\epsilon^1_-(\boldsymbol{p})\epsilon^2_-(\boldsymbol{q})a^\vee$, where $|\mathcal{M}(a; \boldsymbol{p}, b, \boldsymbol{q})|$ is the number of rigid disks in $\mathcal{M}$ counted with sign. Hence the definition of d_{+-} matches the definition of f_1 on $\mathrm{Hom}_-(\epsilon^1_-, \epsilon^2_-)$.

In order to simplify the map d_{+-}, we can choose a Morse function F such that the negative gradient flow of F flows from $*_+$ directly to $*_-$ without going through any critical points. We can further require that the negative gradient flow of F behave the same in a collar neighborhood of the flow line from $*_+$ to $*_-$ as shown in Figure 15.

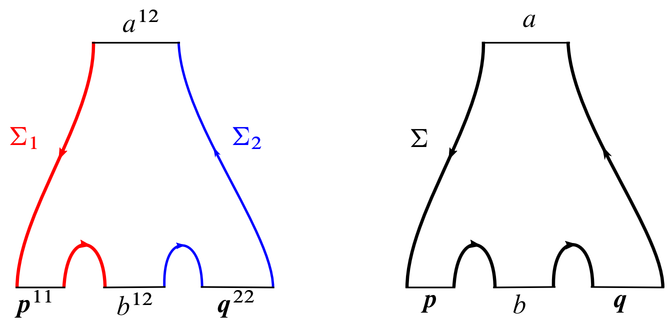

Figure 14: The disk on the left is counted by d_{+-} while the disk on the right is counted by ϕ.

As $x_{\pm}$ and $y_{\pm}$ sit right beside $*_{\pm}$, the negative gradient flow lines of $\widetilde{F}$ flow from x_+ and y_+ directly to x_- and y_-, respectively.

For the element $c^{\vee} \in T$, the map d_{+-} counts the rigid holomorphic disks in $\Sigma_1 \cup \Sigma_2$ that have a negative puncture at the Morse Reeb chord c. For the rigid disk that has a positive puncture at a Morse Reeb chord as well, according to the correspondence (3) in Theorem 5.9, it corresponds to a rigid Morse flow line of $\widetilde{F}$. The indices of $\widetilde{F}$ on y_-, x_-, y_+ and x_+ are 0, 1, 1 and 2, respectively. Therefore $d_{+-}(x_-^{\vee})$ has $x_+^{\vee}$ as a term and $d_{+-}(y_-^{\vee})$ has $y_+^{\vee}$ as a term. If the rigid disk has a positive puncture at a non-Morse mixed Reeb chord a^{12}, we denote it by $u \in \mathcal{M}(a^{12}; \boldsymbol{p}^{11}, c^{12}, \boldsymbol{q}^{22})$. By the correspondence (2) in Theorem 5.9, it corresponds to a rigid generalized disk (u, γ), where $u \in \mathcal{M}(a; \boldsymbol{p}, \boldsymbol{q})$ is a holomorphic disk with boundary on Σ and γ is a Morse flow of $\widetilde{F}$ that flows toward c (see Figure 16). Due to the dimension formula (9) of generalized disks, no rigid disk has a negative puncture at y_- since $\operatorname{Ind}_{\widetilde{F}} y_- = 0$ but $\dim \mathcal{M} \geq 0$. Hence $d_{+-}(y_-^{\vee}) = y_+^{\vee}$, which matches the definition of f_1 on y_k^-.

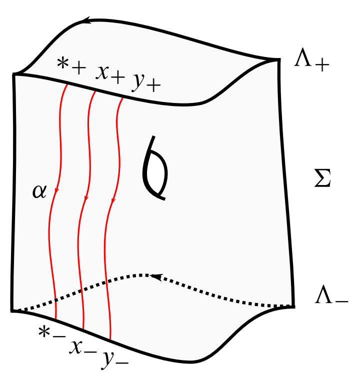

Figure 15: An example of Morse flows of F

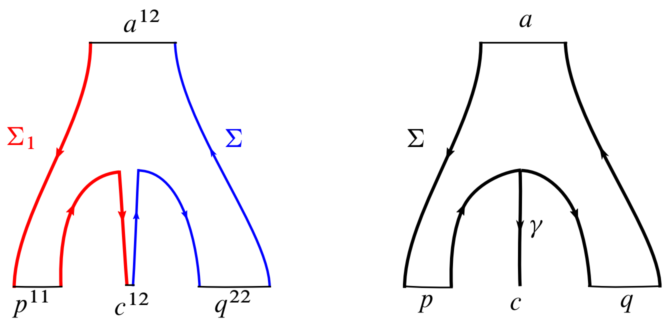

Figure 16: The disk on the left is counted by d_{+-} while the disk on the right is counted by ϕ.

For the element $x_-^\vee$, we know that $f_1(x_+^\vee)$ counts the element $a^\vee$ if t shows up in the image of the DGA map $\phi(a)$. In other words, there exists a rigid holomorphic disk $u \in \mathcal{M}(a; \boldsymbol{p}, \boldsymbol{q})$ with boundary on Σ, where $\boldsymbol{p}$ and $\boldsymbol{q}$ are words of pure Reeb chords of Λ_-, such that u has a nontrivial intersection number with α, where α is the curve from the basepoint $*_+$ to $*_-$. Each rigid holomorphic disk u contributes to $f_1(x_+^\vee)$ a term of $a^\vee$ with coefficient $s(u,\alpha)\epsilon_-^1(\boldsymbol{p})\epsilon_-^2(\boldsymbol{q})a^\vee$, where $s(u,\alpha)$ is the intersection number of the boundary of u and α. We can make the Morse function F satisfy the property that the negative gradient flow line γ of F from x_+ to x_- is parallel to α and of the same orientation. For each intersection point p_i of the boundary of u and γ, denote the part of γ from p_i to c by γ_i. By the correspondence (3) in Theorem 5.9, the rigid generalized disk (u, γ_i) corresponds to a rigid holomorphic disk in $\mathcal{M}(a^{12}; \boldsymbol{p}^{11}, c^{12}, \boldsymbol{q}^{22})$ with boundary on $\Sigma^1 \cup \Sigma^2$, and hence contributes to $d_{+-}(x_-^\vee)$ with a term $s(u,\gamma_i)\epsilon_-^1(\boldsymbol{p}^{11})\epsilon_-^2(\boldsymbol{q}^{22})a^\vee$, where $s(u,\gamma_i)$ is the sign of the intersection. Summing over all the intersections of the boundary of u and γ, the rigid holomorphic disk u contributes $s(u,\gamma)\epsilon_-^1(\boldsymbol{p})\epsilon_-^2(\boldsymbol{q})a^\vee$ to $d_{+-}(x_-^\vee)$, where $s(u,\gamma)$ is the intersection number of the boundary of u and γ. Therefore, we have $d_{+-} = f_1$ on $x_-^\vee$. □

5.4 Aside

In this section, we describe the differential map of the Cthulhu chain complex in terms of holomorphic disks with boundary on Σ and Morse flow lines. This allows us to recover the long exact sequences in [6]. The theorem in this section is stated without rigorous proof. But it will not be used elsewhere in the paper.

In Section 5.3, we only need to describe the rigid disks with boundary on $\Sigma^1 \cup \Sigma^2$ that have punctures at Reeb chords. Hence we only have correspondences for those types of disks. However, the method should work for all the rigid holomorphic disks with boundary on $\Sigma^1 \cup \Sigma^2$ including the disks counted by d_{+0} and d_{0-}. We state the following theorem without proof.

Theorem 5.11 *Let Σ be an exact Lagrangian cobordism from Λ_- to Λ_+ and $\Sigma^1 \cup \Sigma^2$ be a 2–copy of Σ as constructed in Section 5.1. For $i = 1, 2$, assume ϵ_-^i is an augmentation of $\mathcal{A}(\Lambda_-)$ and ϵ_+^i is the augmentation of $\mathcal{A}(\Lambda_+)$ induced by Σ. For η small enough, the Cthulhu chain complex can be decomposed into five parts:*

$$\mathrm{Cth}^k(\Sigma^1, \Sigma^2) = \mathrm{Hom}_-^{k-1}(\epsilon_+^1, \epsilon_+^2) \oplus C^{k-1}_{\mathrm{Morse}} f_+ \oplus C^k_{\mathrm{Morse}} F \oplus \mathrm{Hom}_-^k(\epsilon_-^1, \epsilon_-^2) \oplus C^k_{\mathrm{Morse}} f_-.$$

Under this decomposition, the differential can be written as

$$d = \begin{pmatrix} m_1 & d_{+f_+} & d_{+F} & d_{+-} & d_{+f_-} \\ 0 & d_{f_+} & d_{f_+F} & 0 & d_{f_+f_-} \\ 0 & 0 & d_F & 0 & d_{Ff_-} \\ 0 & 0 & 0 & m_1 & d_{-f_-} \\ 0 & 0 & 0 & 0 & d_{f_-} \end{pmatrix}.$$

Moreover:

1. *The holomorphic disks counted by d_{+F} and d_{+f_-} are in one-to-one correspondence with rigid generalized disks on Σ determined by $(\widetilde{F}, g, J)$.*
2. *The holomorphic disks counted by d_{+-} are in one-to-one correspondence with rigid holomorphic disks with boundary on Σ.*
3. *The holomorphic disks counted by d_{f_+F}, $d_{f_+f_-}$ and d_{Ff_-} are in one-to-one correspondence with the rigid Morse flow lines of $\widetilde{F}$.*

This theorem is similar to the conjectural analytic Lemma 4.11 in [15], which describes the correspondence in the case of exact Lagrangian fillings.

We next discuss how to recover the three long exact sequences in [6] from this chain complex.

(1) Decompose the Cthulhu chain complex as

$$\mathrm{Hom}_-^{k-1}(\epsilon_+^1, \epsilon_+^2) \oplus (C^{k-1}_{\mathrm{Morse}} f_+ \oplus C^k_{\mathrm{Morse}} F) \oplus \mathrm{Hom}_+^k(\epsilon_-^1, \epsilon_-^2).$$

Notice that the chain complex

$$\left(C^{k-1}_{\mathrm{Morse}} f_+ \oplus C^k_{\mathrm{Morse}} F, \begin{pmatrix} d_{f_+} & d_{f_+F} \\ 0 & d_F \end{pmatrix} \right)$$

can be identified with the Morse cochain complex $(C^k_{\mathrm{Morse}} \overline{F}, d_{\overline{F}})$ induced by a Morse function $\overline{F}$, where $\overline{F}$ agrees with $\widetilde{F}$ near Λ_+ and agrees with F elsewhere. Hence

$$H^*(C^k_{\mathrm{Morse}} \overline{F}, d_{\overline{F}}) = H^k(\Sigma, \Lambda_+ \cup \Lambda_-).$$

Therefore, we have the following long exact sequence:

$$\cdots \to H^k(\Sigma, \Lambda_+ \cup \Lambda_-) \to H^k \operatorname{Hom}_-(\epsilon_+^1, \epsilon_+^2) \to H^k \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2) \\ \to H^{k+1}(\Sigma, \Lambda_+ \cup \Lambda_-) \to \cdots,$$

which is Theorem 1.5 of [6].

(2) To obtain another long exact sequence, view the Cthulhu chain complex as a direct sum of $\operatorname{Hom}_-^{k-1}(\epsilon_+^1, \epsilon_+^2)$, $C_{\text{Morse}}^{k-1} f_+ \oplus C_{\text{Morse}}^k F \oplus \operatorname{Hom}_-^k(\epsilon_-^1, \epsilon_-^2)$ and the Morse cochain complex $C_{\text{Morse}}^k f_-$. We have

$$\cdots \to H^{k-1}(\Lambda_-) \to H^k(\Sigma, \Lambda_+ \cup \Lambda_-) \oplus H^k \operatorname{Hom}_-(\epsilon_-^1, \epsilon_-^2) \to H^k \operatorname{Hom}_-(\epsilon_+^1, \epsilon_+^2) \\ \to H^k(\Lambda_-) \to \cdots,$$

which is Theorem 1.6 in [6].

(3) Rewrite the Cthulhu chain complex as

$$\begin{aligned} \operatorname{Cth}^k(\Sigma^1, \Sigma^2) &= \operatorname{Hom}_-^{k-1}(\epsilon_+^1, \epsilon_+^2) \oplus \operatorname{Hom}_-^k(\epsilon_-^1, \epsilon_-^2) \oplus C_{\text{Morse}}^{k-1} f_+ \oplus C_{\text{Morse}}^k F \oplus C_{\text{Morse}}^k f_- \\ &= \operatorname{Hom}_-^{k-1}(\epsilon_+^1, \epsilon_+^2) \oplus \operatorname{Hom}_-^k(\epsilon_-^1, \epsilon_-^2) \oplus C_{\text{Morse}}^k \widetilde{F} \end{aligned}$$

with the differential

$$d = \begin{pmatrix} m_1 & * & * \\ 0 & m_1 & * \\ 0 & 0 & d_{\widetilde{F}} \end{pmatrix}.$$

We have the long exact sequence

$$\cdots \to H^k \operatorname{Hom}_-(\epsilon_-^1, \epsilon_-^2) \to H^k \operatorname{Hom}_-(\epsilon_+^1, \epsilon_+^2) \\ \to H^{k+1}(\Sigma, \Lambda_+) \to H^{k+1} \operatorname{Hom}_-(\epsilon_-^1, \epsilon_-^2) \to \cdots,$$

which is Theorem 1.4 in [6].

One may get other long exact sequences from the Cthulhu chain complex above. One example is

$$\cdots \to H^k \operatorname{Hom}_-(\epsilon_-^1, \epsilon_-^2) \to H^k \operatorname{Hom}_+(\epsilon_+^1, \epsilon_+^2) \\ \to H^k(\Sigma) \to H^{k+1} \operatorname{Hom}_-(\epsilon_-^1, \epsilon_-^2) \to \cdots,$$

which is obtained by decomposing the Cthulhu chain complex as a direct sum of $\operatorname{Hom}_+^{k-1}(\epsilon_+^1, \epsilon_+^2)$, $\operatorname{Hom}_-^k(\epsilon_-^1, \epsilon_-^2)$ and $C_{\text{Morse}}^k F \oplus C_{\text{Morse}}^k f_-$.

5.5 Injectivity

Theorem 5.10 implies that d_{+-} is a chain map and hence it gives the Cthulhu chain complex a stronger algebraic structure. In this section, we use this algebraic information to deduce that the augmentation category map induced by the exact Lagrangian cobordism Σ is injective on the level of equivalence classes of objects. And its induced map on the cohomology category $H^*\mathcal{A}\mathrm{ug}_+$ is faithful.

Notice that $d_{+-} = f_1$ implies that d_{+-} is a chain map and thus induces maps $d^k_{+-}\colon H^k\,\mathrm{Hom}_+(\epsilon^1_-, \epsilon^2_-) \to H^k\,\mathrm{Hom}_+(\epsilon^1_+, \epsilon^2_+)$ for $k \in \mathbb{Z}$. Then we have the following theorem deduced from the double cone structure of the Cthulhu chain complex. We would like to thank the referee for pointing out this theorem.

Theorem 5.12 *Let Σ be an exact Lagrangian cobordism from a Legendrian knot Λ_- to a Legendrian knot Λ_+ with Maslov number 0. For $i = 1, 2$, assume ϵ^i_- is an augmentation of $\mathcal{A}(\Lambda_-)$ and ϵ^i_+ is the augmentation of $\mathcal{A}(\Lambda_+)$ induced by Σ. With the same choice of Morse function as in Theorem 5.10, we have the following statement.*

For fixed $k \in \mathbb{Z}$, the map

$$i^k\colon H^k\,\mathrm{Hom}_+(\epsilon^1_+, \epsilon^2_+) \to H^k\,\mathrm{Hom}_+(\epsilon^1_-, \epsilon^2_-)$$

in the long exact sequence (6) *is injective (resp. surjective) if and only if the map*

$$d^k_{+-}\colon H^k\,\mathrm{Hom}_+(\epsilon^1_-, \epsilon^2_-) \to H^k\,\mathrm{Hom}_+(\epsilon^1_+, \epsilon^2_+)$$

is surjective (resp. injective).

Proof We will first prove that i^k is surjective if and only if d^k_{+-} is injective for fixed k.

Consider the Cthulhu chain complex as a mapping cone of $\Phi\colon \mathrm{Hom}_+(\epsilon^1_-, \epsilon^2_-) \to \mathrm{Cone}(d_{+0})$, where $\Phi = d_{+-} + d_{0-}$. The trivial cohomology of the Cthulhu chain complex implies that Φ induces isomorphisms

$$\Phi^k\colon H^k\,\mathrm{Hom}_+(\epsilon^1_-, \epsilon^2_-) \to H^k\,\mathrm{Cone}(d_{+0}) \quad \text{for } k \in \mathbb{Z}.$$

The following diagram commutes:

$$\begin{tikzcd}
\cdots \arrow[r] & H^k\,\mathrm{Hom}_+(\epsilon^1_+, \epsilon^2_+) \arrow[r, "i_k"] & H^k\,\mathrm{Cone}(d_{+0}) \arrow[r, "j_k"] & H^{k+1}(C_{\mathrm{Morse}}F) \arrow[r] & \cdots \\
& & H^k\mathrm{Hom}_+(\epsilon^1_-, \epsilon^2_-) \arrow[u, "\Phi^k", "\cong"'] \arrow[ur, "d^k_{0-}"'] & &
\end{tikzcd}$$

Comparing this to the long exact sequence (6), we know that $i^k = (\Phi^k)^{-1} \circ i_k$. Notice that d_{+-} is a chain map and thus $d_{+0} \circ d_{0-} = 0$. It is not hard to show that $\Phi^k = d_{+-}^k + d_{0-}^k$. Thus

$$H^k \operatorname{Cone}(d_{+0}) \cong d_{+-}^k(H^k \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2)) \oplus d_{0-}^k(H^k \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2)).$$

Since we are working over the field $\mathbb{F}$, we have the following relation on dimensions:

$$\begin{aligned}\dim(H^k \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2)) \\ &= \dim(H^k \operatorname{Cone}(d_{+0})) \\ &= \dim(d_{+-}^k(H^k \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2))) + \dim(d_{0-}^k(H^k \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2))).\end{aligned}$$

Therefore $\dim(d_{+-}^k(H^k \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2))) \le \dim(H^k \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2))$ and the equality holds if and only if $d_{0-}^k = 0$, which is equivalent to the condition that i^k is surjective. Hence d_{+-}^k is injective if and only if i^k is surjective.

The proof of the statement that i^k is injective if and only if d_{+-}^k is surjective is basically the same if we consider the Cthulhu chain complex as a mapping cone of $\Psi\colon \operatorname{Cone}(d_{0-}) \to \operatorname{Hom}_+(\epsilon_+^1, \epsilon_+^2)$, where $\Psi = d_{+0} + d_{+-}$. □

Thanks to Theorem 5.10, we know that d_{+-} agrees with f_1. Theorem 5.4 shows that i^0 is both injective and surjective. Therefore we have the following corollary.

Corollary 5.13 *Let f^* denote the induced map of f_1 on cohomology. Then f^* restricted on degree-0 cohomology,*

$$f^*\colon H^0 \operatorname{Hom}_+(\epsilon_-^1, \epsilon_-^2) \to H^0 \operatorname{Hom}_+(\epsilon_+^1, \epsilon_+^2),$$

is an isomorphism.

Theorem 5.14 *Let Σ be an exact Lagrangian cobordism with Maslov number 0 from a Legendrian knot Λ_- to a Legendrian knot Λ_+. The A_∞–category map $f\colon \mathcal{A}ug_+(\Lambda_-) \to \mathcal{A}ug_+(\Lambda_+)$ induced by the exact Lagrangian cobordism Σ is injective on the level of equivalence classes of objects.*

In other words, for $i = 1, 2$, assume ϵ_-^i is an augmentation of $\mathcal{A}(\Lambda_-)$ with a single basepoint and ϵ_+^i is the augmentation of $\mathcal{A}(\Lambda_+)$ with a single basepoint induced by Σ. If ϵ_+^1 and ϵ_+^2 are equivalent in $\mathcal{A}ug_+(\Lambda_+)$, then ϵ_-^1 and ϵ_-^2 are equivalent in $\mathcal{A}ug_+(\Lambda_-)$.

Proof Since ϵ_+^1 and ϵ_+^2 are equivalent in $\mathcal{A}ug_+(\Lambda_+)$, there exist

$$[\alpha_+] \in H^0 \operatorname{Hom}_+(\epsilon_+^1, \epsilon_+^2) \quad \text{and} \quad [\beta_+] \in H^0 \operatorname{Hom}_+(\epsilon_+^2, \epsilon_+^1)$$

such that

$$[m_2(\alpha_+,\beta_+)] = [e_{\epsilon_+^2}] \in H^0\operatorname{Hom}_+(\epsilon_+^2,\epsilon_+^2)$$

and

$$[m_2(\beta_+,\alpha_+)] = [e_{\epsilon_+^1}] \in H^0\operatorname{Hom}_+(\epsilon_+^1,\epsilon_+^1).$$

Corollary 5.13 shows that $f^*\colon H^0\operatorname{Hom}_+(\epsilon_-^1,\epsilon_-^2) \to H^0\operatorname{Hom}_+(\epsilon_+^1,\epsilon_+^2)$ is an isomorphism. Hence there exists $[\alpha_-] \in H^0\operatorname{Hom}_+(\epsilon_-^1,\epsilon_-^2)$ such that

$$f^*([\alpha_-]) = [\alpha_+] \in H^0\operatorname{Hom}_+(\epsilon_+^1,\epsilon_+^2).$$

Similarly, there exists $[\beta_-] \in H^0\operatorname{Hom}_+(\epsilon_-^2,\epsilon_-^1)$ such that

$$f^*([\beta_-]) = [\beta_+] \in H^0\operatorname{Hom}_+(\epsilon_+^2,\epsilon_+^1).$$

Moreover, we have

$$\begin{aligned} f^*[m_2(\alpha_-,\beta_-)] &= m_2(f^*([\alpha_-]), f^*([\beta_-])) \\ &= m_2([\alpha_+],[\beta_+]) \\ &= [e_{\epsilon_+^2}] \in H^0\operatorname{Hom}_+(\epsilon_+^2,\epsilon_+^2). \end{aligned}$$

Notice that f sends $y_-^\vee \in \operatorname{Hom}_+^0(\epsilon_-^2,\epsilon_-^2)$ to $y_+^\vee \in \operatorname{Hom}_+^0(\epsilon_+^2,\epsilon_+^2)$ and hence we have $f^*[e_{\epsilon_-^2}] = [e_{\epsilon_+^2}]$. By Corollary 5.13, the map

$$f^*\colon H^0\operatorname{Hom}_+(\epsilon_-^2,\epsilon_-^2) \to H^0\operatorname{Hom}_+(\epsilon_+^2,\epsilon_+^2)$$

is an isomorphism. Hence $[m_2(\alpha_-,\beta_-)] = [e_{\epsilon_-^2}] \in H^0\operatorname{Hom}_+(\epsilon_-^2,\epsilon_-^2)$. Similarly, $[m_2(\beta_-,\alpha_-)] = [e_{\epsilon_-^1}] \in H^0\operatorname{Hom}_+(\epsilon_-^1,\epsilon_-^1)$. Therefore ϵ_-^1 and ϵ_-^2 are equivalent in $\mathcal{A}\mathrm{ug}_+(\Lambda_-)$. □

In addition, the exact Lagrangian cobordism Σ described above also induces a category functor on the cohomology category

$$\tilde{f}\colon H^*\mathcal{A}\mathrm{ug}_+(\Lambda_-) \to H^*\mathcal{A}\mathrm{ug}_+(\Lambda_+),$$

as described in Section 3.1.

We have the following statement.

Theorem 5.15 *Let Σ be an exact Lagrangian cobordism from a Legendrian knot Λ_- to a Legendrian knot Λ_+ with Maslov number 0. The corresponding cohomology category map $\tilde{f}\colon H^*\mathcal{A}\mathrm{ug}_+(\Lambda_-) \to H^*\mathcal{A}\mathrm{ug}_+(\Lambda_+)$ induced by Σ is faithful. Moreover, if $\chi(\Sigma) = 0$, this functor is fully faithful.*

Proof Notice that the category map $\widetilde{f}$ restricted on the level of morphisms is

$$f^*\colon H^*\operatorname{Hom}_+(\epsilon_-^1,\epsilon_-^2)\to H^*\operatorname{Hom}_+(\epsilon_+^1,\epsilon_+^2).$$

The long exact sequence (6) tells us that the i^k are surjective for all $k\in\mathbb{Z}$. By Theorem 5.12, we know that f^* is injective. Thus $\widetilde{f}$ is faithful.

In particular, if $\chi(\Sigma)=0$, Theorem 5.4 implies that the i^k are isomorphisms for all $k\in\mathbb{Z}$. Therefore, by Theorem 5.12, the map f^* is an isomorphism, which implies that $\widetilde{f}$ is fully faithful. □

As a result of Theorem 5.14, there is an induced map from the equivalence classes of augmentations of Λ_- to the equivalence classes of augmentations of $\mathcal{A}\mathrm{ug}_+(\Lambda_+)$. Thus the number of equivalence classes of augmentations of $\mathcal{A}(\Lambda_-)$ is less than or equal to the number of equivalence classes of augmentations of $\mathcal{A}(\Lambda_+)$. However, the equivalence classes of augmentations is difficult to count in general. Ng, Rutherford, Shende and Sivek [32] introduced another way to count objects: the homotopy cardinality of $\pi_{\geq 0}\,\mathcal{A}\mathrm{ug}_+(\Lambda;\mathbb{F}_q)^*$, where $\mathbb{F}_q$ is a finite field. This can be computed using ruling polynomials.

The *homotopy cardinality* is defined by

$$\pi_{\geq 0}\,\mathcal{A}\mathrm{ug}_+(\Lambda;\mathbb{F}_q)^* = \sum_{[\epsilon]\in\mathcal{A}\mathrm{ug}_+(\Lambda;\mathbb{F}_q)/\sim}\frac{1}{|\mathrm{Aut}(\epsilon)|}\cdot\frac{|H^{-1}\operatorname{Hom}_+(\epsilon,\epsilon)|\cdot|H^{-3}\operatorname{Hom}_+(\epsilon,\epsilon)|\cdots}{|H^{-2}\operatorname{Hom}_+(\epsilon,\epsilon)|\cdot|H^{-4}\operatorname{Hom}_+(\epsilon,\epsilon)|\cdots},$$

where $[\epsilon]$ is the equivalence class of ϵ in the augmentation category $\mathcal{A}\mathrm{ug}_+(\Lambda)$ and $|\mathrm{Aut}(\epsilon)|$ is the number of invertible elements in $H^0\operatorname{Hom}_+(\epsilon,\epsilon)$.

Corollary 5.16 *Let Σ be a spin exact Lagrangian cobordism from a Legendrian knot Λ_- to a Legendrian knot Λ_+ with Maslov number 0. Then for any finite field $\mathbb{F}_q$, we have*

$$\pi_{\geq 0}\,\mathcal{A}\mathrm{ug}_+(\Lambda_-;\mathbb{F}_q)^*\leq\pi_{\geq 0}\,\mathcal{A}\mathrm{ug}_+(\Lambda_+;\mathbb{F}_q)^*.$$

Proof Assume $[\epsilon_-]$ is an equivalence class in $\mathcal{A}\mathrm{ug}_+(\Lambda_-;\mathbb{F}_q)$ and $[\epsilon_+]$ is the induced equivalence class in $\mathcal{A}\mathrm{ug}_+(\Lambda_+;\mathbb{F}_q)$. Theorem 5.4 implies

$$H^k\operatorname{Hom}_+(\epsilon_-,\epsilon_-)\cong H^k\operatorname{Hom}_+(\epsilon_+,\epsilon_+)\quad\text{for } k<1.$$

In particular, we have $H^0\operatorname{Hom}_+(\epsilon_-,\epsilon_-)\cong H^0\operatorname{Hom}_+(\epsilon_+,\epsilon_+)$, which implies that $|\mathrm{Aut}(\epsilon_-)|=|\mathrm{Aut}(\epsilon_+)|$.

Notice that $\mathcal{A}\mathrm{ug}_+(\Lambda_+;\mathbb{F}_q)$ may have more equivalence classes than $\mathcal{A}\mathrm{ug}_+(\Lambda_-;\mathbb{F}_q)$. Therefore, we have

$$\pi_{\geq 0}\,\mathcal{A}\mathrm{ug}_+(\Lambda_-;\mathbb{F}_q)^* \leq \pi_{\geq 0}\,\mathcal{A}\mathrm{ug}_+(\Lambda_+;\mathbb{F}_q)^*. \qquad \square$$

From [32, Corollary 2], this cardinality can be related to the ruling polynomial in the following way:

$$\pi_{\geq 0}\,\mathcal{A}\mathrm{ug}_+(\Lambda;\mathbb{F}_q)^* = q^{\mathrm{tb}(\Lambda)/2} R_\Lambda(q^{1/2}-q^{-1/2}).$$

Recall that a *normal ruling* R is a decomposition of the front projection of Λ into embedded disks connected by switches that satisfy certain requirements (see details in [8]). The *ruling polynomial* is defined by

$$R_\Lambda(z) = \sum_R z^{\#(\text{switches})-\#(\text{disks})}.$$

Corollary 5.17 *Suppose there is a spin exact Lagrangian cobordism from a Legendrian knot Λ_- to a Legendrian knot Λ_+ with Maslov number 0. Then the ruling polynomials R_{Λ_-} and R_{Λ_+} satisfy*

$$R_{\Lambda_-}(q^{1/2}-q^{-1/2}) \leq q^{-\chi(\Sigma)/2} R_{\Lambda_+}(q^{1/2}-q^{-1/2})$$

for any q that is a power of a prime number.

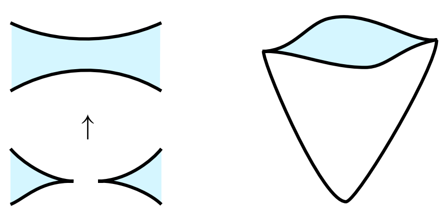

Figure 17: The relation between rulings of Legendrian submanifolds that are related by a pinch move (left) or a minimum cobordism (right)

When Σ is decomposable, ie consists of pinch moves and minimum cobordisms [19], there is a map from the rulings of Λ_- to rulings of Λ_+. For each pinch move or minimal cobordism, any normal ruling of the bottom knot gives a normal ruling of the top knot, as shown in Figure 17. Moreover, different rulings of the bottom knot give different rulings of the top knot. Therefore the ruling polynomials of Λ_+ and Λ_- satisfy the relation in Corollary 5.17. This corollary shows that the result is true even if the cobordism is not decomposable.

One can check the atlas in [9] for the ruling polynomials of small crossing Legendrian knots. This corollary gives a new and easily computable obstruction to the existence of exact Lagrangian cobordisms. We can use Corollary 5.17 to give a new proof of the follow theorem which is a result in [5] and was reproved in [11].

Theorem 5.18 [5] *Lagrangian concordance is not a symmetric relation.*

Proof Consider the Legendrian knot Λ of smooth knot type $m(9_{46})$ with maximum Thurston–Bennequin number and the Legendrian unknot Λ_0 as shown in Figure 18. There is an exact Lagrangian concordance from the Λ_0 to Λ by doing a pinch move at the red (straight) line in Figure 18 and Legendrian isotopy. However, there does not exist an exact Lagrangian concordance from Λ to Λ_0 since the ruling polynomial of Λ is 2 while the ruling polynomial of Λ_0 is 1. □

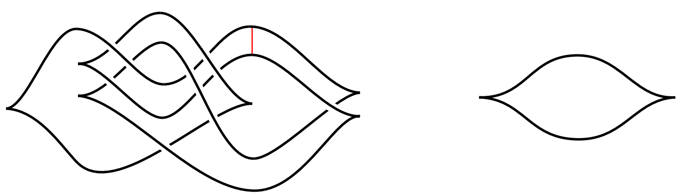

Figure 18: Front projections of the Legendrian knot Λ of knot type $m(9_{46})$ (left) and the Legendrian unknot Λ_0 (right)

References

[1] **J A Baldwin**, **S Sivek**, *Invariants of Legendrian and transverse knots in monopole knot homology*, preprint (2014) arXiv To appear in J. Symplectic Geom.

[2] **F Bourgeois**, **B Chantraine**, *Bilinearized Legendrian contact homology and the augmentation category*, J. Symplectic Geom. 12 (2014) 553–583 MR

[3] **F Bourgeois**, **J M Sabloff**, **L Traynor**, *Lagrangian cobordisms via generating families: construction and geography*, Algebr. Geom. Topol. 15 (2015) 2439–2477 MR

[4] **B Chantraine**, *Lagrangian concordance of Legendrian knots*, Algebr. Geom. Topol. 10 (2010) 63–85 MR

[5] **B Chantraine**, *Lagrangian concordance is not a symmetric relation*, Quantum Topol. 6 (2015) 451–474 MR

[6] **B Chantraine**, **G Dimitroglou Rizell**, **P Ghiggini**, **R Golovko**, *Floer homology and Lagrangian concordance*, from "Proceedings of the Gökova Geometry–Topology Conference" (S Akbulut, D Auroux, T Önder, editors), GGT, Gökova (2015) 76–113 MR

[7] **Y Chekanov**, *Differential algebra of Legendrian links*, Invent. Math. 150 (2002) 441–483 MR

[8] **Y V Chekanov**, *Invariants of Legendrian knots*, from "Proceedings of the International Congress of Mathematicians, II" (T Li, editor), Higher Ed. Press, Beijing (2002) 385–394 MR

[9] **W Chongchitmate**, **L Ng**, *An atlas of Legendrian knots*, Exp. Math. 22 (2013) 26–37 MR

[10] **G Civan**, **P Koprowski**, **J Etnyre**, **J M Sabloff**, **A Walker**, *Product structures for Legendrian contact homology*, Math. Proc. Cambridge Philos. Soc. 150 (2011) 291–311 MR

[11] **C Cornwell**, **L Ng**, **S Sivek**, *Obstructions to Lagrangian concordance*, Algebr. Geom. Topol. 16 (2016) 797–824 MR

[12] **G Dimitroglou Rizell**, *Exact Lagrangian caps and non-uniruled Lagrangian submanifolds*, Ark. Mat. 53 (2015) 37–64 MR

[13] **G Dimitroglou Rizell**, *Lifting pseudo-holomorphic polygons to the symplectisation of $P \times \mathbb{R}$ and applications*, Quantum Topol. 7 (2016) 29–105 MR

[14] **T Ekholm**, *Rational symplectic field theory over $\mathbb{Z}_2$ for exact Lagrangian cobordisms*, J. Eur. Math. Soc. 10 (2008) 641–704 MR

[15] **T Ekholm**, *Rational SFT, linearized Legendrian contact homology, and Lagrangian Floer cohomology*, from "Perspectives in analysis, geometry, and topology" (I Itenberg, B Jöricke, M Passare, editors), Progr. Math. 296, Springer (2012) 109–145 MR

[16] **T Ekholm**, **J B Etnyre**, **J M Sabloff**, *A duality exact sequence for Legendrian contact homology*, Duke Math. J. 150 (2009) 1–75 MR

[17] **T Ekholm**, **J Etnyre**, **M Sullivan**, *The contact homology of Legendrian submanifolds in $\mathbb{R}^{2n+1}$*, J. Differential Geom. 71 (2005) 177–305 MR

[18] **T Ekholm**, **J Etnyre**, **M Sullivan**, *Orientations in Legendrian contact homology and exact Lagrangian immersions*, Internat. J. Math. 16 (2005) 453–532 MR

[19] **T Ekholm**, **K Honda**, **T Kálmán**, *Legendrian knots and exact Lagrangian cobordisms*, J. Eur. Math. Soc. 18 (2016) 2627–2689 MR

[20] **T Ekholm**, **T Kálmán**, *Isotopies of Legendrian 1–knots and Legendrian 2–tori*, J. Symplectic Geom. 6 (2008) 407–460 MR

[21] **Y Eliashberg**, *Invariants in contact topology*, Doc. Math. (1998) 327–338 MR

[22] **Y Eliashberg**, **A Givental**, **H Hofer**, *Introduction to symplectic field theory*, Geom. Funct. Anal. (2000) 560–673 MR

[23] **J B Etnyre**, **L L Ng**, **J M Sabloff**, *Invariants of Legendrian knots and coherent orientations*, J. Symplectic Geom. 1 (2002) 321–367 MR

[24] **A Floer**, *Morse theory for Lagrangian intersections*, J. Differential Geom. 28 (1988) 513–547 MR

[25] **E Getzler**, **J D S Jones**, *A_∞–algebras and the cyclic bar complex*, Illinois J. Math. 34 (1990) 256–283 MR

[26] **B Keller**, *Introduction to A–infinity algebras and modules*, Homology Homotopy Appl. 3 (2001) 1–35 MR

[27] **C Leverson**, *Augmentations and rulings of Legendrian knots*, J. Symplectic Geom. 14 (2016) 1089–1143 MR

[28] **M Menke**, *On the augmentation categories of positive braid closures*, preprint (2015) arXiv

[29] **D Nadler**, **E Zaslow**, *Constructible sheaves and the Fukaya category*, J. Amer. Math. Soc. 22 (2009) 233–286 MR

[30] **L L Ng**, *Computable Legendrian invariants*, Topology 42 (2003) 55–82 MR

[31] **L Ng**, **D Rutherford**, *Satellites of Legendrian knots and representations of the Chekanov–Eliashberg algebra*, Algebr. Geom. Topol. 13 (2013) 3047–3097 MR

[32] **L Ng**, **D Rutherford**, **V Shende**, **S Sivek**, *The cardinality of the augmentation category of a Legendrian link*, preprint (2015) arXiv To appear in Math. Res. Lett.

[33] **L Ng**, **D Rutherford**, **V Shende**, **S Sivek**, **E Zaslow**, *Augmentations are sheaves*, preprint (2015) arXiv

[34] **J M Sabloff**, **L Traynor**, *Obstructions to Lagrangian cobordisms between Legendrians via generating families*, Algebr. Geom. Topol. 13 (2013) 2733–2797 MR

[35] **P Seidel**, *Fukaya categories and Picard–Lefschetz theory*, Eur. Math. Soc., Zürich (2008) MR

[36] **J D Stasheff**, *Homotopy associativity of H–spaces, I*, Trans. Amer. Math. Soc. 108 (1963) 275–292 MR

[37] **J D Stasheff**, *Homotopy associativity of H–spaces, II*, Trans. Amer. Math. Soc. 108 (1963) 293–312 MR

Mathematics Department, Duke University
Durham, NC, United States

yu.pan@duke.edu

http://sites.google.com/site/yupanduke/

Received: 1 August 2016 Revised: 21 December 2016

Algebraic & Geometric Topology 17 (2017) 1871–1916

Tethers and homology stability for surfaces

ALLEN HATCHER
KAREN VOGTMANN

Homological stability for sequences $G_n \to G_{n+1} \to \cdots$ of groups is often proved by studying the spectral sequence associated to the action of G_n on a highly connected simplicial complex whose stabilizers are related to G_k for $k < n$. When G_n is the mapping class group of a manifold, suitable simplicial complexes can be made using isotopy classes of various geometric objects in the manifold. We focus on the case of surfaces and show that by using more refined geometric objects consisting of certain configurations of curves with arcs that tether these curves to the boundary, the stabilizers can be greatly simplified and consequently also the spectral sequence argument. We give a careful exposition of this program and its basic tools, then illustrate the method using braid groups before treating mapping class groups of orientable surfaces in full detail.

20J06, 57M07

Introduction

Many classical groups occur in sequences G_n with natural inclusions $G_n \to G_{n+1}$. Examples include the symmetric groups Σ_n, linear groups such as GL_n, the braid groups B_n, mapping class groups $M_{n,1}$ of surfaces with one boundary component, and automorphism groups of free groups $\mathrm{Aut}(F_n)$. A sequence of groups is said to be *homologically stable* if the natural inclusions induce isomorphisms on homology $H_i(G_n) \to H_i(G_{n+1})$ for n sufficiently large with respect to i. All of the sequences of groups mentioned above are homologically stable. This terminology is also slightly abused when there is no natural inclusion, such as for the mapping class groups M_n of closed surfaces and outer automorphism groups of free groups $\mathrm{Out}(F_n)$; in this case, we say the series is homologically stable if the i^{th} homology is independent of n for n sufficiently large with respect to i.

Homology stability is a very useful property. It sometimes allows one to deduce properties of the limit group $G_\infty = \lim_{\to} G_n$ from properties of the groups G_n; the classical example of this is Quillen's proof that various K–groups are finitely generated. It is also useful in the opposite direction: it is sometimes possible to compute invariants

Published: 17 July 2017 DOI: 10.2140/agt.2017.17.1871

of the limit group G_∞, which by stability are invariants of the groups G_n; an example of this is the computation by Madsen and Weiss of the stable homology of the mapping class group. Finally, there is the obvious advantage that homology computations which are unmanageable for n large can sometimes be done in G_n for n small.

In unpublished work from the 1970s, Quillen introduced a general method of proving stability theorems, which was used by many authors in subsequent years (the earliest examples include Wagoner [23], Vogtmann [22], Charney [5], van der Kallen [17] and Harer [9]). The idea is to find a highly connected complex X_n on which G_n acts such that stabilizers of simplices are isomorphic to G_m for $m < n$. One then examines a slight variant of the equivariant homology spectral sequence for this action; this has

$$E^1_{p,q} = \begin{cases} H_q(G_n, \mathbb{Z}) & \text{for } p = -1, \\ \bigoplus_{\sigma \in \Sigma_p} H_q(\mathrm{stab}(\sigma); \mathbb{Z}_\sigma) & \text{for } p \geq 0, \end{cases}$$

where Σ_p is a set of representatives of orbits of p–simplices. The fact that X_n is highly connected implies that this spectral sequence converges to 0 for $p+q$ small compared to n, and the fact that simplex stabilizers are smaller groups G_m means that the map $H_i(G_{n-1}) \to H_i(G_n)$ induced by inclusion occurs as a d^1 map in the spectral sequence. If one assumes the quotient X_n/G_n is highly connected and one or two small conditions of a more technical nature are satisfied, then an induction argument on i can be used to prove that this d^1 map is an isomorphism for n and i in the approximate range $n > 2i$.

This is the ideal situation, but in practice the original proofs of homology stability were often more complicated because the complexes X_n chosen had simplex stabilizers that were not exactly the groups G_m for $m < n$. For the groups $\mathrm{Aut}(F_n)$ and $\mathrm{Out}(F_n)$, a way to avoid the extra complications was developed by Hatcher, Vogtmann and Wahl [14; 15] with further refinements and extensions in Hatcher and Wahl [16]. The idea was to use variations of the original complexes studied by Hatcher and Vogtmann [12; 13] that included more data. In [14] this extra data consisted of supplementary 2–spheres in the ambient 3–manifold that were called "enveloping spheres", while in [15; 16] this extra data was reformulated in terms of arcs joining 2–spheres to basepoints in the boundary of the manifold. These arcs could be interpreted as "tethering" the spheres to the boundary.

In the present paper we show how this tethering idea can be used in the case of mapping class groups of surfaces. As above, tethers are arcs to a point in the boundary, while at their other end they attach either to individual curves in the surface or to ordered pairs of curves intersecting transversely in one point. We call such an ordered pair (a, b) a *chain* (see Figure 1) and use the term "curve" always to mean a simple closed curve.

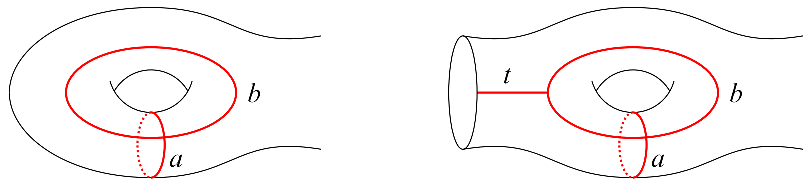

Figure 1: Chain on a closed surface and tethered chain on a surface with boundary

The classical curve complex $C(S)$ of a compact orientable surface S has vertices corresponding to isotopy classes of nontrivial curves in S (where nontrivial means neither bounding a disk nor isotopic to a boundary component of S) and a set of vertices spans a simplex if the curves can be chosen to be disjoint. Of particular interest is the subcomplex $C^0(S)$ formed by simplices corresponding to coconnected curve systems, that is, systems with connected complement.

In a similar way we can define a complex $\mathrm{Ch}(S)$ of chains, with simplices corresponding to isotopy classes of systems of disjoint chains; note that such a system is automatically coconnected and there are no trivial chains to exclude. When S has nonempty boundary we can also form complexes $\mathrm{TC}(S)$ and $\mathrm{TCh}(S)$ of systems of disjoint tethered curves or tethered chains. In the case of $\mathrm{TC}(S)$ we assume the curves without their tethers form coconnected curve systems, and in the case of $\mathrm{TCh}(S)$ we assume the tether to a chain attaches to the b–curve. A further variant that is useful for proving homology stability is a complex $\mathrm{DTC}(S)$ of *double-tethered* curves, by which we mean curves with two tethers attached at the same point of the curve but on opposite sides, and curve systems are again assumed to be coconnected. Note that shrinking the tether of a tethered chain to the point where it attaches to ∂S converts the b–curve to a double tether for the a–curve. More precise definitions for these complexes are given in Section 5, including extra data specifying where the tethers attach in ∂S.

Our main new result is:

Theorem *If S is a compact orientable surface of genus g then the complexes* $\mathrm{TC}(S)$, $\mathrm{DTC}(S)$, $\mathrm{Ch}(S)$ *and* $\mathrm{TCh}(S)$ *are all* $\frac{1}{2}(g-3)$*–connected.*

Recall that a space X is r–connected if $\pi_i(X)=0$ for $i\leq r$, which makes sense even if r is not an integer. Thus r–connected means the same as $\lfloor r\rfloor$–connected. In particular, (-1)–connected means nonempty (every map of $\partial D^0=\varnothing$ to X extends to a map of D^0 to X) and r–connected for $r<-1$ is an empty condition.

The mapping class group of S acts on all these complexes. A nice feature of the action on $\mathrm{TCh}(S)$ is that the stabilizer of a vertex is exactly the mapping class group for a surface with genus one less but the same number of boundary components. This is because cutting S along a tethered chain reduces the genus by one without changing the number of boundary components. Similarly, the stabilizer of a k–simplex is the mapping class group of a surface with genus reduced by $k+1$ and the same number of boundary components. This makes $\mathrm{TCh}(S)$ ideal for the spectral sequence argument proving homology stability with respect to increasing genus with a fixed positive number of boundary components. Actually it turns out to be slightly more efficient to use the complex $\mathrm{DTC}(S)$, or a variant of it where the double tethers attach to basepoints in two different components of ∂S and the ordering of the tethers at these basepoints satisfies a compatibility condition. With this complex a single spectral sequence suffices to prove both that the homology stabilizes with respect to genus (namely, H_i of the mapping class group is independent of g for $g \geq 2i+2$) and that the stable homology does not depend on the number of boundary components as long as this number is positive. In order to extend this to closed surfaces we need to work with a complex that does not involve tethers, and we use a version of $\mathrm{Ch}(S)$ in which chains are oriented and systems of oriented chains are ordered. (Even if one is interested only in closed surfaces it is necessary to consider the case of nonempty boundary in order to have a way to compare mapping class groups in different genus.)

The best stable dimension range that these simple sorts of spectral sequence arguments can yield has slope 2, as in the inequality $g \geq 2i+2$. This is not the optimal range, which has slope $\frac{3}{2}$, arising from more involved spectral sequence arguments. See Boldsen [3], Randal-Williams [20] and Wahl [24] for details.

The complexes of chains and tethered chains that we show are highly connected have found other recent applications as well in Putman and Sam [18] and Wahl and Randal-Williams [26]. In higher dimensions the natural analog of a tethered chain is a pair of k–spheres in a smooth manifold M^{2k} intersecting transversely in a single point, together with an arc tethering one of the spheres to a basepoint in ∂M. These tethered sphere-pairs play a central role in recent work of Galatius and Randal-Williams [8] on homology stability for $B\,\mathrm{Diff}(M)$ for certain $2k$–dimensional manifolds M with $k > 2$, including the base case that M is obtained from a connected sum of copies of $S^k \times S^k$ by deleting the interior of a $2k$–ball.

Here is an outline of the paper. In Section 1 we present the basic spectral sequence argument and in Section 2 we lay out the tools used to prove the key connectivity results. In Section 3 we give a warm-up example, illustrating the method in a particularly simple case, proving Arnol'd's homology stability theorem for braid groups. In Section 4 we give new, simpler proofs of results due to Harer about curve complexes and arc complexes that will be used in Section 5 to prove the main new connectivity statements. Finally in Section 6 we deduce homology stability for mapping class groups.

Remark A draft version of this paper dating from 2006 and treating several other classes of groups has been informally circulated for a number of years. This current version focuses only on mapping class groups, significantly simplifies several of the proofs in the earlier version, and also corrects a couple of errors. We thank Alexander Jasper for bringing one of these errors to our attention.

1 The basic spectral sequence argument

In this section we give the simplest form of the spectral sequence argument for proving homology stability of a sequence of group inclusions $\cdots \to G_n \to G_{n+1} \to G_{n+2} \to \cdots$. The input for the spectral sequence will be a simplicial action of G_n on a simplicial or semisimplicial complex X_n for each n. To deduce stability we will make the following assumptions, which are stronger than necessary for stability but simplify the arguments and are satisfied in all but one of our applications. The only exception arises in the proof of Theorem 6.2, where a short extra argument is required.

(1) X_n has dimension $n-1$ and the action of G_n is transitive on simplices of each dimension.

(2) The stabilizer of a vertex is conjugate to G_{n-1}, and more generally the stabilizer of a p–simplex is conjugate to G_{n-p-1}. Moreover, the stabilizer of a simplex fixes the simplex pointwise.

(3) If e is an edge of X_n with vertices v and w, then there is an element of G_n taking v to w which commutes with all elements of the stabilizer of e.

The dimension range in which homology stability holds will depend on the connectivity of X_n, which must grow linearly with n. The best result that the method can yield is that $H_i(G_{n-1}) \to H_i(G_n)$ is an isomorphism for $n > 2i + c$ and a surjection for $n = 2i + c$ for some constant c. In the cases which occur in this paper we have the following stable ranges:

Theorem 1.1 *Suppose the action of G_n on X_n satisfies conditions* (1)–(3) *for each n. Then:*

(a) *If X_n is $(n-3)$–connected for each n then the stabilization $H_i(G_{n-1}) \to H_i(G_n)$ is an isomorphism for $n > 2i+1$ and a surjection for $n = 2i+1$.*

(b) *If X_n is $\frac{1}{2}(n-3)$–connected for each n then the stabilization $H_i(G_{n-1}) \to H_i(G_n)$ is an isomorphism for $n > 2i+2$ and a surjection for $n = 2i+2$.*

Proof For $G = G_n$ let E_*G be a free resolution of $\mathbb{Z}$ by $\mathbb{Z}[G]$–modules, and let

$$\cdots \to C_p \to C_{p-1} \to \cdots \to C_0 \to C_{-1} = \mathbb{Z} \to 0$$

be the augmented simplicial chain complex of $X = X_n$. The action of G on X makes C_* into a complex of $\mathbb{Z}[G]$–modules, so we can take the tensor product over $\mathbb{Z}[G]$ to form a double complex $C_* \otimes_G E_*G$. Filtering this double complex horizontally and then vertically or vice versa gives rise to two spectral sequences, both converging to the same thing (see eg [4, VII.3]).

Using the horizontal filtration, the $E^1_{p,q}$ term of the associated spectral sequence is formed by taking the p^{th} homology of $C_* \otimes_G E_qG$. If we assume X is highly connected, say $c(X)$–connected, then the complex C_* is exact through dimension $c(X)$. Since E_qG is free, $C_* \otimes_G E_qG$ is exact in the same range, so $E^2_{p,q} = 0$ for $p \le c(X)$. In particular the spectral sequence converges to 0 in the range $p + q \le c(X)$, so the same will be true for the other spectral sequence as well.

For the second spectral sequence, if we begin by filtering vertically instead of horizontally the associated $E^1_{p,q}$ term becomes $H_q(G; C_p)$. By Shapiro's lemma (see eg [4, page 73]) this reduces to

$$E^1_{p,q} = \bigoplus_{\sigma \in \Sigma_p} H_q(\text{stab}(\sigma); \mathbb{Z}_\sigma),$$

where Σ_p is a set of representatives for orbits of p–simplices (if we consider a "(-1)–simplex" to be empty, with stabilizer all of G), and $\mathbb{Z}_\sigma$ is $\mathbb{Z}$ twisted by the orientation action of $\text{stab}(\sigma)$ on σ. In our case we assume the action is transitive on p–simplices so there is only one term in the direct sum. We also assume $\text{stab}(\sigma)$ is conjugate to G_{n-p-1} and fixes σ pointwise, so $\mathbb{Z}_\sigma$ is an untwisted $\mathbb{Z}$ and the E^1 terms become simply $E^1_{p,q} = H_q(G_{n-p-1})$ with untwisted $\mathbb{Z}$ coefficients understood.

Returning to the general case, the q^{th} row of the E^1 page is the augmented chain complex of the quotient X/G with coefficients in the system $\{H_q(\text{stab}(\sigma))\}$. The d^1–differentials in this chain complex can be described explicitly as follows. For a simplex $\sigma \in \Sigma_p$, the restriction of d^1 to the summand $H_q(\text{stab}(\sigma))$ will be the alternating sum of partial boundary maps $d^1_i\colon H_q(\text{stab}(\sigma)) \to H_q(\text{stab}(\tau))$, where $\tau \in \Sigma_{p-1}$ is the orbit representative of the i^{th} face $\partial_i\sigma$ and d^1_i is induced by the inclusion $\text{stab}(\sigma) \to \text{stab}(\partial_i\sigma)$ followed by the conjugation that takes this stabilizer to $\text{stab}(\tau)$.

Homology stability is proved by induction on the homology dimension i, starting with the trivial case $i = 0$. The sort of result we seek is that the stabilization map $H_i(G_{n-1}) \to H_i(G_n)$ is an isomorphism for $n > \varphi(i)$ and a surjection for $n = \varphi(i)$, for a linear function φ of positive slope.

The map $d = d^1\colon E^1_{0,i} \to E^1_{-1,i}$ in the second spectral sequence constructed above is the map on homology induced by the inclusion of a vertex stabilizer into the whole group;

by assumption this is the map $H_i(G_{n-1}) \to H_i(G_n)$ induced by the standard inclusion $G_{n-1} \to G_n$, so this is the map we are trying to prove is an isomorphism. In the situation we are considering the E^1 page of the spectral sequence has the following form:

$$
\begin{array}{r|ccccccccc}
i & H_i(G_n) & \xleftarrow{d} & H_i(G_{n-1}) & \longleftarrow & H_i(G_{n-2}) & \longleftarrow & \cdots & & \\
 & & & & & & & & & \\
i-1 & \cdots & \longleftarrow & H_{i-1}(G_{n-1}) & \longleftarrow & H_{i-1}(G_{n-2}) & \longleftarrow & H_{i-1}(G_{n-3}) & \longleftarrow & \cdots \\
 & & & & & & & & & \\
i-2 & & & & & \cdots & \longleftarrow & H_{i-2}(G_{n-3}) & \longleftarrow & \cdots \\
 & & & & & & & & & \\
 & & & & & & & & & \\
q=0 & H_0(G_n) & \longleftarrow & H_0(G_{n-1}) & \longleftarrow & H_0(G_{n-2}) & \longleftarrow & H_0(G_{n-3}) & \longleftarrow & \cdots \\
\hline
 & p=-1 & & 0 & & 1 & & 2 & & \cdots
\end{array}
$$

We first consider the argument for showing that d is surjective. If i is less than the connectivity $c(X)$ of X, then the terms $E^{\infty}_{p,q}$ must be zero for $p+q \le i-1$, and in particular $E^{\infty}_{-1,i}$ must be zero. We will show that every differential d^r with target $E^r_{-1,i}$ for $r > 1$ is the zero map because its domain is the zero group, so the only differential that can do the job of killing $E^*_{-1,i}$ is d, which must therefore be onto. Thus it will suffice to show that $E^2_{p,q} = 0$ for $p+q \le i$ and $q < i$. These groups are the reduced homology groups of X/G with coefficients in the system of groups $\{H_q(\operatorname{stab}(\sigma))\}$. We will argue that these coefficient groups can be replaced by $H_q(G)$, with a suitable induction hypothesis, so that $E^2_{p,q} = \widetilde{H}_p(X/G; H_q(G))$, still assuming $p+q \le i$ and $q < i$. Thus the groups $E^2_{p,q}$ with $p+q \le i$ and $q < i$ will be zero once we know that the connectivity $c(X/G)$ is large enough, namely $c(X/G) \ge p$. Since we have $p+q \le i$ and $q \ge 0$, the condition $c(X/G) \ge p$ can be reformulated as $c(X/G) \ge i$.

As explained earlier, the d^1 differentials are built from maps induced by inclusion followed by conjugation. These maps fit into commutative diagrams

$$
\begin{tikzcd}
H_q(\operatorname{stab}(\sigma)) \arrow[r] \arrow[d] & H_q(\operatorname{stab}(\tau)) \arrow[d] \\
H_q(G) \arrow[r] & H_q(G)
\end{tikzcd}
$$

where the vertical maps are induced by inclusion and the lower map is induced by conjugation in G, hence is the identity. If the vertical maps are isomorphisms, we can then replace the coefficient groups in the q^{th} row of the E^1 page by the constant groups $H_q(G)$. In our case with E^1 page displayed above, we would like the group $H_{i-1}(G_{n-3})$ and the groups to the left of it to be in the stable range, isomorphic to $H_{i-1}(G)$. Actually we can get by with slightly less, just having $H_{i-1}(G_{n-2})$ and the terms to the left of it isomorphic to the stable group and having $H_{i-1}(G_{n-3})$ mapping onto the stable group, since this is enough to guarantee that the homology of the chain complex at $H_{i-1}(G_{n-2})$ will be zero. Thus we want the relation $\varphi(i) \geq \varphi(i-1)+2$. The corresponding relation for smaller values of i will take care of lower rows, by the same argument.

To summarize, we have shown that the stabilization $H_i(G_{n-1}) \to H_i(G_n)$ will be surjective if $\varphi(i) \geq \varphi(i-1)+2$, assuming $i-1 \leq c(X_n)$ and $i \leq c(X_n/G_n)$.

To prove that d is injective the argument is similar, but with one extra step. If $i \leq c(X_n)$ the term $E^\infty_{0,i}$ will be zero, and then it will suffice to show that all differentials with target $E^r_{0,i}$ are zero, so the only way for $E^1_{0,i}$ to die is if d is injective. We can argue that the terms $E^2_{p,q}$ are zero for $p+q \leq i+1$ and $q < i$ just as before, assuming again that $\varphi(i) \geq \varphi(i-1)+2$ but with the inequality $i \leq c(X_n/G_n)$ replaced by $i+1 \leq c(X_n/G_n)$ since we have shifted one unit to the right in the spectral sequence. The extra step we need for injectivity of d is showing that the differential $d^1\colon E^1_{1,i} \to E^1_{0,i}$ is zero. This will follow from the assumption that for each edge e of X_n there is an element g of G_n taking one of the endpoints v of e to the other endpoint w such that g commutes with $\mathrm{stab}(e)$. This guarantees that d^1 vanishes on the summand of $E^1_{1,i}$ corresponding to e by our earlier description of d^1. Namely, conjugation by g fixes $\mathrm{stab}(e)$ and sends $\mathrm{stab}(v)$ to $\mathrm{stab}(w)$. If v_0 is the vertex chosen to represent the vertex orbit and if $h_v v = v_0$ and $h_w w = v_0$, then the identifications of $\mathrm{stab}(v)$ and $\mathrm{stab}(w)$ with $\mathrm{stab}(v_0)$ differ by conjugation by $h_w g h_v^{-1}$ so we have the following commutative diagram:

$$\begin{tikzcd}
 & \mathrm{stab}(v) \arrow[r, "c_{h_v}"] \arrow[d, "c_g"] & \mathrm{stab}(v_0) \arrow[d, "c_{h_w g h_v^{-1}}"] \\
\mathrm{stab}(e) \arrow[ur, "i"] \arrow[dr, "i"'] & \mathrm{stab}(w) \arrow[r, "c_{h_w}"] & \mathrm{stab}(v_0)
\end{tikzcd}$$

Since $h_w g h_v^{-1}$ is an element of $\mathrm{stab}(v_0)$, conjugation by it induces the identity on $H_*(\mathrm{stab}(v_0))$ and the previous diagram induces a commutative diagram:

$$\begin{tikzcd}
 & H_*(\mathrm{stab}(v)) \arrow[dd] \arrow[dr] & \\
H_*(\mathrm{stab}(e)) \arrow[ur] \arrow[dr] & & H_*(\mathrm{stab}(v_0)) \\
 & H_*(\mathrm{stab}(w)) \arrow[ur] &
\end{tikzcd}$$

Thus we see that the stabilization $H_i(G_{n-1}) \to H_i(G_n)$ will be injective whenever $\varphi(i) \geq \varphi(i-1)+2$, assuming $i \leq c(X_n)$ and $i+1 \leq c(X_n/G_n)$.

The connectivity of the quotient X_n/G_n is not hard to compute. Since we assume the action of G_n is transitive on simplices of each dimension and the stabilizer of a simplex fixes it pointwise, X_n/G_n can be identified with the quotient of the standard simplex Δ^{n-1} obtained by identifying all of its k–dimensional faces for each k, where the identification preserves the ordering of the vertices. Thus X_n/G_n is a semisimplicial complex (or Δ–complex) with one k–simplex for each $k \leq n-1$. It is easy to see that X_n/G_n is simply connected. Its simplicial chain complex has a copy of $\mathbb{Z}$ in each dimension $k \leq n-1$ with boundary maps that are alternately zero and isomorphisms. Therefore the reduced homology groups of X_n/G_n are trivial below dimension $n-1$, while $H_{n-1}(X_n/G_n)$ is trivial when n is odd and $\mathbb{Z}$ when n is even. Thus X_n/G_n is $(n-2)$–connected

The condition $\varphi(i) \geq \varphi(i-1)+2$ is satisfied if we choose $\varphi(i) = 2i + c$ for any constant c. It remains to determine c.

To get surjectivity from the spectral sequence argument we need $c(X_n/G_n) \geq i$ and $c(X_n) \geq i-1$. For injectivity we need one more degree of connectivity for each. Consider first the inequalities involving $c(X_n/G_n)$. We know $c(X_n/G_n) = n-2$ so we need $n \geq i+2$ for surjectivity and $n \geq i+3$ for injectivity. We want surjectivity for $n \geq \varphi(i) = 2i + c$ for all $i \geq 1$ (and injectivity for $n \geq 2i+c+1$), so any $c \geq 1$ works.

There remain the conditions $c(X_n) \geq i-1$ for surjectivity and $c(X_n) \geq i$ for injectivity. In case (a) we have $c(X_n) = n-3$ giving the same $n \geq i+2$ for surjectivity and $n \geq i+3$ for injectivity as before, so $\varphi(i) = 2i+1$ still works. For case (b) we have $c(X_n) = \frac{1}{2}(n-3)$, so we need $n \geq 2i+1$ for surjectivity and $n \geq 2i+3$ for injectivity; in particular, taking $\varphi(i) = 2i+2$ works for both. □

2 Connectivity tools

All of the complexes we will consider are of a certain type, which we shall call, somewhat informally, *geometric complexes*. Such a complex X is a simplicial complex whose vertices correspond to isotopy classes of some type of nontrivial geometric object (for example arcs or curves in a surface, or combinations thereof), where *trivial* has different meanings in different contexts. A collection of vertices $v_0, \dots, v_k$ spans a k–simplex if representatives for the vertices can be chosen which are pairwise disjoint, and perhaps also satisfy some auxiliary conditions. The corresponding set of isotopy classes defining the simplex of X is also called a *system*, and the set of systems forms a partially ordered set (poset) $\widehat{X}$ under inclusion, whose geometric realization is the barycentric subdivision X' of X.

In this section we lay out a few general tools we will use for proving that various geometric complexes are highly connected.

2.1 Link arguments: rerouting disks to avoid bad simplices

We would like to relate n–connectedness of a simplicial complex X to n–connectedness of a subcomplex Y. We do this by finding conditions under which the relative homotopy groups $\pi_i(X, Y)$ are zero in some range $i \le n$, so that the desired connectivity statements can be deduced from the long exact sequence of homotopy groups for (X, Y). Thus we wish to deform a map $f\colon (D^i, \partial D^i) \to (X, Y)$ to have image in Y, staying fixed on ∂D^i. We may assume f is simplicial with respect to some triangulation of D^i, and then the idea is to deform f by performing a sequence of local alterations in the open star of one simplex at a time, until f is finally pushed into Y. This method of improving the map is called a *link argument*. We write $\mathrm{lk}(\sigma)$ for the link of a simplex σ and $\mathrm{st}(\sigma)$ for the star; if the ambient complex X needs to be specified we write $\mathrm{lk}_X(\sigma)$ and $\mathrm{st}_X(\sigma)$.

We first identify a set of simplices in $X - Y$ as *bad* simplices, satisfying the following two conditions:

(1) Any simplex with no bad faces is in Y, where by a "face" of a simplex we mean a subsimplex spanned by any nonempty subset of its vertices, proper or not.

(2) If two faces of a simplex are both bad, then their join is also bad.

We call simplices with no bad faces *good* simplices. Bad simplices may have good faces, or faces which are neither good nor bad. If σ is a bad simplex we say a simplex τ in $\mathrm{lk}(\sigma)$ is *good for* σ if any bad face of $\tau * \sigma$ is contained in σ. The simplices which are good for σ form a subcomplex of $\mathrm{lk}(\sigma)$, which we denote by G_σ.

Proposition 2.1 *Let X, Y and G_σ be as above. Suppose that for some integer $n \ge 0$ the subcomplex G_σ of X is $(n-\dim(\sigma)-1)$–connected for all bad simplices σ. Then the pair (X, Y) is n–connected, ie $\pi_i(X, Y) = 0$ for all $i \le n$.*

Proof We will show how to deform a map $f\colon (D^i, \partial D^i) \to (X, Y)$ to have image in Y, staying fixed on ∂D^i, provided that $i \le n$. We may assume f is simplicial with respect to some triangulation of D^i. Let μ be a maximal simplex of D^i such that $\sigma = f(\mu)$ is bad (so in particular μ is not contained in ∂D^i). Then $f(\mathrm{lk}(\mu)) \subset \mathrm{lk}(\sigma)$ is contained in G_σ, since otherwise there is some $\nu \in \mathrm{lk}(\mu)$ and face σ_0 of σ such that $f(\nu) * \sigma_0$ is bad, in which case, by property (2), $(f(\nu) * \sigma_0) * \sigma = f(\nu) * \sigma = f(\nu * \mu)$ is bad, contradicting maximality of μ.

We can assume the triangulation of D^i gives the standard PL structure on D^i, so $\mathrm{lk}(\mu)$ is homeomorphic to S^{i-k-1}, where $k = \dim(\mu) \ge \dim(\sigma)$. Since G_σ is $(n-k-1)$–connected and $i \le n$, the restriction of f to $\mathrm{lk}(\mu)$ can be extended to $g\colon D^{i-k} \to G_\sigma$,

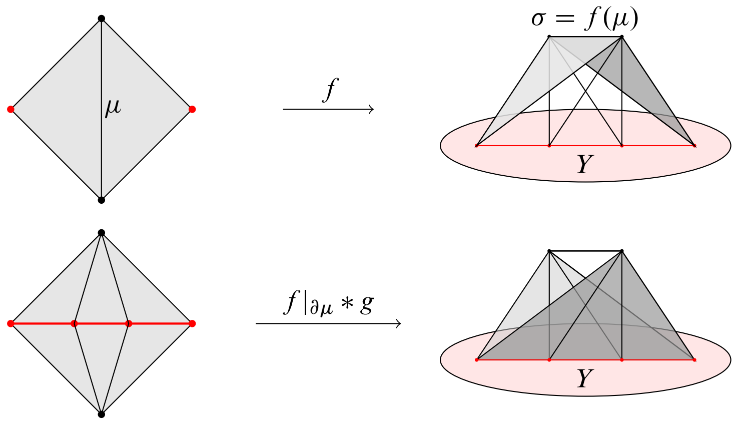

Figure 2: Retriangulation of $\mathrm{st}(\mu)$ and new definition of f

which we may take to be simplicial for some triangulation of D^{i-k} extending the given triangulation on $S^{i-k-1} = \mathrm{lk}(\mu)$. We retriangulate $\mathrm{st}(\mu)$ as $\partial\mu * D^{i-k}$ and redefine f on this new triangulation to be $f|_{\partial\mu} * g$ (see Figure 2).

The new map is homotopic to the old map, and agrees with the old map outside the interior of $\mathrm{st}(\mu)$, in particular on ∂D^i. Since simplices in G_σ are good for σ, no simplices in the interior of $\mathrm{st}(\mu)$ have bad images. Since the original triangulation of D^i was finite, in this way we can eventually eliminate all k–simplices of D^i with bad images without introducing simplices of higher dimension with bad images. Repeating the process in the new triangulation of D^i for simplices of dimension $k-1$, then $k-2$, etc, we eventually eliminate all bad simplices from the image, so that by property (1) the image lies in Y. $\square$

We give two applications of this proposition which we will use in the rest of the paper.

Corollary 2.2 *Let Y be a subcomplex of a simplicial complex X, and suppose $X-Y$ has a set of bad simplices satisfying* (1) *and* (2) *above. Then:*

(a) *If X is n–connected and G_σ is $(n-\dim(\sigma))$–connected for all bad simplices σ, then Y is n–connected.*

(b) *If Y is n–connected and G_σ is $(n-\dim(\sigma)-1)$–connected for all bad simplices σ, then X is n–connected.*

Proof Both statements follow from Proposition 2.1 using the long exact sequence of homotopy groups for (X,Y). This is immediate for (b), while for (a) one should replace the n in the proposition by $n+1$. $\square$

Given any simplicial complex X and a set of labels S, we can form a new simplicial complex X^S whose simplices consist of the simplices of X with vertices labeled by elements of S. Thus there are $|S|^{k+1}$ k–simplices of X^S for each k–simplex of X.

Corollary 2.3 *Let X be a simplicial complex and S a set of labels. If X is n–connected and the link of each k–simplex in X is $(n-k-1)$–connected, then X^S is n–connected. In the other direction, if X^S is n–connected then so is X, without any condition on the links.*

Proof Choosing a label $s_0 \in S$, we can regard X as the subcomplex of X^S consisting of simplices with all labels equal to s_0. There is then a retraction $r\colon X^S \to X$ which changes all labels to s_0. This implies the second statement of the corollary. For the first statement we will apply Corollary 2.2(b). Call a simplex of X^S *bad* if none of its vertex labels is equal to s_0. It is immediate that the set of bad simplices satisfies (1) and (2). If σ is a bad simplex, then a simplex in $\mathrm{lk}(\sigma)$ is good for σ if and only if all of its labels are s_0, so that G_σ is isomorphic to $\mathrm{lk}_X(r(\sigma))$ and Corollary 2.2(b) applies. □

Example 2.4 If X is the p–simplex Δ^p, one might think the lemma could be applied for all n to conclude that $(\Delta^p)^S$ was contractible. However, it can only be applied for $n \le p-1$, since for $n = p$ the hypothesis would say that the link of the whole simplex is (-1)–connected, ie nonempty, which is not the case. In fact $(\Delta^p)^S$ is the join of $p+1$ copies of the discrete set S, so it is p–dimensional and exactly $(p-1)$–connected if S has more than one element.

2.2 Homotopy equivalence of posets

The *geometric realization* of a poset P is the simplicial complex with one k–simplex for each totally ordered chain $p_0 < \cdots < p_k$ of $k+1$ elements $p_i \in P$. An order-preserving map (poset map) between posets induces a simplicial map of their geometric realizations. When we attribute some topological property to a poset or poset map we mean that the corresponding space or simplicial map has that property.

For a poset map $\phi\colon P \to Q$ the fiber $\phi_{\le q}$ over an element $q \in Q$ is defined to be the subposet of P consisting of all $p \in P$ with $\phi(p) \le q$. The fiber $\phi_{\ge q}$ is defined analogously. The following statement is known as Quillen's fiber lemma and is a special case of his Theorem A [19]. We supply an elementary proof.

Proposition 2.5 *A poset map $\phi\colon P \to Q$ is a homotopy equivalence if all fibers $\phi_{\le q}$ are contractible, or if all fibers $\phi_{\ge q}$ are contractible.*

Proof There is no difference between the two cases, so let us assume the fibers $\phi_{\geq q}$ are contractible. We construct a map $\psi\colon Q \to P$ inductively over the skeleta of Q as follows. For a vertex q_0 we let $\psi(q_0)$ be any vertex in $\phi_{\geq q_0}$, which is nonempty since it is contractible. For an edge $q_0 < q_1$ both $\psi(q_0)$ and $\psi(q_1)$ then lie in $\phi_{\geq q_0}$ and we let ψ map this edge to any path in $\phi_{\geq q_0}$ from $\psi(q_0)$ to $\psi(q_1)$. Extending ψ over higher simplices $q_0 < \cdots < q_k$ is done similarly, mapping them to $\phi_{\geq q_0}$ extending the previously constructed map on the boundary of the simplex.

We claim that ψ is a homotopy inverse to ϕ. The composition $\phi\psi$ sends each simplex $q_0 < \cdots < q_k$ to the subcomplex $Q_{\geq q_0}$. These subcomplexes are contractible, having minimum elements, so one can construct a homotopy from $\phi\psi$ to the identity inductively over skeleta of Q. Similarly $\psi\phi$ is homotopic to the identity since it sends each simplex $p_0 < \cdots < p_k$ to the contractible subcomplex $\phi_{\geq \phi(p_0)}$. □

We will often apply this proposition to the poset $\widehat{X}$ of simplices in some simplicial complex X. The geometric realization of this poset is the barycentric subdivision X' of the complex. The following lemma characterizes the poset fibers:

Lemma 2.6 *Let $f\colon X \to Y$ be a simplicial map of simplicial complexes, $\widehat{X}$ the poset of simplices in X, $\widehat{Y}$ the poset of simplices in Y and $\widehat{f}\colon \widehat{X} \to \widehat{Y}$ the induced poset map. Then for each simplex σ of Y we have the following relationships:*

(i) *$\widehat{f}_{\leq\sigma}$ is homeomorphic to $f^{-1}(\sigma)$.*

(ii) *$\widehat{f}_{\geq\sigma}$ is homotopy equivalent to $\widehat{f}^{-1}(\sigma)$.*

(iii) *$\widehat{f}^{-1}(\sigma)$ is homeomorphic to $f^{-1}(y)$, where y is the barycenter of σ.*

Proof Statement (i) is immediate from the definitions: $\widehat{f}_{\leq\sigma}$ is the set of all simplices τ such that $f(\tau)$ is a face of σ.

On the other hand, $\widehat{f}_{\geq\sigma}$ is the set of all simplices τ such that $f(\tau)$ has σ as a face. Since f is a simplicial map, some face of τ maps to σ; let τ_σ be the (unique) maximal such face. The map $\tau \mapsto \tau_\sigma$ is a poset map $\widehat{f}_{\geq\sigma} \to \widehat{f}^{-1}(\sigma)$ whose upper fibers are contractible, having unique minimal elements. Thus $\widehat{f}_{\geq\sigma}$ is homotopy equivalent to $\widehat{f}^{-1}(\sigma)$, giving statement (ii). Part (iii) is clear from the definitions. □

The following is an immediate consequence of Proposition 2.5 and Lemma 2.6:

Corollary 2.7 *Let $f\colon X \to Y$ be a simplicial map of simplicial complexes. If $f^{-1}(\sigma)$ is contractible for all simplices σ or if $f^{-1}(y)$ is contractible for all barycenters y, then f is a homotopy equivalence.*

Remark For a simplicial map, contractibility of the fibers over barycenters implies contractibility of all fibers since the fibers over an open simplex are all homeomorphic. Other types of maps for which contractibility of fibers implies homotopy equivalence or at least weak homotopy equivalence include fibrations, quasifibrations, and microfibrations (see [27] for the last case). The corollary implies that simplicial maps with contractible fibers are quasifibrations, but they need not be fibrations or microfibrations, as shown by the simple example of vertical projection of the letter L onto its base segment.

2.3 Fiber connectivity

Lemma 2.8 *Let $f\colon X \to Y$ be a simplicial map of simplicial complexes. Suppose that Y is n–connected and the fibers $f^{-1}(y)$ over the barycenters y of all k–simplices in Y are $(n{-}k)$–connected. Then X is n–connected.*

Proof Given a map $g\colon S^i \to X$, which we can assume is simplicial, we want to extend this to a map $G\colon D^{i+1} \to X$ if $i \le n$. In order to do this, we first consider the composition $h = fg\colon S^i \to Y$. Since Y is n–connected, we can extend h to a simplicial map $H\colon D^{i+1} \to Y$. We will use H to construct G, which we will do inductively on the skeleta of the barycentric subdivision D' of D^{i+1}.

We begin by replacing all complexes and maps by the associated posets of simplices and poset maps:

$$\begin{array}{ccc} \widehat{S}^i & \xrightarrow{\widehat{g}} & \widehat{X} \\ \downarrow & & \downarrow \widehat{f} \\ \widehat{D}^{i+1} & \xrightarrow{\widehat{H}} & \widehat{Y} \end{array}$$

Let τ be a vertex of D', so τ can be viewed as a simplex of D^{i+1} or as an element of $\widehat{D}^{i+1}$. Since H is simplicial, $\sigma = H(\tau)$ has dimension at most $i+1 \le n+1$ in Y. By the hypothesis and Lemma 2.6, $\widehat{f}_{\ge\sigma}$ is at least (-1)–connected, ie it is nonempty, so choose $x \in \widehat{f}_{\ge\sigma}$ and set $G(\tau) = x$. We can assume this agrees with the given g for $\tau \in \partial D'$.

Now assume we have defined G on the $(k{-}1)$–skeleton of D', and let $\tau_0 < \cdots < \tau_k$ be a k–simplex of D'. Let $\sigma_i = H(\tau_i)$, and note that $\widehat{f}_{\ge\sigma_j} \subset \widehat{f}_{\ge\sigma_0}$ for all j. By construction, then, G maps the boundary of the simplex to $\widehat{f}_{\ge\sigma_0}$. Since H is a simplicial map, it can only decrease the dimension of a simplex, so $\dim(\sigma_0) \le i+1-k \le n+1-k$, and consequently $\widehat{f}_{\ge\sigma_0}$ is at least $(k{-}1)$–connected. Therefore we can extend G over the interior of the k–simplex $\tau_0 < \cdots < \tau_k$, agreeing with the given G on $\partial D'$. This gives the induction step in the construction of G. □

2.4 Flowing into a subcomplex

In this section we abstract the essential features of a surgery technique from [11] for showing that certain complexes of arcs on a surface are contractible, in order to more conveniently apply the method to several different situations later in the paper.

Let Y be a subcomplex of a simplicial complex X. If $F\colon X \times I \to X$ is a deformation retraction into Y then each $x \in X$ gives a path $F(x,t)$ for $0 \le t \le 1$ starting at x and ending in Y. In nice cases these paths fit together to give a flow on the complement of Y. What we want to do is to work backwards, constructing a deformation retraction by first constructing a set of flow lines. Our flow lines will intersect each open simplex of X which is not contained in Y either transversely or in a family of parallel line segments. To specify these line segments, for each simplex $\sigma \in X - Y$ we choose a preferred vertex $v = v_\sigma$ and a simplex Δv in the link of v in X such that $\sigma * \Delta v$ is a simplex of X; then $\sigma * \Delta v$ is foliated by line segments parallel to the line from v_σ to the barycenter of Δv_σ (see Figure 3). To show that the flow ends up in Y we measure progress by means of a *complexity function* assigning a nonnegative integer to each vertex of X and taking strictly positive values on vertices not in Y. This can be extended to be defined for all simplices of X, where the complexity of a simplex is the sum of the complexities of its vertices.

Lemma 2.9 *Let Y be a subcomplex of a simplicial complex X with a complexity function c as above. Suppose that for each vertex $v \in X - Y$ we have a rule for associating a simplex Δv in the link of v in X, and for each simplex σ of X not contained in Y we have a rule for picking one of its vertices $v_\sigma \in X - Y$ so that*

(i) *the join $\sigma * \Delta v_\sigma$ is a simplex of X,*

(ii) $c(\Delta v) < c(v)$,

(iii) *if τ is a face of σ which contains v_σ, then $v_\tau = v_\sigma$.*

Then Y is a deformation retract of X.

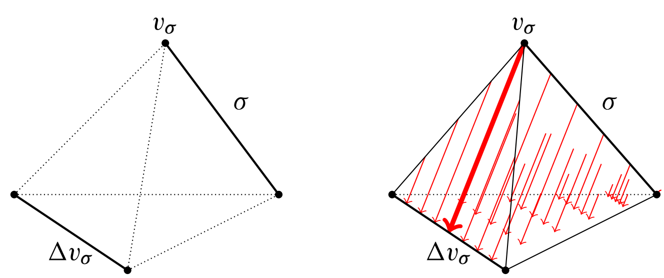

Figure 3: A simplex σ, its preferred vertex v_σ and flow lines in $\Delta v_\sigma * \sigma$

Proof For each simplex σ not contained in Y we construct flow lines in the simplex $\sigma * \Delta v_\sigma$ as described above, starting at $x \in \sigma$ and running parallel to the line from v_σ to the barycenter of Δv_σ. In terms of barycentric coordinates in the simplex $\sigma * \Delta v_\sigma$, viewed as weights on its vertices, we are shifting the weight on v_σ to equally distributed weights on the vertices of Δv_σ, keeping the weights of other vertices fixed. When we follow the resulting flow on $\sigma * \Delta v_\sigma$, all points that actually move end up with smaller complexity by condition (ii). Thus after a finite number of such flows across simplices, each point of σ follows a polygonal path ending in Y. Condition (iii) guarantees that the resulting flow is continuous on X, where we fix a standard Euclidean metric on each simplex and let each point flow at constant speed so as to reach Y at time 1. □

Surgery flows In this paper we will use Lemma 2.9 on various complexes of arcs and curves on surfaces. The complexity function will count the number of nontrivial intersection points with a fixed arc, curve or set of curves, and the simplex Δv will be obtained using the surgery technique from [11] to decrease the number of intersection points. The vertex v_σ will be an "innermost" or "outermost" arc or curve of σ, depending on the situation. In order for this surgery process to be well-defined we must first put each arc or curve system $\{t_0, \dots, t_k\}$ into normal form with respect to some fixed arc, curve, or curve system t, so that each t_i has minimal possible intersection with t in its isotopy class. In all cases we consider these normal forms are easily shown to exist; furthermore they are unique up to isotopy through normal forms, apart from the special situation that one t_i is isotopic to t, in which case this t_i can be isotoped across t from one side to the other without always being in normal form during the isotopy.

2.5 Ordered complexes

In this subsection we prove a proposition that will be used at the very end of the paper, when we extend the proof of homology stability for mapping class groups of nonclosed surfaces to the case of closed surfaces. We remark that this extension can also be proved without using the proposition, at the expense of complicating the spectral sequence argument and introducing an infinite-dimensional auxiliary complex. This was the method used in three earlier papers in analogous situations: [12, page 53; 14, end of Section 6; 16, proof of Theorem 5.1].

For a simplicial complex X let $\langle X\rangle$ be the ordered version of X, the semisimplicial complex whose k–simplices are the k–simplices of X with orderings of their vertices. Thus there are $(k+1)!$ k–simplices of $\langle X\rangle$ for each k–simplex of X. For example if X is a 1–simplex then $\langle X\rangle$ has two vertices connected by two distinct edges.

Forgetting orderings gives a natural projection $\langle X\rangle \to X$. This has cross-sections obtained by choosing an ordering of all the vertices of X and using this to order the vertices of each simplex of X. Thus X is a retract of $\langle X\rangle$, so high connectivity of $\langle X\rangle$ implies the same high connectivity for X. We will be interested in the converse question of when high connectivity of X implies high connectivity of $\langle X\rangle$. It is clear that $\langle X\rangle$ is connected if X is, but the example of a 1–simplex shows that this does not extend to 1–connectedness. We therefore need some conditions on X, conditions that will be satisfied in our application.

Generalizing a property of spheres with PL triangulations, a simplicial complex of dimension n is called *Cohen–Macaulay* if it is $(n-1)$–connected and the link of each of its k–simplices is $(n-k-2)$–connected. If we drop the condition that n is the dimension of the complex and only require it to be $(n-1)$–connected with the link of each k–simplex $(n-k-2)$–connected, then we have the notion of *weakly Cohen–Macaulay (wCM) of level n*. In the existing literature (eg [16]), the term "level" is replaced by "dimension", although this may be misleading since there is no restriction on the actual dimension.

A simple observation is that a complex X is wCM of level n if and only its n–skeleton is wCM of level n. This is because the homotopy groups $\pi_i(X)$ for $i\le n-1$ depend only on the n–skeleton, and a similar statement holds also for links in X. Thus X being wCM of level n is equivalent to its n–skeleton being Cohen–Macaulay (of dimension n).

Note that a complex which is wCM of level n is automatically wCM of level m for each $m<n$. There is no need to require n to be an integer, but if it is not, then wCM of level n is the same as wCM of level $\lfloor n\rfloor$, the greatest integer $\le n$, so allowing n to be nonintegral is just a matter of convenience.

Another observation is that if X is wCM of level n and σ is a k–simplex in X then the link of σ is wCM of level $n-k-1$. This is because for τ an l–simplex in $\mathrm{lk}(\sigma)$ we have $\mathrm{lk}_{\mathrm{lk}(\sigma)}(\tau)=\mathrm{lk}_X(\sigma*\tau)$ with $\sigma*\tau$ having dimension $k+l+1$, so $\mathrm{lk}_X(\sigma*\tau)$ has connectivity $n-(k+l+1)-2=(n-k-1)-l-2$.

Proposition 2.10 *If a simplicial complex X is wCM of level n then the ordered complex $\langle X\rangle$ is $(n-1)$–connected.*

In addition to the proof given below we give a different proof in the appendix, following instead the approach in Proposition 2.14 of [26]. We include both proofs since they are of similar length and each has its own virtues.

Proof As notation, we will use lowercase Greek letters for simplices of X, while simplices in $\langle X\rangle$ will be written as the ordered string of their vertices, eg $x_0x_1\dots x_k$, sometimes abbreviated to $\boldsymbol{x}=x_0x_1\dots x_k$.

By induction on n it suffices to show that $\pi_{n-1}(\langle X\rangle)=0$. Given a map $f\colon \partial D^n \to \langle X\rangle$, compose it with the projection $\langle X\rangle \to X$ to get a map $\partial D^n \to X$. Since X is wCM of level n, this can be extended to a map $D^n \to X$. Composing this extension with a section gives a map $D^n \to \langle X\rangle$, whose restriction $g\colon \partial D^n \to \langle X\rangle$ has the same projection as f. Since g is homotopically trivial, it suffices to show f is homotopic to g.

In order to construct a homotopy we cover X by the stars $\mathrm{st}_X(\sigma)$ of its simplices and consider the corresponding cover of $\langle X\rangle$ by the ordered complexes $\langle \mathrm{st}_X(\sigma)\rangle$. We first show that each of these is $(n-1)$–connected, and then use this fact to build the homotopy between f and g.

Claim *For each k–simplex σ in X, $\langle \mathrm{st}_X(\sigma)\rangle$ is $(n-1)$–connected.*

Proof We may assume X has dimension n since higher-dimensional simplices have no effect on the relevant connectivities, as noted earlier. Choose a vertex $a\in\sigma$. Then $\mathrm{st}_X(\sigma)$ is the cone $a*Y$ on $Y=\tau*\mathrm{lk}_X(\sigma)$, where τ is the face of σ opposite a. Since τ is wCM of level $k-1$ and $\mathrm{lk}_X(\sigma)$ is wCM of level $n-k-1$ as noted earlier, it follows that Y is wCM of level $n-1$. By induction the proposition is therefore true for Y (and all of its links).

Filter $\langle \mathrm{st}_X(\sigma)\rangle$ by subcomplexes Δ_i, where Δ_i is the union of all ordered n–simplices $x_0\dots x_n$ in $\langle \mathrm{st}_X(\sigma)\rangle$ with $a=x_j$ for some $j\le i$. The lower-dimensional simplices of Δ_i thus have the form $x_0\dots x_m$ with either $x_j=a$ for some $j\le i$ or no $x_j=a$. We will show that each Δ_i is $(n-1)$–connected by induction on i. The first subcomplex Δ_0 is the union of all ordered simplices of the form $ay_1\dots y_n$ with $y_1\dots y_n\in\langle Y\rangle$, ie it is the cone $a*\langle Y\rangle$ so is contractible. Each Δ_i for $i>0$ is obtained from Δ_{i-1} by attaching all n–simplices of $\langle \mathrm{st}_X(\sigma)\rangle$ of the form $x_0\dots x_{i-1}ay_{i+1}\dots y_n$. If we fix the ordered simplex $\boldsymbol{x}=x_0\dots x_{i-1}$ and let the y_j vary we obtain a subcomplex $\Delta_i(\boldsymbol{x})$ of Δ_i, ie $\Delta_i(\boldsymbol{x})$ is the union of the ordered n–simplices of $\langle \mathrm{st}_X(\sigma)\rangle$ starting with $\boldsymbol{x}a$.

The subcomplex $\Delta_i(\boldsymbol{x})$ decomposes as the join $\boldsymbol{x}a*\langle \mathrm{lk}_Y(\eta)\rangle$, where η is the (unordered) projection of $\boldsymbol{x}$ to X. In particular $\Delta_i(\boldsymbol{x})$ is contractible since $\boldsymbol{x}a$ is contractible. The intersection of $\Delta_i(\boldsymbol{x})$ with Δ_{i-1} is $\partial(\boldsymbol{x}a)*\langle \mathrm{lk}_Y(\eta)\rangle$ since the only way a face of a simplex $x_0\dots x_{i-1}ay_{i+1}\dots y_n$ can lie in Δ_{i-1} is if at least one of the vertices $x_0,\dots,x_{i-1},a$ is deleted. Since $\partial(\boldsymbol{x}a)$ is an $(i-1)$–sphere, $\partial(\boldsymbol{x}a)*\langle \mathrm{lk}_Y(\eta)\rangle$ is the i–fold suspension of $\langle \mathrm{lk}_Y(\eta)\rangle$ (since join with S^0 is suspension and join is associative) so the connectivity of $\partial(\boldsymbol{x}a)*\langle \mathrm{lk}_Y(\eta)\rangle$ is i more than the connectivity of $\langle \mathrm{lk}_Y(\eta)\rangle$. By induction $\langle \mathrm{lk}_Y(\eta)\rangle$ is $((n-i-1)-1)$–connected, so $\partial(\boldsymbol{x}a)*\langle \mathrm{lk}_Y(\eta)\rangle$ is $(n-2)$–connected. An application of the Mayer–Vietoris sequence and the van Kampen theorem then shows that $\Delta_i(\boldsymbol{x})\cup\Delta_{i-1}$ is $(n-1)$–connected.

If we fix an ordered simplex $\boldsymbol{z}=z_0\dots z_{i-1}$ different from $\boldsymbol{x}$, then the intersection of $\Delta_i(\boldsymbol{z})$ and $\Delta_i(\boldsymbol{x})$ is contained in Δ_{i-1} since simplices in the intersection can only be obtained by deleting at least one vertex of $\boldsymbol{x}$ and of $\boldsymbol{z}$ (and possibly other vertices). We can then apply the above argument inductively to show that attaching finitely many complexes $\Delta_i(\boldsymbol{x})$ to Δ_{i-1} preserves the connectivity $n-1$. Since homotopy groups commute with direct limits, it follows that the entire subspace Δ_i is $(n-1)$–connected. Since $\Delta_n = \langle \mathrm{st}_X(\sigma) \rangle$, the claim is established. $\square$

We now proceed to build our homotopy $f \simeq g$. The semisimplicial complex $\langle X \rangle$ has the property that the vertices of each simplex are all distinct, so its barycentric subdivision $\langle X \rangle'$ is a simplicial complex. We view f and g as maps $S^{n-1} \to \langle X \rangle'$, which we may take to be simplicial with respect to some triangulation of S^{n-1}. We build the homotopy inductively on the skeleta of S^{n-1}.

If v is a vertex of S^{n-1} then $f(v)$ and $g(v)$ project to the same vertex of X', ie to the barycenter of some simplex σ of X. Since $\langle \mathrm{st}_X(\sigma) \rangle$ is $(n-1)$–connected there is a path in $\langle \mathrm{st}_X(\sigma) \rangle$ connecting $f(v)$ to $g(v)$, and we use this to define our homotopy on $v \times I$.

Now let s be any simplex of S^{n-1} and assume we have already defined a homotopy $f \simeq g$ on ∂s. The projection of $f(s)$ (and hence $g(s)$) to X is a simplex of X', ie a chain $\sigma_0 \subset \cdots \subset \sigma_k$ of simplices of X. The stars of these simplices satisfy the reverse inclusions $\mathrm{st}_X(\sigma_0) \supset \cdots \supset \mathrm{st}_X(\sigma_k)$, hence the same is true for the ordered versions of these stars. We may assume by induction that the homotopy from f to g on ∂s takes place in $\langle \mathrm{st}_X(\sigma_0) \rangle$. Since $\langle \mathrm{st}_X(\sigma_0) \rangle$ is $(n-1)$–connected and $\dim(\partial s) \le n-2$, we can extend the homotopy over the interior of s so that its image lies in $\langle \mathrm{st}_X(\sigma_0) \rangle$. This finishes the induction step. $\square$

3 A simple example: the braid group

As a warm up for our main case of mapping class groups let us first show how the method described in this paper gives a simple proof of homology stability for the classical braid groups B_n, where we are viewing B_n as the mapping class group of an n–punctured disk.

We start by constructing a suitable "tethered" complex with a B_n–action. In fact, in this case the tethers will be all there is to the complex. Consider a fixed disk D with d distinguished points $b_1, \dots, b_d$ on the boundary and n marked points or punctures $p_1, \dots, p_n$ in the interior. A *tether* is an arc in D connecting some p_i to some b_j and disjoint from the other p_k and b_k. A *system of tethers* is a collection of tethers which are disjoint except at their endpoints, and with no two of the tethers isotopic. See Figure 4 for an example.

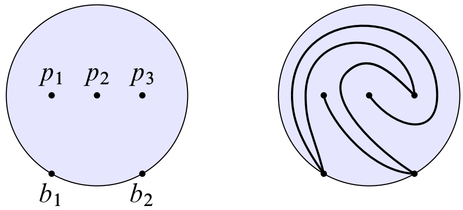

Figure 4: A system of tethers in a 3–punctured disk with two boundary points

Define the tether complex $T = T_{n,d}$ to be the geometric complex having one k–simplex for each isotopy class of systems of $k+1$ tethers, where the face relation is given by omitting tethers.

Proposition 3.1 *T is contractible.*

Proof We choose a single fixed tether t, then use a surgery flow to deform T into the star of the vertex t. The flow will decrease the complexity of a system σ (in normal form with respect to t), which we define to be the total number of points in the intersection of the interiors of σ and t.

If s is a tether which intersects t at an interior point, let x be the intersection point which is closest along t to the end b_i of t. Perform surgery on s by cutting it at x and moving both new endpoints down to b_i (see Figure 5). This creates two new arcs which can be isotoped to be disjoint from s except at their endpoints. One of these arcs joins b_i to a puncture, and one joins b_i to some (possibly different) b_j. Define Δs to be the arc connecting b_i to a puncture. Note that Δs has smaller complexity than s.

The conditions of Lemma 2.9 are now met, with $X = T$ and the star of t as the subcomplex Y, by defining v_σ to be the tether in σ containing the point of $\mathrm{int}(\sigma) \cap \mathrm{int}(t)$ closest to b_i along t. Thus T deformation retracts to the star of t, which is contractible, hence T is contractible. $\square$

A system of tethers $\tau = \{t_1, \dots, t_k\}$ is *coconnected* if the complement $D - \tau$ is connected. Note that a system is coconnected if and only if each arc in the system

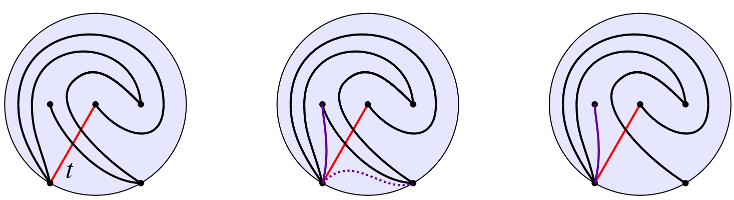

Figure 5: Surgery on a system of tethers using a fixed tether t

ends at a different puncture. Let $T^0 = T^0_{n,d}$ be the subcomplex of $T_{n,d}$ consisting of isotopy classes of coconnected tether systems.

Proposition 3.2 *The complex $T^0 = T^0_{n,1}$ is contractible.*

Proof We prove that T^0 is contractible by induction on the number n of punctures. If $n = 1$ then T^0 is a single point. For the induction step we will use a link argument (Corollary 2.2) to show the inclusion map $T^0 \hookrightarrow T = T_{n,1}$ is a homotopy equivalence, so we need to specify which simplices of T are bad. We define a simplex of T to be *bad* if each tethered puncture has at least two tethers. We check (1) every simplex in T which is not in T^0 has a bad face, and (2) if σ and τ are two bad faces of a simplex of T, the join $\sigma * \tau$ is also bad.

If σ is a bad simplex, we also need to identify the subcomplex G_σ of $\mathrm{lk}(\sigma)$ consisting of simplices which are good for σ. In our case $\tau \in \mathrm{lk}(\sigma)$ is good for σ if and only if τ consists of single tethers to punctures which are not used by σ. The subcomplex G_σ decomposes as a join $G_\sigma = T^0(P_1) * T^0(P_2) * \cdots * T^0(P_r)$, where $P_1, \dots, P_r$ are the components of the space obtained by cutting D open along σ and each $T^0(P_i)$ is either empty or isomorphic to $T^0_{n_i,d_i}$ for some $n_i < n$. Two tethers in σ going to the same puncture bound a disk in D. A minimal such disk must be a component P_i with at least one puncture in its interior (since isotopic tethers are not allowed) and only one distinguished boundary point. Thus $T^0(P_i) \cong T^0_{n_i,1}$ is contractible by induction on n, and the entire join G_σ is contractible. The hypotheses of Corollary 2.2(a) are satisfied and we conclude that T^0 is contractible since T is. □

Theorem 3.3 *The stabilization $H_i(B_{n-1}) \to H_i(B_n)$ is an isomorphism for $n > 2i+1$ and a surjection for $n = 2i + 1$.*

Proof We use the spectral sequence constructed in Section 1 for the action of B_n on the contractible complex $T^0 = T^0_{n,1}$. Recall that this action arises from regarding B_n as the group of isotopy classes of diffeomorphisms of the disk that are the identity on the boundary and permute the punctures p_i. We verify conditions (1)–(3) at the beginning of Section 1.

(1) T^0 has dimension $n-1$ and the action of B_n has only one orbit of k–simplices for each k.

(2) To see that the stabilizer of a k–simplex fixes the simplex pointwise, note that a set of $k+1$ tethers coming out of the basepoint in the boundary of the disk has a natural ordering determined by an orientation of the disk at the basepoint, and this ordering is preserved by any diffeomorphism of the disk that is the identity on the boundary. The stabilizer of a k–simplex is therefore isomorphic to B_{n-k-1}.

(3) For an edge of T^0 corresponding to a pair of tethers there is a diffeomorphism of the disk supported in a neighborhood of the two tethers that interchanges the punctures at the ends of the tethers and takes the first tether to the second or vice versa. This diffeomorphism gives an element of B_n commuting with the stabilizer of the edge.

Theorem 1.1(a) now gives the result since T^0 is contractible by Proposition 3.2. □

In this simple example we can in fact deduce more from the spectral sequence without much work:

Theorem 3.4 *When n is odd the stabilization $H_i(B_{n-1}) \to H_i(B_n)$ is an isomorphism for all i. Also, $H_i(B_n) = 0$ for $i \geq n$ (for n of either parity).*

Proof We look more closely at the spectral sequence used in the proof of Theorem 3.3 above. The E^1 page has the following form:

$$
\begin{array}{r|ccccccccc}
i & H_i(B_n) & \longleftarrow & H_i(B_{n-1}) & \xleftarrow{0} & H_i(B_{n-2}) & \xleftarrow{\cong} & \cdots & & \\
i-1 & \cdots & \longleftarrow & H_{i-1}(B_{n-1}) & \xleftarrow{0} & H_{i-1}(B_{n-2}) & \xleftarrow{\cong} & H_{i-1}(B_{n-3}) & \xleftarrow{0} & \cdots \\
i-2 & & & & & \cdots & \xleftarrow{\cong} & H_{i-2}(B_{n-3}) & \xleftarrow{0} & \cdots \\
 & & & & & & & & & \\
q=0 & H_0(B_n) & \longleftarrow & H_0(B_{n-1}) & \xleftarrow{0} & H_0(B_{n-2}) & \xleftarrow{\cong} & H_0(B_{n-3}) & \xleftarrow{0} & \cdots \\
\hline
 & p=-1 & & 0 & & 1 & & 2 & & \cdots
\end{array}
$$

To see that the differentials are alternately zeros and isomorphisms as shown, note first that the observation we used in the proof of Theorem 3.3 to verify condition (3) holds more generally to show that for any system σ of $k+1 \geq 2$ tethers there is a diffeomorphism of the disk permuting the punctures and supported in a neighborhood of the tethers that takes any subset of k of the tethers to any other set of k of the tethers, preserving their natural order and commuting with $\mathrm{stab}(\sigma)$. This implies that each of the $p+1$ terms of the map $d^1\colon E^1_{p,q} = H_q(\mathrm{stab}(\sigma)) \to E^1_{p-1,q}$ is the same, so, for p odd, d^1 is zero, and for p even, d^1 is the map induced by inclusion. If we assume n is odd then by induction on n starting with the trivial case $n=1$ the differential $d^1\colon E^1_{p,q} \to E^1_{p-1,q}$ is an isomorphism for p even, $p > 0$. In particular, for the rightmost nonvanishing column, which is the $p = n-1$ column since T^0 has dimension $n-1$, the d^1 differentials originating in this column are isomorphisms since n is odd. (The only nonzero term in this column is $H_0(B_0) = \mathbb{Z}$ since B_0, like B_1, is the trivial group.) Thus in the E^2 page all the terms to the right of the $p = 0$ column vanish. Since the spectral sequence converges to zero and no differentials beyond the E^1 page can be nonzero, it follows that the differentials $d^1\colon H_i(B_{n-1}) \to H_i(B_n)$ must be isomorphisms for all i, which finishes the induction step to prove the first part of the theorem.

For the second statement of the theorem we again look at the E^1 page of the spectral sequence. The groups along the diagonal $p + q = n - 1$ are the groups $H_j(B_j)$. By induction on n all the terms on or above this diagonal are zero except possibly in the $p = -1$ column, where the groups on or above the diagonal are $H_i(B_n)$ for $i \geq n$. Since the spectral sequence converges to zero, all of these terms must vanish as well. □

The fact that $H_i(B_n)$ vanishes for $i \geq n$ is also a consequence of the well-known fact that there is an Eilenberg–Mac Lane space $K(B_n, 1)$ which is a CW complex of dimension $n - 1$ (see eg [6]). Arnol'd [2] proved the statements in the two preceding theorems by methods not involving spectral sequences.

Arnol'd also computed the homology of the pure braid subgroup $P_n \subset B_n$ in [1] and it does not stabilize, even at the level of H_1, which is free abelian of rank $\binom{n}{2}$, as can be seen already from a presentation for P_n. When the action of B_n on T^0 is restricted to P_n it is no longer transitive on simplices of each dimension, and in particular not on vertices. We note that homology stability can sometimes still be proved using an action which is not transitive on simplices, as long as the number of orbits is independent of the stabilization parameters. However, the spectral sequence argument becomes more complicated if there is more than one orbit.

4 Curve and arc complexes

For a compact orientable surface $S = S_{g,s}$ of genus g with s boundary components the classical *curve complex* $C(S)$ has as its vertices the isotopy classes of embedded curves (circles) in S which are nontrivial, ie do not bound a disk and are not isotopic to a component of ∂S. A set of vertices of $C(S)$ spans a simplex if the corresponding curves can be isotoped to be all disjoint, so they form a *curve system*. We will be particularly interested in the subcomplex $C^0(S)$ whose simplices are the isotopy classes of coconnected curve systems, ie systems with connected complement. We will show that $C^0(S)$ is highly connected by showing that $C(S)$ is highly connected and using a link argument to deduce the connectivity for $C^0(S)$. These results are due originally to Harer [9] and we follow the same overall strategy while simplifying the proofs of several of the individual steps.

4.1 Curves on surfaces with nonempty boundary

To prove $C(S)$ is highly connected when ∂S is nonempty the idea is to compare $C(S)$ with three other complexes in a sequence

$$A(S,\partial_0 S) \supset A_\infty(S,\partial_0 S) \simeq S(S,\partial_0 S) \simeq C(S).$$

The case that S is closed will be deduced from the nonclosed case.

We start by defining $A(S,\partial_0 S)$. An *arc system* on a bounded surface S is a set of disjoint embedded arcs with endpoints on the boundary ∂S, such that no arc is isotopic to an arc in ∂S and no two arcs in a system are isotopic to each other, where all isotopies of arcs are required to keep their endpoints in ∂S. We choose a component $\partial_0 S$ of ∂S and define the complex $A(S,\partial_0 S)$ as the geometric complex whose k–simplices are the isotopy classes of systems of $k+1$ arcs whose endpoints all lie in $\partial_0 S$.

Proposition 4.1 *The complex $A(S,\partial_0 S)$ is contractible whenever it is nonempty, ie when S is not a disk or annulus.*

Proof This is an application of Lemma 2.9, using surgery to flow into the star of a fixed "target" arc a. The complexity of a system that intersects a minimally within its isotopy class is defined as the number of intersection points with a. To do the surgery we first choose an orientation for a. An arc b crossing a is cut into two arcs at the point where it meets a nearest the terminal point of a, and the two new endpoints are moved to this terminal point to produce a new arc system Δb meeting a in one fewer point than b. The function $\sigma \mapsto b_\sigma$ assigns to a system σ the arc of σ meeting a at the point closest to the terminal point of a. □

We define an arc system in $A(S,\partial_0 S)$ to be *at infinity* if it has some complementary component which is neither a disk nor an annular neighborhood of a boundary component. (The terminology comes from the fact, observed by Harer, that arc systems at infinity can be identified with rational points in the boundary of the Teichmüller space of the surface.) Arc systems at infinity form a subcomplex $A_\infty(S,\partial_0 S)$. A calculation using Euler characteristics shows that it takes at least $(2g+s-1)$ arcs to cut S into disks and annuli when $s=1$, so in this case $A_\infty(S,\partial_0 S)$ contains the entire $(2g+s-3)$–skeleton of $A(S,\partial_0 S)$. When $s>1$ there are similar statements with $2g+s-1$ replaced by $2g+s-2$ and $2g+s-3$ replaced by $2g+s-4$. The inclusion $A_\infty(S,\partial_0 S) \hookrightarrow A(S,\partial_0 S)$ induces an injection on π_i when $A_\infty(S,\partial_0 S)$ contains the $(i+1)$–skeleton of $A(S,\partial_0 S)$, so we deduce:

Corollary 4.2 *$A_\infty(S,\partial_0 S)$ is $(2g+s-4)$–connected if $s=1$ and $(2g+s-5)$–connected if $s>1$.*

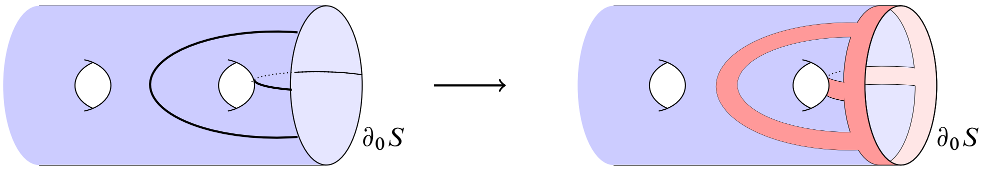

Figure 6: The map $\widehat{A}_\infty(S, \partial_0 S) \to \widehat{S}(S, \partial_0 S)$

Now define the *subsurface complex* $S(S, \partial_0 S)$ to be the geometric realization of the poset $\widehat{S}(S, \partial_0 S)$ of isotopy classes of compact connected subsurfaces F of S such that one component of ∂F is $\partial_0 S$ and the other components of ∂F that are not contained in ∂S form a nonempty curve system in S, possibly containing parallel copies of the same curve. In particular, no component of $\partial F - \partial S$ bounds a disk in S or is isotopic to a component of ∂S.

To each arc system α with $\partial\alpha \subset \partial_0 S$ we can associate a subsurface $F(\alpha)$ of S by first taking a regular neighborhood N of $\alpha \cup \partial_0 S$ and then adjoining any components of $S - N$ that are disks or annuli with one boundary circle contained in ∂S (see Figure 6). Thus the simplices of $A_\infty(S, \partial_0 S)$ correspond to systems α for which $F(\alpha) \neq S$, and $\alpha \mapsto F(\alpha)$ is a map $f\colon \widehat{A}_\infty(S, \partial_0 S) \to \widehat{S}(S, \partial_0 S)$, where $\widehat{A}_\infty(S, \partial_0 S)$ denotes the poset of simplices in $A_\infty(S, \partial_0 S)$. This map is a poset map since $\alpha \subset \beta$ implies $F(\alpha) \subset F(\beta)$.

Proposition 4.3 *The map* $f\colon \widehat{A}_\infty(S, \partial_0 S) \to \widehat{S}(S, \partial_0 S)$ *is a homotopy equivalence.*

Proof We apply Quillen's fiber lemma, Proposition 2.5. If F is a subsurface of S then $f_{\leq F}$ is all arc systems α with $F(\alpha) \subset F$, so this is $\widehat{A}(F, \partial_0 S)$. Since F is not a disk or annulus, $\widehat{A}(F, \partial_0 S)$ is contractible by Proposition 4.1. $\square$

Given a curve system γ, let the subsurface $F(\gamma) \subset S$ be the component of the complement of a regular neighborhood of γ containing $\partial_0 S$ (see Figure 7). Note that if $\gamma \subset \gamma'$ then $F(\gamma) \supset F(\gamma')$. Thus if $\widehat{C}(S)$ denotes the poset of simplices of $C(S)$, then the association $\gamma \mapsto F(\gamma)$ defines a poset map $g\colon \widehat{C}(S) \to \widehat{S}(S, \partial_0 S)$ with respect to the reverse ordering on $\widehat{S}(S, \partial_0 S)$ defined by $F_1 \leq F_2$ if $F_1 \supset F_2$.

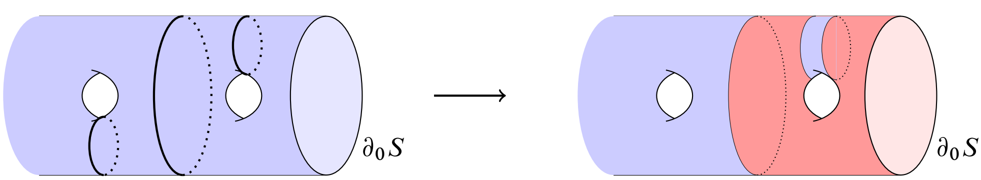

Figure 7: The map $\widehat{C}(S) \to \widehat{S}(S, \partial_0 S)$

Proposition 4.4 *The map $g\colon \widehat{C}(S) \to \widehat{S}(S, \partial_0 S)$ is a homotopy equivalence.*

Proof We again apply Proposition 2.5. For a subsurface F in $\widehat{S}(S, \partial_0 S)$, the fiber $g_{\leq F}$ consists of curve systems in $S - F$, where curves are allowed to be parallel to curves of the system $\gamma(F) = \partial F - \partial S$. In particular, $\gamma(F)$ is in the fiber, and $\gamma(F)$ can be added to any curve system in the fiber, so the poset maps $\gamma \mapsto \gamma \cup \gamma(F) \mapsto \gamma(F)$ give a deformation retraction of $g_{\leq F}$ to the point $\gamma(F)$. □

Corollary 4.5 *If ∂S is not empty, then $C(S)$ is $(2g+s-4)$–connected if $s = 1$ and $(2g+s-5)$–connected if $s > 1$.*

Corollary 4.6 *If S has genus 0, then $C(S)$ is homotopy equivalent to a wedge of spheres of dimension $s-4$.*

Proof If S has genus 0, $C(S)$ is $(s-5)$–connected by the preceding corollary, and it is $(s-4)$–dimensional, so it is homotopy equivalent to a wedge of spheres of dimension $s-4$. □

Remark In fact $C(S)$ is homotopy equivalent to a wedge of spheres in all cases. When $g > 0$ the dimension of the spheres is $2g+s-3$ if $s > 0$ and $2g-2$ if $s = 0$. This was proved by Harer [10, Theorems 3.3 and 3.5]. Thus the connectivity statements derived above are best possible when $s = 1$ but one below best possible when $s > 1$. However, these stronger results are not needed for the proof of homology stability.

4.2 Curves on closed surfaces

There is a map $\phi\colon C(S_{g,1}) \to C(S_{g,0})$ induced by filling in the boundary circle of $S_{g,1}$ with a disk. We remark that the dimension of $C(S_{g,1})$ is one more than that of $C(S_{g,0})$ when $g > 1$ since maximal curve systems cut S into pairs of pants. For $g = 1$ the map $C(S_{1,1}) \to C(S_{1,0})$ is an isomorphism.

Proposition 4.7 *The map $\phi\colon C(S_{g,1}) \to C(S_{g,0})$ is a homotopy equivalence for each $g \geq 1$.*

The weaker statement that $\phi_*\colon \pi_k C(S_{g,1}) \to \pi_k C(S_{g,0})$ is surjective for all k suffices to prove that $C(S_{g,0})$ is $(2g-3)$–connected, which is all we will need for homology stability. The surjectivity of ϕ_* has a short proof using a little hyperbolic geometry, as follows: Choose a hyperbolic structure on $S_{g,0}$ in the nontrivial cases $g \geq 2$. Given a map $f\colon S^k \to C(S_{g,0})$ which we may assume is simplicial in some triangulation of S^k, the images $f(v)$ of all the vertices v in S^k can be represented by geodesics.

These are unique in their homotopy classes and are disjoint for sets of vertices spanning a simplex in $f(S^k)$. Then a lift of f to $C(S_{g,1})$ is obtained by deleting a disk in $S_{g,0}$ disjoint from this finite set of geodesics.

To obtain the full strength of Proposition 4.7, here is a proof that uses only topological techniques:

Proof We may assume $g \geq 2$. It will suffice to show that for each simplex σ of $C(S_{g,0})$ the subcomplex $F_\sigma = \phi^{-1}(\sigma)$ of $C(S_{g,1})$ is contractible, by Proposition 2.5 and Lemma 2.6. To begin, choose a curve system $\widetilde{\sigma}$ in $S_{g,1}$ with $\phi(\widetilde{\sigma}) = \sigma$. Enlarge $\widetilde{\sigma}$ to a maximal curve system δ cutting $S_{g,1}$ into pairs of pants. Let P be the pair of pants containing $\partial S_{g,1}$ and let d_1 and d_2 be the other two boundary circles of P. We can choose δ so that d_1 is a curve of $\widetilde{\sigma}$.

We may assume that all curve systems γ in $S_{g,1}$ are in normal form with respect to δ, so γ intersects δ transversely in the minimum number of points among all systems isotopic to γ. This minimality is equivalent to the "no bigon" condition that S contains no disk whose boundary consists of an arc in γ and an arc in δ. If two systems in normal form with respect to δ are isotopic, then they are isotopic through systems transverse to δ, except that curves in γ isotopic to curves in δ can be pushed from one side of δ to the other and such an isotopy cannot be transverse to δ at all times.

If a curve system γ is in normal form with respect to δ then each component arc of $\gamma \cap P$ either crosses P from d_1 to d_2, or it enters P, goes around $\partial S_{g,1}$, and leaves by crossing the same d_i that it crossed when it entered P. An arc of the latter type we call a *return arc*. Note that all return arcs of γ must have their endpoints on the same d_i.

We will use surgery to flow from F_σ into the subcomplex F_σ^{nr} consisting of curve systems with no return arcs. Let c be a curve in normal form with respect to δ that contains return arcs. Let b be the innermost of these return arcs, the one closest to $\partial S_{g,1}$. Pushing b across $\partial S_{g,1}$ converts c into a new curve Δc which can be isotoped to be disjoint from c. (See Figure 8). Alternatively, we can view Δc as the result of surgering c along an arc of ∂P to produce two curves, one of which is isotopic to $\partial S_{g,1}$ and is discarded. The curve Δc may not be in normal form with respect to δ, but it can be made so by an isotopy eliminating bigons one by one. Note that Δc is in F_σ if c is since the two curves become isotopic when ∂S_0 is capped off with a disk.

If γ is a curve system in F_σ with at least one return arc, define c_γ to be the curve in γ containing the innermost return arc of γ. Pushing c_γ across $\partial S_{g,1}$ as above then yields the curve Δc_γ. If we define the complexity of a curve system to be the number of intersection points with ∂P, we then have the ingredients to apply Lemma 2.9, producing a deformation retraction of F_σ to F_σ^{nr}.

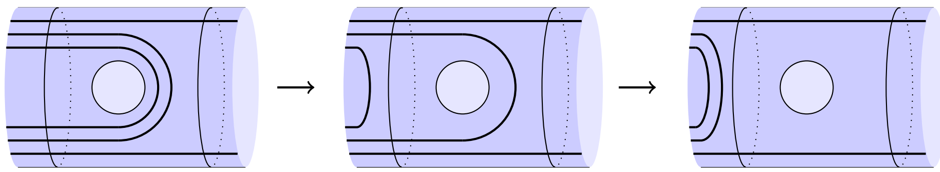

Figure 8: The flow in the proof of Proposition 4.7

We claim that F_σ^{nr} is just a single simplex, the simplex spanned by $\widetilde{\sigma}$ and d_2. To see this, observe that if γ is a simplex in F_σ^{nr} in normal form with respect to δ, and with any of its curves parallel to d_1 or d_2 pushed outside P, then we can obtain the normal form for $\phi(\gamma)$ by simply deleting P from $S_{g,1}$ and identifying d_1 and d_2 in such a way as to match up the endpoints of any arcs of $\gamma \cap P$. In fact there can be no such arcs since $\phi(\gamma)$ is a face of σ, hence γ is a face of $\widetilde{\sigma} * d_2$. Thus F_σ^{nr} is a simplex and it follows that F_σ is contractible. □

4.3 Coconnected curve systems

The complex $C^0(S)$ of coconnected curve systems on a surface of genus g has dimension $g-1$. As we will observe in Remark 4.9 below, the top-dimensional homology group $H_{g-1}(C^0(S))$ is nonzero, so the best one could hope is that $C^0(S)$ is $(g-2)$–connected, and indeed it is:

Proposition 4.8 *The complex $C^0(S)$ of coconnected curve systems on a surface S of genus g is $(g-2)$–connected.*

Proof This is a link argument, an application of Corollary 2.2 with $X = C(S)$ and $Y = C^0(S)$. To begin we need to single out the bad simplices of $C(S)$. To each curve system we associate a dual graph, with a vertex for each complementary component of the system and an edge for each curve. Thus a curve system is coconnected if and only if its dual graph has one vertex and all edges are loops. A *bad simplex* in $C(S)$ is a system of curves for which *no* edges of the dual graph are loops. This is equivalent to saying that each curve in the system separates the complement of the other curves. It is easy to see that conditions (1) and (2) in Section 2.1 are satisfied for this notion of badness. For a bad simplex σ, the complex G_σ is the join of the complexes $C^0(S_i)$ for the components S_i of the surface S_σ obtained by cutting S open along σ. Either the genus g_i of S_i is smaller than g or S_i has fewer boundary components than S so we may proceed by induction on the lexicographically ordered pair (g,s). Since $g_\sigma = \sum_i g_i$ by definition and the quantity "connectivity plus two" is additive for joins, it follows that we may assume inductively that G_σ is $(g_\sigma-2)$–connected. The induction can start with the obvious cases that the genus is zero or one.

Cutting S along the curves of σ can decrease the total genus by at most $\dim(\sigma)$ since each cut decreases genus by at most one and cutting along the last curve cannot decrease genus since σ is bad. Thus $g_\sigma \geq g - \dim(\sigma)$. (This estimate is best possible in the case that the dual graph to σ consists of two vertices joined by a number of edges.) It follows that the connectivity of G_σ, which inductively is at least $g_\sigma - 2$, is also at least $g - 2 - \dim(\sigma)$ so the connectivity hypothesis on links in Corollary 2.2(a) is satisfied with $n = g - 2$.

The last thing to check to apply the corollary is that the larger complex $C(S)$ is $(g-2)$–connected. We can assume $g > 0$ since the proposition is trivial when $g = 0$. Then Corollary 4.5 and Proposition 4.7 imply that $C(S)$ is $(2g-3)$–connected, and we have $g - 2 \leq 2g - 3$ when $g \geq 1$. $\square$

Remark 4.9 There is an easy argument showing that $H_{g-1}(C^0(S_{g,s}))$ is nonzero for all $g \geq 1$ and $s \geq 0$. Choosing g disjoint copies of $S_{1,1}$ in $S_{g,s}$ gives an embedding of the join of g copies of $C^0(S_{1,1})$ into $C^0(S_{g,s})$ as a subcomplex. The complex $C^0(S_{1,1}) = C(S_{1,1})$ is an infinite discrete set, so the join is homotopy equivalent to the wedge of an infinite number of copies of S^{g-1}. The inclusion map of the join into $C^0(S_{g,s})$ induces an injection on H_{g-1} since both complexes have dimension $g-1$ so no nontrivial $(g-1)$–dimensional cycle in the join can bound in $C^0(S_{g,s})$. Thus $H_{g-1}(C^0(S_{g,s}))$ is nontrivial, and in fact is free abelian of infinite rank since it is the kernel of the boundary map from the free abelian group of simplicial $(g-1)$–chains to the simplicial $(g-2)$–chains, and a subgroup of a free abelian group is free abelian.

There is an oriented version of $C^0(S)$ whose simplices are isotopy classes of coconnected systems of curves together with choices of orientations for these curves. Call the resulting complex $C^0_\pm(S)$.

Corollary 4.10 *The complex $C^0_\pm(S)$ is $(g-2)$–connected.*

Proof Choose an arbitrary orientation for each isotopy class of nonseparating curves in S. Then the two possible orientations correspond to the labels $+$ and $-$ and the result is immediate from Corollary 2.3. $\square$

Remark 4.11 There are also versions of $C^0(S)$ and $C^0_\pm(S)$ in which simplices correspond to *ordered* coconnected systems of curves or oriented curves. These too have the same connectivity as $C^0(S)$ by Proposition 2.10.

5 Tethered curves and chains

This section represents the heart of the paper, where we introduce the geometric complexes that encode more information than is given by curves or arcs alone. The main work is in showing that the new complexes are roughly half as highly connected as $C^0(S)$, but this is enough for the spectral sequence arguments. The various complexes we will consider fit into a commutative diagram:

$$\begin{tikzcd} \mathrm{TCh}(S) \arrow[r] \arrow[d] & \mathrm{DTC}(S) \arrow[r] & \mathrm{TC}(S) \arrow[d] \\ \mathrm{Ch}(S) \arrow[rr] & & C^0(S) \end{tikzcd}$$

The maps are forgetful maps except for the upper left horizontal map which is an injection. We will start with the known connectivity of the complex $C^0(S)$ in the lower right corner, then proceed around the diagram in the counterclockwise direction to show each complex in turn is highly connected. Except for the one injection, each step will involve two stages: first enlarge the domain complex to a complex for which a surgery flow can be used to show that the fibers of the extended forgetful map are contractible, then use a link argument to shrink back to the original source complex. For the injection we need only a link argument to show that the image (and hence the domain) is highly connected.

5.1 Tethered curves

Let P be a nonempty finite collection of disjoint open intervals and circles in ∂S. A *tether* for a simple closed curve c in S is an arc in S with one endpoint in c and the other in P, the interior of the arc being disjoint from c and from ∂S. Define a complex $\mathrm{TC}(S,P)$ whose k–simplices are isotopy classes of systems of $k+1$ disjoint tethered curves such that the complement of the system of tethered curves is connected. This last condition is equivalent to the curves by themselves forming a coconnected system since, after cutting S open along the curves, each newly created boundary circle is connected to P by at most one tether arc, so cutting along these arcs cannot disconnect the surface. Note that tethering a curve gives it a normal orientation, pointing away from curve in the direction of the tether. An orientation of S converts the normal orientation of the curve to a tangential orientation.

Proposition 5.1 *The complex* $\mathrm{TC}(S,P)$ *is* $\frac{1}{2}(g-3)$*–connected, where* g *is the genus of* S.

Proof We will include $\mathrm{TC}(S,P)$ into a larger complex $\mathrm{TC}(S,P^*)$, then project this larger complex onto $C^0_\pm(S)$, the oriented version of $C^0(S)$. We show that the projection is a homotopy equivalence using a surgery argument on barycentric fibers. This shows that $\mathrm{TC}(S, P^*)$ is $(g-2)$–connected, and then a link argument will show that $\mathrm{TC}(S, P)$ is $\frac{1}{2}(g-3)$–connected.

The complex $\mathrm{TC}(S, P^*)$ has the same vertices as $\mathrm{TC}(S, P)$ but the simplices correspond to coconnected curve systems with at least one but possibly more tethers to each curve. All the tethers must be disjoint, and all tethers to the same curve must attach on the same side of the curve (at distinct points), giving the curve the same orientation.

The projection $\mathrm{TC}(S, P^*) \to C^0_\pm(S)$ forgets the tethers but remembers the orientations they induce on curves. The fiber of this projection over the barycenter of a simplex σ of $C^0_\pm(S)$ consists of all tether systems for σ attaching on the "positive" sides of the curves in σ. Choose one such system, which for simplicity we take to lie in $\mathrm{TC}(S, P)$ so that it consists of a single tether t_i to each curve c_i in σ. We can deform the barycentric fiber into the star of this tethered system by a surgery flow. The surgeries are performed first using the tether t_1, surgering toward P until all tethers are disjoint from t_1, then using t_2 in similar fashion, and so on. Each surgery cuts a tether into two arcs, one of which has both ends in P and which we discard, while the other is a tether which we keep. To make the surgery process well-defined on isotopy classes one must first put the tethers being surgered into normal form with respect to the fixed tethers t_i; this minimizes the number of intersection points with t_i by eliminating bigons and "half-bigons".

The surgery flow shows that the barycentric fiber over σ is contractible, so the projection $\mathrm{TC}(S, P^*) \to C^0_\pm(S)$ is a homotopy equivalence by Lemma 2.8. Thus $\mathrm{TC}(S, P^*)$ is $(g-2)$–connected by Corollary 4.10.

We now use a link argument to analyze the inclusion of $\mathrm{TC}(S, P)$ into $\mathrm{TC}(S, P^*)$. To see how the connectivity number $\frac{1}{2}(g-3)$ arises, let us try to show the inclusion is $(Ag+B)$–connected for yet-to-be-determined constants A and B. A *bad simplex* in $\mathrm{TC}(S, P^*)$ is one that corresponds to a system of tethered curves in which each curve has at least two tethers; in particular a vertex cannot be bad. If σ is a bad simplex, the surface S_σ obtained by cutting S open along the system of curves and tethers corresponding to σ may have several components S_i, and P is cut into pieces $P_i \subset \partial S_i$. The components S_i all have smaller genus than S and the subcomplex G_σ is the join of the complexes $\mathrm{TC}(S_i, P_i)$, so we can argue by induction on the genus.

To apply Corollary 2.2(a) we need to arrange that G_σ is $(Ag + B - k)$–connected for $k = \dim(\sigma)$. Cutting a surface along a multitethered curve system decreases genus by at most one for each tether, so if g_σ denotes the genus of S_σ (ie the sum of the genera g_i of the components S_i) we have $g_\sigma \geq g - k - 1$. Suppose we know by induction on genus that G_σ is $(Ag_\sigma + B)$–connected. The inequality we need is

$Ag_\sigma + B \ge Ag + B - k$, ie $Ag_\sigma \ge Ag - k$. We have observed that $g_\sigma \ge g - k - 1$, so it suffices to have $A(g-k-1) \ge Ag - k$, which simplifies to $A \le k/(k+1)$. We only need this for $k \ge 1$ since vertices cannot be bad, so $A = \frac{1}{2}$ works for all $k \ge 1$. We therefore choose $A = \frac{1}{2}$.

To apply Corollary 2.2(a) we also need $\mathrm{TC}(S, P^*)$ to be $\left(\frac{1}{2}g+B\right)$–connected, which means $\frac{1}{2}g + B \le g - 2$, the connectivity of $\mathrm{TC}(S, P^*)$. This inequality reduces to $B \le \frac{1}{2}g - 2$. We can assume $g \ge 1$ since the proposition is trivially true when $g = 0$. When $g = 1$ the inequality $B \le \frac{1}{2}g - 2$ is $B \le -\frac{3}{2}$, so we maximize B by choosing $B = -\frac{3}{2}$ and then $B \le \frac{1}{2}g - 2$ for all $g \ge 1$.

Thus our candidate for $Ag + B$ is $\frac{1}{2}(g-3)$. It remains to verify the induction step by showing that the join of the $\frac{1}{2}(g_i{-}3)$–connected complexes $\mathrm{TC}(S_i, P_i)$ is $\frac{1}{2}(g_\sigma{-}3)$–connected. We know that connectivity plus two is additive for joins, but this assumes the connectivities are integers and here they could be fractions. This means we need to use the floor function $\lfloor \cdot \rfloor$ for connectivities in order to apply the connectivity-plus-two fact. Thus we let

$$f(g) = \left\lfloor \tfrac{1}{2}(g-3) \right\rfloor + 2$$

and we wish to verify that $f(g_\sigma) \le \sum_i f(g_i)$.

We have $f(g) = \frac{1}{2}g + \frac{1}{2}$ if g is odd and $\frac{1}{2}g$ if g is even. If $g_\sigma = \sum_i g_i$ is odd then at least one g_i is odd, so $\sum_i f(g_i) \ge \left(\sum_i \frac{1}{2}g_i\right) + \frac{1}{2} = f(g_\sigma)$. If g_σ is even, we only need to notice that $\sum_i f(g_i) \ge \sum_i \frac{1}{2}g_i = f(g_\sigma)$. □

5.2 Double-tethered curves

We now consider a complex $\mathrm{DTC}(S, P, Q)$ of double-tethered curve systems. Here Q is a second nonempty finite collection of disjoint open intervals and circles in ∂S (we allow P and Q to overlap or even coincide), and a *double tether* for a curve c is an ordered pair of tethers attaching at the same point of c but on opposite sides, with the first tether going to a point in P and the second to a point in Q. The two tethers must be disjoint except at their common attaching point in c (see Figure 9). It is often useful to think of a double tether as a single oriented arc going from P to Q and crossing the curve at a single point. Note that a Dehn twist along the curve c acts nontrivially on the isotopy classes of double tethers for c, in contrast with the situation for single tethers.

A k–simplex of $\mathrm{DTC}(S, P, Q)$ is by definition an isotopy class of systems of $k+1$ disjoint double-tethered curves such that the complement of the system is connected. As before, this last condition is equivalent to the curves by themselves forming a coconnected system, since cutting along the tethers after cutting along the curves cannot disconnect the surface.

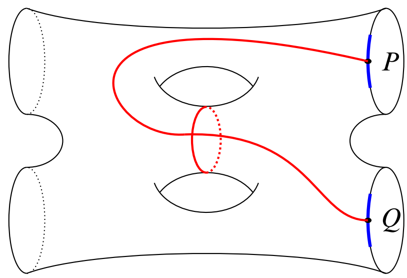

Figure 9: Double-tethered curve

Proposition 5.2 *The complex* $\mathrm{DTC}(S, P, Q)$ *is* $\frac{1}{2}(g-3)$*–connected, where* g *is the genus of* S*.*

Proof The proof follows closely the proof of Proposition 5.1. As before we include $\mathrm{DTC}(S, P, Q)$ into a larger complex $\mathrm{DTC}(S, P, Q^*)$, then project to $\mathrm{TC}(S, P)$. The complex $\mathrm{DTC}(S, P, Q^*)$ has the same vertices as $\mathrm{DTC}(S, P, Q)$; however higher-dimensional simplices of $\mathrm{DTC}(S, P, Q^*)$ correspond to isotopy classes of coconnected curve systems with one tether from each curve to P and at least one but possibly more tethers to Q, where all the tethers for a given curve attach at the same point of the curve and all the Q–tethers attach on the opposite side from the P–tether. All the tethers to P and Q in the system are disjoint from each other and from the curves except at the points where they attach to a curve. Faces of simplices in $\mathrm{DTC}(S, P, Q^*)$ are obtained by deleting one Q–tether to a curve if there are several, or by deleting the whole double-tethered curve if there is only one Q–tether to it. The vertices of the simplex are thus the double-tethered curves contained in the given system of curves and tethers.

The projection $\mathrm{DTC}(S, P, Q^*) \to \mathrm{TC}(S, P)$ forgets the tethers to Q, keeping only the single tether from each curve to P. As before, the barycentric fiber over a simplex of $\mathrm{TC}(S, P)$ can be contracted by surgery into the star of a fixed system in $\mathrm{DTC}(S, P, Q)$. Thus $\mathrm{DTC}(S, P, Q^*)$ is $\frac{1}{2}(g-3)$–connected.

We now use a link argument exactly as in the proof of Proposition 5.1 to deduce that $\mathrm{DTC}(S, P, Q)$ is $\frac{1}{2}(g-3)$–connected. The key point is that cutting S along a simplex of $\mathrm{DTC}(S, P, Q^*)$ decreases the genus by at most one for each tether to Q. This is because cutting along a nonseparating curve decreases genus by one, then cutting along the single tether to P does not decrease the genus further, nor does cutting along the first tether to Q, and cutting along each additional tether to Q can decrease genus by at most one. □

For the spectral sequence proof of homology stability we will use a certain subcomplex of $\mathrm{DTC}(S,P,Q)$ defined when P and Q are disjoint single intervals. To define this subcomplex we first choose orientations for P and Q. For a simplex of $\mathrm{DTC}(S,P,Q)$ the orientation of P induces an ordering of the double tethers of this simplex. Likewise the orientation of Q induces a possibly different ordering of the double tethers. The simplices for which the two orderings are in fact the same form a subcomplex of $\mathrm{DTC}(S,P,Q)$, which we denote $\mathrm{DTC}^m(S,P,Q)$, with the superscript indicating *matching* orderings.

Proposition 5.3 *The complex* $\mathrm{DTC}^m(S,P,Q)$ *is* $\frac{1}{2}(g-3)$*–connected.*

Proof This will be a link argument, following the idea of the proof of Theorem 4.9 of [25]. A simplex of $\mathrm{DTC}(S,P,Q)$ has vertices a set of pairs $(c_0,d_0),\dots,(c_k,d_k)$ consisting of curves c_i with double tethers d_i. We may assume these are listed in the order specified by the orientation of P. The ordering determined by the orientation of Q differs from this ordering by a permutation π of $\{0,1,\dots,k\}$, with π the identity exactly when the simplex is in $\mathrm{DTC}^m(S,P,Q)$. If π is not the identity let i be the smallest index such that $\pi(i)\neq i$. We call the given simplex *bad* if $i=0$. We can write the simplex uniquely as the join of a simplex $\langle(c_0,d_0),\dots,(c_{i-1},d_{i-1})\rangle$ which is good, ie in $\mathrm{DTC}^m(S,P,Q)$, and a simplex $\langle(c_i,d_i),\dots,(c_k,d_k)\rangle$ which is bad, where either of these two subsimplices could be empty. This notion of badness satisfies the two conditions in Section 2.1.

For a bad simplex $\sigma=\langle(c_0,d_0),\dots,(c_k,d_k)\rangle$ the subcomplex G_σ of simplices that are good for σ can be identified with $\mathrm{DTC}^m(S_\sigma,P_\sigma,Q_\sigma)$, where S_σ is the (connected) surface obtained by cutting S along σ and P_σ and Q_σ are the subintervals of P and Q up to the point where the first (with respect to the orientations of P and Q) tethers of σ attach. Cutting S open along each double tethered curve decreases genus by one, so by induction on genus we may assume that $\mathrm{DTC}^m(S_\sigma,P_\sigma,Q_\sigma)$ is $\frac{1}{2}(g-(k+1)-3)$–connected. Since there are no bad 0–simplices we have $k\geq 1$ and $\frac{1}{2}(g-(k+1)-3)\geq\frac{1}{2}(g-3)-k$. The result now follows from Corollary 2.2(a). □

5.3 Chains and tethered chains

The connectivity results obtained so far are enough to prove homology stability for mapping class groups of surfaces with nonempty boundary. However if a surface is closed there is no natural place for tethers to go, and we instead consider complexes of *chains*, where a chain is an ordered pair (a,b) of simple closed curves intersecting transversely in a single point, together with an orientation on b. The geometric complex of chains is denoted $\mathrm{Ch}(S)$.

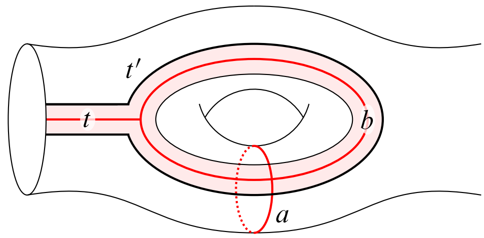

Figure 10: Double tether associated to a tethered chain

Remark 5.4 Forgetting the orientation on b gives a retraction of $\mathrm{Ch}(S)$ to a complex of unoriented chains, which has the same connectivity by an application of Corollary 2.3.

We will prove that $\mathrm{Ch}(S)$ is highly connected for all surfaces S, with or without boundary, but we start with a complex of tethered chains $\mathrm{TCh}(S_{g,s}, P)$ for a surface with nonempty boundary. Here each chain has one tether connecting the positive side of the (oriented) b–curve of the chain to a point in some finite collection P of disjoint open intervals (but no circles) in ∂S. Since the tether is only allowed to attach to the positive side of the b–curve, specifying the tether determines the orientation of b.

Proposition 5.5 *The complex* $\mathrm{TCh}(S_{g,s}, P)$ *is* $\frac{1}{2}(g-3)$*–connected.*

Proof Let $((a,b),t)$ be a vertex of $\mathrm{TCh}(S_{g,s}, P)$. A small neighborhood N of $b\cup t$ is homeomorphic to an annulus, one of whose boundary components intersects ∂S in an arc. Deleting the interior of this arc from this component of ∂N leaves a double tether t' for the a–curve (see Figure 10).

Thus the double tether t' and the a–curve give a vertex of the complex $\mathrm{DTC}(S_{g,s}, P)$ of double-tethered coconnected curve systems where the double tethers have both ends in P. In $\mathrm{DTC}(S_{g,s}, P)$ we do not orient the double tethers, in contrast with the double tethers in $\mathrm{DTC}(S_{g,s}, P, P)$, which do have specified orientations. This map on vertices $((a,b),t)\mapsto(a,t')$ extends to a simplicial embedding

$$\mathrm{TCh}(S_{g,s}, P)\hookrightarrow \mathrm{DTC}(S_{g,s}, P).$$

The image consists of simplices with the special property that the two ends of each double tether used in the simplex are adjacent in P. We denote this image by $\mathrm{DTC}^a(S_{g,s}, P)$, with the superscript indicating the adjacency of the two ends of a double tether. Thus it suffices to prove that $\mathrm{DTC}^a(S_{g,s}, P)$ is $\frac{1}{2}(g-3)$–connected. We will do this by a link argument similar to the one for Proposition 5.3. This will use the fact that the larger complex $\mathrm{DTC}(S_{g,s}, P)$ is $\frac{1}{2}(g-3)$–connected, which follows by embedding it in $\mathrm{DTC}(S_{g,s}, P, P)$ as a retract by arbitrarily choosing orientations for all the double tethers of vertices of $\mathrm{DTC}(S_{g,s}, P)$; the retraction is obtained by replacing all orientations by these arbitrarily chosen ones, and $\mathrm{DTC}(S_{g,s}, P, P)$ is $\frac{1}{2}(g-3)$–connected by Proposition 5.2.

For the link argument assume first that P is a single interval and choose an orientation for P. This allows us to order the ends of the double tethers in each simplex of $\mathrm{DTC}(S_{g,s}, P)$. Define a simplex of $\mathrm{DTC}(S_{g,s}, P)$ to be *bad* if its first double-tether end in P is not immediately followed by the other end of this double tether. (Note that vertices cannot be bad.) Each simplex in $\mathrm{DTC}(S_{g,s}, P)$ is then the join of two of its faces, the first face consisting of a string (possibly empty) of adjacently double-tethered curves whose tether ends form an initial segment of the sequence of all the tether ends, and the second face a bad simplex whose tether ends form the rest of the sequence. The two conditions for badness in Section 2.1 are easily checked.

For a bad k–simplex σ of $\mathrm{DTC}(S_{g,s}, P)$ the subcomplex G_σ of simplices that are good for σ can be identified with $\mathrm{DTC}^a(S_\sigma, P_\sigma)$ where S_σ is the surface obtained by cutting S along σ and P_σ is the part of P up to the first attaching point for the tethers of σ. By induction on genus we may assume G_σ is $\frac{1}{2}(g-(k+1)-3)$–connected, hence $\left(\frac{1}{2}(g-3)-k\right)$–connected since $k \geq 1$. The result for the case that P is a single interval then follows from Corollary 2.2(a).

Now we treat the case of a more general P consisting of several disjoint intervals. Let P_0 be one of these intervals. We will apply a link argument for $\mathrm{TCh}(S_{g,s}, P)$ and its subcomplex $\mathrm{TCh}(S_{g,s}, P_0)$. Define a k–simplex σ of $\mathrm{TCh}(S_{g,s}, P)$ to be *bad* if all of its tethers attach to points in $P - P_0$. Clearly the two conditions for badness are satisfied, and G_σ is $\mathrm{TCh}(S_\sigma, P_0)$ for S_σ the surface obtained by cutting $S_{g,s}$ along σ. We have shown that $\mathrm{TCh}(S_\sigma, P_0)$ is $\frac{1}{2}(g-k-1-3)$–connected, hence $\left(\frac{1}{2}(g-3)-k-1\right)$–connected. By Corollary 2.2(b), since $\mathrm{TCh}(S_{g,s}, P_0)$ is $\left(\frac{1}{2}(g-3)\right)$–connected, so is $\mathrm{TCh}(S_{g,s}, P)$. □

Finally we consider the complex $\mathrm{Ch}(S_{g,s})$ of oriented chains.

Proposition 5.6 $\mathrm{Ch}(S_{g,s})$ *is* $\frac{1}{2}(g-3)$*–connected.*

Proof We first treat the cases $s > 0$. Consider the complex $\mathrm{TCh}(S_{g,s}, P)$ with P a collection of disjoint open intervals in $\partial S_{g,s}$. We enlarge $\mathrm{TCh}(S_{g,s}, P)$ to a complex $\mathrm{TCh}(S_{g,s}, P^*)$ by allowing multiple tethers to each chain, all attaching at the same point of the b–curve of the chain and on the same side of the curve, each tether being otherwise disjoint from all other tethers and chains. There is a projection

$$\mathrm{TCh}(S_{g,s}, P^*) \to \mathrm{Ch}(S_{g,s}),$$

obtained by forgetting the tethers and orienting chains according to which side of the b–curves the tethers attach to. The fibers of this projection are contractible by the usual surgery argument, so it suffices to show that $\mathrm{TCh}(S_{g,s}, P^*)$ is $\frac{1}{2}(g-3)$–connected.

We do this by a link argument. The *bad* k–simplices σ in $\mathrm{TCh}(S_{g,s}, P^*)$ are those whose chains all have at least two tethers. The complex G_σ is the join of the complexes $\mathrm{TCh}(S_i, P_i)$ where cutting $(S_{g,s}, P)$ along σ produces a pair (S_σ, P_σ) with components (S_i, P_i). Cutting along a chain reduces genus by one and creates a new boundary circle, and then cutting along the first tether to the chain does not reduce the genus further, while cutting along each subsequent tether to the chain reduces genus by at most one more. Thus the total genus of S_σ is at least $g-k-1$. It follows as in the last paragraph of the proof of Proposition 5.1 that G_σ is $\frac{1}{2}(g-k-1-3)$–connected, hence $\big(\frac{1}{2}(g-3)-k-1\big)$–connected. Since $\mathrm{TCh}(S_{g,s}, P)$ is $\big(\frac{1}{2}(g-3)\big)$–connected by Proposition 5.5, we can apply Corollary 2.2(b) to deduce that $\mathrm{TCh}(S_{g,s}, P^*)$ has this connectivity as well. This proves the proposition when $s > 0$.

For the case $s = 0$ we use hyperbolic geometry. The cases $g \le 1$ are trivial, so we can assume $g \ge 2$ and fix a hyperbolic structure on $S_{g,0}$. Each nontrivial isotopy class of curves in $S_{g,0}$ contains a unique geodesic representative, and the geodesic representatives of two isotopy classes intersect the minimum number of times within the isotopy classes. Furthermore, if two curves intersect minimally, then one can choose isotopies to geodesics such that the number of intersections remains minimal throughout the isotopies. Thus each simplex in $\mathrm{Ch}(S_{g,0})$ has a unique geodesic representative.

There is a simplicial map $\mathrm{Ch}(S_{g,1}) \to \mathrm{Ch}(S_{g,0})$ induced by filling in $\partial S_{g,1}$ with a disk. Given a simplicial map $f\colon S^i \to \mathrm{Ch}(S_{s,0})$ we can choose a disk D in $S_{g,0}$ disjoint from the finitely many geodesic representatives for the chains that are images of vertices of S^i. Deleting the interior of D, we then have a lift of f to $\mathrm{Ch}(S_{g,1})$. This lift is nullhomotopic if $i \le \frac{1}{2}(g-3)$. Composing with the projection to $\mathrm{Ch}(S_{g,0})$ then gives a nullhomotopy of f, so $\mathrm{Ch}(S_{g,0})$ is $\frac{1}{2}(g-3)$–connected. □

Remark 5.7 One may ask whether the case $s = 0$ can be proved by a purely topological argument. In Proposition 4.7 this was done for the analogous projection $C(S_{g,1}) \to C(S_{g,0})$ by showing (in essence) that its fibers, which are one-dimensional when $g \ge 2$, are contractible. However, the fibers of $\mathrm{Ch}(S_{g,1}) \to \mathrm{Ch}(S_{g,0})$ are zero-dimensional and infinite when $g \ge 2$, so we cannot expect the same approach to work here.

Example 5.8 Consider the case that S is closed of genus 2, so $\mathrm{Ch}(S)$ is one-dimensional. A chain (a,b) in S has neighborhood bounded by a separating curve $c = c(a,b)$. The link of (a,b) in $\mathrm{Ch}(S)$ consists of all chains (a',b') in the genus one surface on the other side of c. These chains also have $c(a',b')=c$, so it follows that all chains (a,b) in this connected component of $\mathrm{Ch}(S)$ have the same curve $c(a,b)$. The connected components of $\mathrm{Ch}(S)$ thus correspond to nontrivial separating curves on S. Each connected component is the join of two copies of the infinite zero-dimensional complex $\mathrm{Ch}(S_{1,1})$. Thus $\mathrm{Ch}(S)$ is not homotopy equivalent to a wedge of spheres of a single dimension, in contrast with the situation for the curve complexes $C(S)$ and $C^0(S)$. Note also that the connectivity bound $\frac{1}{2}(g-3)$ is best possible in this case, where it asserts only that $\mathrm{Ch}(S)$ is nonempty.

Remark 5.9 As a partial generalization of this example one can say that for S a surface of arbitrary genus $g\geq 1$ the group $H_{g-1}(\mathrm{Ch}(S))$ is free abelian of infinite rank. This follows as in Remark 4.9 by embedding g disjoint copies of $S_{1,1}$ in S, which gives an embedding of the join of g copies of the infinite discrete set $\mathrm{Ch}(S_{1,1})$ in $\mathrm{Ch}(S)$. This join has dimension $g-1$, the same dimension as $\mathrm{Ch}(S)$, so the embedding of the join is injective on H_{g-1}.

6 The stability theorems

In the first part of this section we apply the spectral sequence for the action of the mapping class group of S on a double-tethered curve complex to prove homology stability with respect to genus when the number of boundary components is fixed and nonzero. As a bonus, the proof also shows that the stable homology groups do not depend on the number of boundary components of S when this number is nonzero.

After this we show that the homology in the case of closed surfaces is isomorphic to that for nonclosed surfaces, in the stable dimension range. This uses a second spectral sequence, this one for the action of the mapping class group on the complex $\langle\mathrm{Ch}(S)\rangle$ of ordered chains in S.

6.1 Surfaces with nonempty boundary

Let S be a surface of genus g with $s\geq 1$ boundary components, and let $M_{g,s}$ be the mapping class group of S, where diffeomorphisms and isotopies between them are required to restrict to the identity on each boundary circle. There are two stabilization maps $M_{g,s}\to M_{g+1,s}$ induced by inclusions $\alpha,\beta\colon S_{g,s}\to S_{g+1,s}$, shown in Figure 11. For α one attaches $S_{1,2}$ to $S_{g.s}$ along one boundary circle, assuming $s\geq 1$, while for β one attaches $S_{0,4}$ to $S_{g,s}$ along two boundary circles, assuming $s\geq 2$. The inclusions α and β induce homomorphisms of the corresponding mapping class groups by extending diffeomorphisms via the identity on the attached surface. It is a standard fact that these induced homomorphisms are injective; see for example Theorem 3.18 in [7].

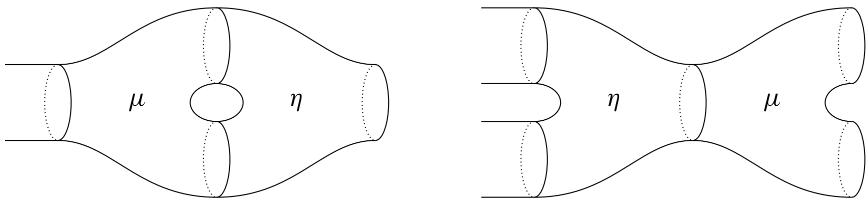

Figure 11: α (left) and β (right) stabilizations

We can factor both α and β as compositions of two inclusions μ and η, each attaching a pair of pants $S_{0,3}$, with the μ attachment being along one boundary circle of $S_{0,3}$ and η along two boundary circles. The difference between α and β is the order of the attachments: For α it is μ followed by η while for β it is the reverse.

For the associated mapping class groups we can stabilize with respect to g for fixed s by iterating α arbitrarily often, or we can do the same using β. The α stabilization is the one usually considered rather than β, probably because it is the more obvious stabilization and only requires $s \geq 1$. We can also iterate μ arbitrarily often to stabilize with respect to s for fixed g, but η can only be iterated finitely often so it is not exactly a stabilization.

Theorem 6.1 *The stabilizations*

$$\alpha_*\colon H_i(M_{g,s}) \to H_i(M_{g+1,s}), \quad \mu_*\colon H_i(M_{g,s}) \to H_i(M_{g,s+1})$$

are isomorphisms for $g \geq 2i+2$ *and* $s \geq 1$.

Proof We first give an easy argument that reduces both cases in the theorem to the statement that the stabilization $\beta_*\colon H_i(M_{g,s}) \to H_i(M_{g+1,s})$ is an isomorphism for $g > 2i+1$ and a surjection for $g = 2i+1$. Consider the three maps

$$H_i(M_{g,s}) \xrightarrow{\mu_*} H_i(M_{g,s+1}) \xrightarrow{\eta_*} H_i(M_{g+1,s}) \xrightarrow{\mu_*} H_i(M_{g+1,s+1}).$$

The composition of the first two maps is α_* and the composition of the second two is β_*. If β_* is surjective for $g \geq 2i+1$ then so is the second μ_* in that range. On the other hand μ_* is always injective since it has a left inverse obtained by attaching a disk to one of the free boundary circles of the attached $S_{0,3}$, so that the net result of the two attachments is attaching an annulus along one boundary circle, and this induces the identity map on $M_{g,s}$. Thus the second μ_* is an isomorphism for $g \geq 2i+1$. It follows that η_* is surjective for $g \geq 2i+1$. The first μ_* is now an isomorphism for $g \geq 2i+2$, so α_* is surjective in that range. For injectivity of α_*, if β_* is injective for $g \geq 2i+2$ then so is η_*, hence also α_* since μ_* is always injective.

To prove stability for β_* we will apply part (b) of Theorem 1.1 for the action of $M_{g,s}$ on the complex $\mathrm{DTC}^m(S_{g,s}, P, Q)$ of systems of double-tethered curves with matching orderings along P and Q, where P and Q are single intervals in different components of $\partial S_{g,s}$. Here we assume $s \geq 2$ in order for β to be defined. Since P and Q are single intervals we can use a slightly different, equivalent definition of $\mathrm{DTC}^m(S_{g,s}, P, Q)$ in which basepoints $x_1 \in P$ and $x_2 \in Q$ are chosen in advance and all tethers are required to have their P–endpoints at x_1 and their Q–endpoints at x_2, but otherwise satisfy the same conditions as before. Note that orderings of the tethers at x_1 and x_2 are still well-defined, specified by orientations of P and Q.

We need to check that the conditions (1)–(3) at the beginning of Section 1 hold for this action. First we check that the action is transitive on simplices of each dimension. The mapping class group clearly acts transitively on ordered coconnected systems of k oriented curves. To see that this holds also when matched systems of double tethers are added, we use the orientation on a curve determined by its tethers which specify a P–side of the curve and a Q–side, and we use the ordering of the curves specified by the ordering of the tethers at P and Q, which agree since we assume tethers satisfy the matching condition. Then transitivity on the double tethers for a fixed coconnected ordered oriented curve system can be seen by first cutting S along the curves in the system to get a surface F, and then observing that the mapping class group of F acts transitively on systems of k arcs in F joining x_1 to basepoints $p_1, \dots, p_k$ in the k ordered P–circles of $\partial F - \partial S$ together with k arcs joining x_2 to basepoints $q_1, \dots, q_k$ in the k ordered Q–circles of $\partial F - \partial S$, where in both cases the orderings of the arcs at x_1 and x_2 agree with the specified orderings of the circles at their other endpoints. This can be seen inductively by first making any two arcs from x_1 to p_1 agree after a diffeomorphism, then making any two arcs from x_1 to p_2 starting on the same side of the first arc at x_1 agree after a diffeomorphism fixing the first arc, etc.

To see that the inclusion of the stabilizer of a vertex σ into $M_{g,s}$ is the map induced by β, note first that a diffeomorphism in the stabilizer can be isotoped to fix the double-tethered curve pointwise, not just setwise, since it fixes x_1 and x_2. Then it can be isotoped to be the identity in a closed neighborhood N_σ of the union of σ and the components of ∂S containing x_1 and x_2 (see Figure 12). This N_σ is diffeomorphic to $S_{0,4}$ since it has Euler characteristic -2 and four boundary circles. Furthermore, N_σ attaches to the complementary surface S_σ along two circles of ∂N_σ. Thus the inclusion of the stabilizer of σ is the β stabilization. (If x_1 and x_2 were in the same component of ∂S, the neighborhood N_σ would be a copy of $S_{1,2}$ and the inclusion of the vertex stabilizer would be the α stabilization.)

More generally the inclusion of the stabilizer of a k–simplex σ is a $(k+1)$–fold iterate of β stabilizations since cutting S along each double-tethered curve of σ in succession

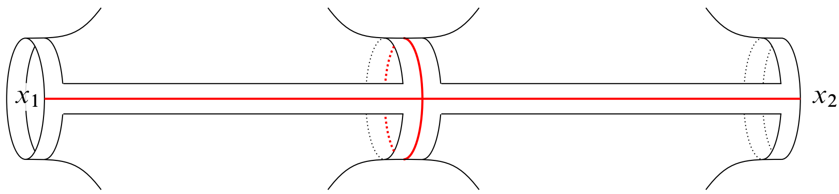

Figure 12: The neighborhood N_σ for a vertex σ of $\mathrm{DTC}(S)$

gives $k+1$ β stabilizations. All such inclusions of stabilizers of k–simplices into $M_{g,s}$ are conjugate since the action is transitive on k–simplices.

For condition (2), the stabilizer of a simplex fixes the simplex pointwise since the order of tethers at a basepoint cannot be permuted by an orientation-preserving diffeomorphism of the surface. To check condition (3) note that a neighborhood of two double-tethered curves defining an edge of $\mathrm{DTC}^m(S_{g,s})$, with the double tethers going from x_1 to x_2, is a copy of $S_{1,2}$ (this is not the same as the neighborhood N_σ in the preceding paragraph since we do not include neighborhoods of the boundary circles containing x_1 and x_2). The mapping class group $M_{1,2}$ acts transitively on vertices of $\mathrm{DTC}^m(S_{1,2})$, so there is a diffeomorphism of S supported in a neighborhood of the two given double-tethered curves that sends the first to the second, or vice versa. □

Remark If we choose P and Q to be disjoint intervals in the same component of ∂S, the same proof as above shows that the stabilization $\alpha_*\colon H_i(M_{g,s}) \to H_i(M_{g+1,s})$ is an isomorphism for $g \geq 2i+2$ and a surjection for $g = 2i+1$, a slight improvement over the preceding theorem. The advantage of using the β stabilization is that one also gets μ stability for free.

6.2 Closed surfaces

It remains to deal with the projection $M_{g,1} \to M_{g,0}$ induced by filling in the boundary circle of $S_{g,1}$ with a disk. More generally we will consider the map $\kappa\colon M_{g,s+1} \to M_{g,s}$ induced by capping off a boundary circle with a disk.

Theorem 6.2 *For each $s \geq 0$, the map $\kappa_*\colon H_i(M_{g,s+1}) \to H_i(M_{g,s})$ is an isomorphism for $g > 2i+3$ and a surjection for $g = 2i+3$.*

Proof We will apply Theorem 1.1(b) to the action of $M_{g,s}$ on $\langle \mathrm{Ch}(S_{g,s}) \rangle$, the complex of ordered systems of oriented chains. The orderings and orientations guarantee that the stabilizer of a k–simplex is exactly $M_{g-k-1,s+k+1}$. The inclusion of this stabilizer is the $(k+1)$–fold iterate of the composition $\lambda = \kappa\alpha$. We already know that α induces an isomorphism on homology in a stable range, so proving this for κ is equivalent to proving it for λ (and surjectivity of λ_* implies surjectivity of κ_*).

Conditions (1) and (2) for Theorem 1.1 are obviously satisfied. Unfortunately condition (3) fails for 1–simplices of $\langle \mathrm{Ch}(S_{g,s}) \rangle$ since there is no diffeomorphism of $S_{g,s}$ moving one chain onto another disjoint chain and supported in a neighborhood of the two chains. However there is a weakening of condition (3) that is satisfied and is strong enough to make the argument for injectivity of the differential $d\colon H_i(M_{g-1,s+1}) \to H_i(M_{g,s})$ still work. (The argument for surjectivity did not use (3).) If we enlarge the neighborhood of the two chains by adding a neighborhood of an arc joining them, producing a surface $T \subset S$ diffeomorphic to $S_{2,1}$, then there is a diffeomorphism supported in T interchanging the two chains and preserving their orientations. If we denote the two vertices of $\langle \mathrm{Ch}(S_{g,s}) \rangle$ corresponding to the two chains by v and w, with e either of the two edges joining them, then we have the following diagram:

$$\begin{tikzcd}
 & & H_i(\mathrm{stab}(v)) \arrow[d] \arrow[dr] & \\
H_i(\mathrm{stab}(T)) \arrow[r] \arrow[urr, bend left] \arrow[drr, bend right] & H_i(\mathrm{stab}(e)) \arrow[ur] \arrow[dr] & & H_i(\mathrm{stab}(v_0)) \\
 & & H_i(\mathrm{stab}(w)) \arrow[ur] &
\end{tikzcd}$$

The four triangles in the diagram commute, except possibly the one just to the left of the vertical arrow. Also, the large triangle formed by the two curved arrows and the vertical map commutes. The horizontal map is the μ_* stabilization in Theorem 6.1 so it is an isomorphism provided that $g-2 \geq 2i+2$, ie $g \geq 2i+4$. This implies that the whole diagram is in fact commutative in this range. This suffices to deduce injectivity of the differential $d = \lambda\colon H_i(M_{g-1,s+1}) \to H_i(M_{g,s})$ when $g > \varphi(i) = 2i + c$ for $c = 3$, recalling that $c \geq 2$ was sufficient in the original argument for Theorem 1.1(b).

It remains only to check that $\langle \mathrm{Ch}(S_{g,s}) \rangle$ is $\frac{1}{2}(g-3)$–connected, which will follow from Proposition 2.10 if $\mathrm{Ch}(S_{g,s})$ is wCM of level $\frac{1}{2}(g-1)$. We know that $\mathrm{Ch}(S_{g,s})$ is $\frac{1}{2}(g-3)$–connected by Proposition 5.6, and likewise the link of a k–simplex of $\mathrm{Ch}(S_{g,s})$ is $\frac{1}{2}(g-(k+1)-3)$–connected. We have $\frac{1}{2}(g-(k+1)-3) \geq \frac{1}{2}(g-1)-k-2$, so the result follows. □

Appendix

Here we give a different proof of Proposition 2.10 using the argument from Proposition 2.14 in [26]. The main step is a version of Theorem 2.4 of [8] which we give as a lemma below that is of interest in its own right. This gives conditions under which a simplicial map $f\colon Y \to X$ can be homotoped to be injective on individual simplices of some subdivision of Y. We in fact need a relative version of this in which f and the triangulation are kept fixed on a subcomplex $Z \subset Y$. In this case the best one could hope for is that f is *simplexwise injective relative to* Z, meaning that either of the following two equivalent conditions is satisfied:

- If an edge $[v, w]$ has $f(v) = f(w)$, then $[v, w] \in Z$.
- For each vertex v of $Y - Z$ we have $f(\mathrm{lk}(v)) \subset \mathrm{lk}(f(v))$.

Lemma A.1 *Let X be a wCM complex of level n, Y a finite simplicial complex of dimension at most n, $f_0\colon Y \to X$ a simplicial map and Z a subcomplex of Y. Then f_0 is homotopic fixing Z to a new map f_1 that is simplicial with respect to a new triangulation of Y subdividing the old one and unchanged on Z such that f_1 is simplexwise injective relative to Z in the new triangulation of Y.*

Proof The proof is by induction on n using a link argument, where the induction starts with the trivial case $n = 0$. Define a simplex σ of Y to be *bad* if for each vertex v of σ there is another vertex w of σ with $f_0(v) = f_0(w)$. Our goal is to eliminate all bad simplices that are not contained in Z, and in particular all bad edges not contained in Z.

If there are any bad simplices not contained in Z, let σ be one of maximal dimension k (note that $k > 0$ since vertices cannot be bad). Since f_0 is simplicial we have $f_0(\mathrm{lk}(\sigma)) \subset \mathrm{st}(f_0(\sigma))$, but by maximality of σ we actually have $f_0(\mathrm{lk}(\sigma)) \subset \mathrm{lk}(f_0(\sigma))$. (If $v \in \mathrm{lk}(\sigma)$ maps to $f_0(v) \in f_0(\sigma)$, then $\sigma * v$ is bad, contradicting maximality of σ.)

Since σ is bad, $f_0(\sigma)$ is a simplex of dimension at most $k-1$, hence $\mathrm{lk}(f_0(\sigma))$ is wCM of level $n-k$ since we assumed X is wCM of level n. We assumed also that Y has dimension at most n, so $\mathrm{lk}(\sigma)$ has dimension at most $n-k-1$. Therefore the map $f_0\colon \mathrm{lk}(\sigma) \to \mathrm{lk}(f_0(\sigma))$ is nullhomotopic and we can extend it to a map $g_0\colon b * \mathrm{lk}(\sigma) \to \mathrm{lk}(f_0(\sigma))$, where b is the barycenter of σ. Since $k > 0$ we have $n-k < n$ and we can apply induction to deform g_0 to a map g_1 which agrees with g_0 on $\mathrm{lk}\,\sigma$ and is simplexwise injective relative to $\mathrm{lk}(\sigma)$ in some subdivision of $b * \mathrm{lk}(\sigma)$ that is unchanged on $\mathrm{lk}(\sigma)$. This homotopy extends over $\mathrm{st}(\sigma) = \sigma * \mathrm{lk}(\sigma) = \partial\sigma * b * \mathrm{lk}(\sigma)$ by taking the join with the constant homotopy of $f_0\colon \partial\sigma \to f_0(\sigma)$. The resulting map $f_1\colon \mathrm{st}(\sigma) \to f_0(\sigma) * \mathrm{lk}(f_0(\sigma)) = \mathrm{st}(f_0(\sigma))$ is simplexwise injective relative to $\partial\sigma * \mathrm{lk}(\sigma)$. We now have two maps f_0 and f_1 from $\mathrm{st}(\sigma)$ to $\mathrm{st}(f_0(\sigma))$ which agree on $\partial\sigma * \mathrm{lk}(\sigma)$. Since $\mathrm{st}(f_0(\sigma))$ is contractible, these two maps are homotopic by a homotopy which fixes $\partial\sigma * \mathrm{lk}(\sigma)$. Extend this homotopy by the constant homotopy outside $\mathrm{st}(\sigma)$. The resulting map $f_1\colon Y \to X$ now has no bad simplices in the (subdivided) $\mathrm{st}(\sigma)$ except those in $\partial\sigma * \mathrm{lk}(\sigma)$ that were present before the modification. The process can now be repeated for other bad simplices of dimension k not contained in Z until they are all eliminated. Then we proceed to $(k-1)$–simplices, etc. $\square$

Remark The proof used the connectivity assumptions on links of simplices in X but not that X itself is $(n-1)$–connected. Thus we really only need local conditions on X, as one might expect.

We now give the alternative proof of Proposition 2.10. Recall that this states that if a simplicial complex X is wCM of level n then its ordered version $\langle X\rangle$ is $(n-1)$–connected. (It follows that $\langle X\rangle$ is wCM of level n as well, using the natural extension of this notion to semisimplicial complexes.)

By induction on n it suffices to show that a map $\partial D^n \to \langle X\rangle$ can be extended over D^n. We first show that any map $\partial D^n \to \langle X\rangle$ is homotopic to a simplicial map. We cannot directly appeal to the simplicial approximation theorem here because $\langle X\rangle$ is only semisimplicial. (The simplicial approximation theorem does generalize to the semisimplicial setting as shown in Theorem 5.1 of [21], but we do not need the full strength of this.)

Given $\langle f\rangle\colon \partial D^n \to \langle X\rangle$, let $f\colon \partial D^n \to X$ be the composition of $\langle f\rangle$ with the projection $p\colon \langle X\rangle \to X$. Since X is $(n-1)$–connected we can extend f to $F\colon D^n \to X$. Since X *is* a simplicial complex we can use the simplicial approximation theorem to get a homotopy from F to a map G that is simplicial in a PL triangulation of D^n subdividing any given triangulation. We assume that this homotopy is constructed as in the standard proof of simplicial approximation, in which case the restriction of the homotopy to ∂D^n lifts to a homotopy of $\langle f\rangle$. This is because the homotopy from F to G has the property that if $F(x)$ lies in the interior of a simplex σ of X, then $G(x)$ also lies in σ (possibly in $\partial\sigma$) and the homotopy just moves $F(x)$ along the linear path to $G(x)$ in σ. If x lies in ∂D^n then σ lifts to an ordered simplex containing $\langle f\rangle(x)$ and the path from $F(x)$ to $G(x)$ also lifts to this ordered simplex.

Since we are free to deform the original map $\langle f\rangle\colon \partial D^n \to \langle X\rangle$ by any homotopy before extending it over D^n, we may therefore assume that $\langle f\rangle$ is simplicial from the start and that we have a simplicial extension $F\colon D^n \to X$ of $f = p\langle f\rangle$.

We now apply the preceding lemma with $(Y, Z) = (D^n, \partial D^n)$ to obtain a new map $F\colon D^n \to X$ that is simplexwise injective on a subdivided triangulation of D^n, relative to ∂D^n. This F can be lifted to $\langle X\rangle$ in the following way. Choose a total ordering for the interior vertices of D^n. This total ordering gives an ordering on the vertices of each simplex τ which has no vertices in ∂D^n, and these orderings are compatible with passing to faces. Since F is injective on each such simplex τ the ordering on τ carries over to an ordering of $F(\tau)$ compatible with passing to faces. This gives a continuous lift $\langle F\rangle$ of F on interior simplices of D^n.

We already have a lift $\langle f\rangle$ of F on ∂D^n. It remains to lift F on the simplices of D^n that meet ∂D^n but are not contained in it. If every such simplex is the join of a boundary simplex σ with an interior simplex τ then the ordering on $f(\sigma)$ given by $\langle f\rangle$ extends to an ordering on $F(\sigma * \tau)$ by orienting each edge with one end in $f(\sigma)$ and the other in $F(\tau)$ towards $F(\tau)$, ie ordering all vertices of $f(\sigma)$ before all vertices of $F(\tau)$. This works since the lemma guarantees that no such edge is collapsed by F to a single vertex.

It is possible that simplices meeting ∂D^n are *not* joins of boundary and interior simplices, for example an edge passing through the interior of D^n might have both vertices in ∂D^n. To avoid this situation, note first that the join property is preserved under subdivision. If we start at the very beginning with a triangulation of D^n that has this property, for example by coning off a triangulation of ∂D^n, then the initial simplicial approximation step in the proof gives a subdivision of this triangulation, and applying the lemma produces a further subdivision.

References

[1] **V I Arnol'd**, *The cohomology ring of the group of the colored braid group*, Mat. Zametki 5 (1969) 227–231 MR In Russian; translated in Math. Notes 5 (1969) 138–140

[2] **V I Arnol'd**, *Certain topological invariants of algebraic functions*, Trudy Moskov. Mat. Obšč. 21 (1970) 27–46 MR In Russian; translated in Trans. Moscow Math. Soc. 21 (1970) 30–52

[3] **S K Boldsen**, *Improved homological stability for the mapping class group with integral or twisted coefficients*, Math. Z. 270 (2012) 297–329 MR

[4] **K S Brown**, *Cohomology of groups*, Graduate Texts in Mathematics 87, Springer (1982) MR

[5] **R M Charney**, *Homology stability for* GL_n *of a Dedekind domain*, Invent. Math. 56 (1980) 1–17 MR

[6] **R Charney**, **M W Davis**, *Finite* $K(\pi,1)$*s for Artin groups*, from "Prospects in topology" (F Quinn, editor), Ann. of Math. Stud. 138, Princeton Univ. Press (1995) 110–124 MR

[7] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Mathematical Series 49, Princeton Univ. Press (2012) MR

[8] **S Galatius**, **O Randal-Williams**, *Homological stability for moduli spaces of high dimensional manifolds, I*, preprint (2014) arXiv

[9] **J L Harer**, *Stability of the homology of the mapping class groups of orientable surfaces*, Ann. of Math. 121 (1985) 215–249 MR

[10] **J L Harer**, *The virtual cohomological dimension of the mapping class group of an orientable surface*, Invent. Math. 84 (1986) 157–176 MR

[11] **A Hatcher**, *On triangulations of surfaces*, Topology Appl. 40 (1991) 189–194 MR Updated version at http://www.math.cornell.edu/ hatcher/Papers/TriangSurf.pdf

[12] **A Hatcher**, *Homological stability for automorphism groups of free groups*, Comment. Math. Helv. 70 (1995) 39–62 MR

[13] **A Hatcher**, **K Vogtmann**, *Cerf theory for graphs*, J. London Math. Soc. 58 (1998) 633–655 MR

[14] **A Hatcher**, **K Vogtmann**, *Homology stability for outer automorphism groups of free groups*, Algebr. Geom. Topol. 4 (2004) 1253–1272 MR

[15] **A Hatcher**, **K Vogtmann**, **N Wahl**, *Erratum to* [14], Algebr. Geom. Topol. 6 (2006) 573–579 MR

[16] **A Hatcher**, **N Wahl**, *Stabilization for mapping class groups of* 3*–manifolds*, Duke Math. J. 155 (2010) 205–269 MR

[17] **W van der Kallen**, *Homology stability for linear groups*, Invent. Math. 60 (1980) 269–295 MR

[18] **A Putman**, **S V Sam**, *Representation stability and finite linear groups*, preprint (2014) arXiv

[19] **D Quillen**, *Higher algebraic* K*–theory, I*, from "Algebraic K–theory, I: Higher K–theories" (H Bass, editor), Lecture Notes in Math. 341, Springer (1973) 85–147 MR

[20] **O Randal-Williams**, *Resolutions of moduli spaces and homological stability*, J. Eur. Math. Soc. 18 (2016) 1–81 MR

[21] **C P Rourke**, **B J Sanderson**, Δ*–sets, I: Homotopy theory*, Quart. J. Math. Oxford Ser. 22 (1971) 321–338 MR

[22] **K Vogtmann**, *Homology stability for* $\mathrm{O}_{n,n}$, Comm. Algebra 7 (1979) 9–38 MR

[23] **J B Wagoner**, *Stability for homology of the general linear group of a local ring*, Topology 15 (1976) 417–423 MR

[24] **N Wahl**, *Homological stability for the mapping class groups of non-orientable surfaces*, Invent. Math. 171 (2008) 389–424 MR

[25] **N Wahl**, *Homological stability for mapping class groups of surfaces*, from "Handbook of moduli, III" (G Farkas, I Morrison, editors), Adv. Lect. Math. 26, International Press, Somerville, MA (2013) 547–583 MR

[26] **N Wahl**, **O Randal-Williams**, *Homological stability for automorphism groups*, preprint (2014) arXiv

[27] **M Weiss**, *What does the classifying space of a category classify?*, Homology Homotopy Appl. 7 (2005) 185–195 MR

Mathematics Department, Cornell University
Ithaca, NY 14853, United States

Mathematics Institute, University of Warwick
Coventry, CV4 7AL, United Kingdom

hatcher@math.cornell.edu, kvogtmann@gmail.com

https://www.math.cornell.edu/~hatcher/,
http://www2.warwick.ac.uk/fac/sci/maths/people/staff/karen_vogtmann/

Received: 1 November 2016 Revised: 23 January 2017

Guidelines for Authors

Submitting a paper to Algebraic & Geometric Topology

Papers must be submitted using the upload page at the AGT website. You will need to choose a suitable editor from the list of editors' interests and to supply MSC codes.

The normal language used by the journal is English. Articles written in other languages are acceptable, provided your chosen editor is comfortable with the language and you supply an additional English version of the abstract.

Preparing your article for Algebraic & Geometric Topology

At the time of submission you need only supply a PDF file. Once accepted for publication, the paper must be supplied in LaTeX, preferably using the journal's class file. More information on preparing articles in LaTeX for publication in AGT is available on the AGT website.

`arXiv` papers

If your paper has previously been deposited on the `arXiv`, we will need its `arXiv` number at acceptance time. This allows us to deposit the DOI of the published version on the paper's `arXiv` page.

References

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited at least once in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used, but will be converted to the house style (see a current issue for examples).

Figures

Fuzzy or sloppily drawn figures will not be accepted. For labeling figure elements consider the pinlabel LaTeXpackage, but other methods are fine if the result is editable. If you're not sure whether your figures are acceptable, check with graphics@msp.org.

Proofs

Page proofs will be made available to authors (or to the designated corresponding author) in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

ALGEBRAIC & GEOMETRIC TOPOLOGY

Volume 17 Issue 3 (pages 1283–1916) 2017

1472-2747(2017)17:3;1-T