BORDERED FLOER HOMOLOGY FOR MANIFOLDS WITH TORUS BOUNDARY VIA IMMERSED CURVES

JONATHAN HANSELMAN, JACOB RASMUSSEN, AND LIAM WATSON

ABSTRACT. This paper gives a geometric interpretation of bordered Heegaard Floer homology for manifolds with torus boundary. If M is such a manifold, we show that the type D structure $\widehat{CFD}(M)$ may be viewed as a set of immersed curves decorated with local systems in ∂M. These curves-with-decoration are invariants of the underlying three-manifold up to regular homotopy of the curves and isomorphism of the local systems. Given two such manifolds and a homeomorphism h between the boundary tori, the Heegaard Floer homology of the closed manifold obtained by gluing with h is obtained from the Lagrangian intersection Floer homology of the curve-sets. This machinery has several applications: We establish that the dimension of $\widehat{HF}$ decreases under a certain class of degree one maps (pinches) and we establish that the existence of an essential separating torus gives rise to a lower bound on the dimension of $\widehat{HF}$. In particular, it follows that a prime rational homology sphere Y with $\widehat{HF}(Y) < 5$ must be geometric. Other results include a new proof of Eftekhary's theorem that L-space homology spheres are atoroidal; a complete characterisation of toroidal L-spaces in terms of gluing data; and a proof of a conjecture of Hom, Lidman, and Vafaee on satellite L-space knots.

1. INTRODUCTION

Bordered Floer homology is a suite of invariants, introduced by Lipshitz, Ozsváth and Thurston [39], assigned to three-manifolds with boundary. These invariants are particularly well-adapted to cut-and-paste techniques. In the case of manifold with torus boundary, this theory has been developed in various ways yielding effective combinatorial tools for studying certain classes of toroidal three-manifolds [10, 20, 19, 21, 23, 24, 35]. The goal of this paper is to provide a geometric interpretation of bordered Floer homology for manifolds with torus boundary in terms of one-dimensional objects in the boundary of the manifold.

1.1. **Bordered invariants as immersed curves.** If M is a closed orientable 3-manifold with torus boundary, we define T_M to be the complement of 0 in $H_1(\partial M; \mathbb{R})/H_1(\partial M; \mathbb{Z})$. The punctured torus T_M can be identified with the complement of a point z in ∂M, and this identification is well-defined up to isotopy.

In order to define the bordered invariant $\widehat{CFD}(M, \alpha, \beta)$, we must choose a parametrization (α, β) of ∂M. That is, α and β are the cores of the 1-handles in a handle decomposition of ∂M. Starting from $\widehat{CFD}(M, \alpha, \beta)$ we will define a collection $\boldsymbol{\gamma} = \{\gamma^1, \dots, \gamma^n\}$ of immersed closed curves $\gamma^i \colon S^1 \looparrowright T_M$, each decorated with a local system (V_i, Φ_i) consisting of a finite dimensional vector space over $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$ and an automorphism $\Phi_i \colon V_i \to V_i$; we denote this data by $\widehat{HF}(M)$.

Theorem 1.1. *The collection of decorated curves $\widehat{HF}(M)$ is a well defined invariant of M, up to regular homotopy of curves and isomorphism of local systems. In particular, $\widehat{HF}(M)$ does not depend on the choice of parametrization (α, β).*

This data is equivalent to bordered Floer homology: $\widehat{HF}(M)$ is determined by $\widehat{CFD}(M, \alpha, \beta)$ and, for any choice of parametrization (α, β), $\widehat{CFD}(M, \alpha, \beta)$ can be recovered from $\widehat{HF}(M)$.

Now suppose that M_0 and M_1 are manifolds as above, and that $Y = M_0 \cup_h M_1$ is the closed manifold obtained by gluing them together by the orientation reversing homeomorphism $h \colon \partial M_1 \to \partial M_0$. The Heegaard Floer homology $\widehat{HF}(Y)$ can be recovered from $\gamma_0 = \widehat{HF}(M_0)$ and $\gamma_1 = \bar{h}(\widehat{HF}(M_1))$, where the homeomorphism $\bar{h}$ is the composition of h with the elliptic involution on ∂M_0. (In fact, the invariant $\widehat{HF}(M)$ is fixed by the elliptic involution on ∂M; this symmetry is established in the companion to this paper [22]. Thus $\bar{h}$ can be replaced by h in the definition of γ_1.)

Theorem 1.2. *For $Y = M_0 \cup_h M_1$ as above,*
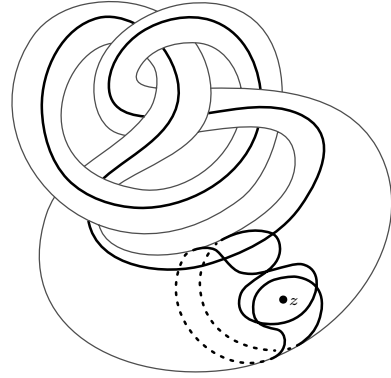
$$\widehat{HF}(Y) \cong HF(\gamma_0, \gamma_1)$$

where $HF(\cdot, \cdot)$ is an appropriately defined version of the Lagrangian intersection Floer homology in T_{M_0}.

1.2. **Trivial local systems.** In order to illustrate the content of Theorem 1.1 and Theorem 1.2, it is instructive to consider the case where the local systems are trivial. In this case the associated vector spaces are 1-dimensional and can be dropped from the notation; the resulting invariants are simply (systems of) immersed curves. As a very simple example, the invariant $\widehat{HF}(D^2 \times S^1)$ consists of a single closed circle parallel to the longitude $\lambda = \partial D^2 \times \{\text{pt}\}$.
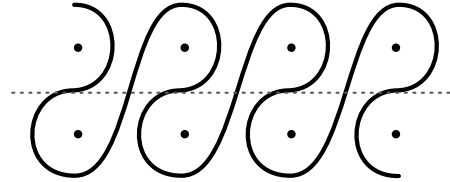
A more interesting example is illustrated in Figure 1: The invariant associated with the complement of the figure eight knot has two components. When $\widehat{HF}(M)$ is more complicated, it is generally easier to represent it by drawing some curves in the cover $\widetilde{T}_M = H_1(M; \mathbb{R}) \backslash H_1(M; \mathbb{Z}) \cong \mathbb{R}^2 \setminus \mathbb{Z}^2$ whose image in T_M is $\widehat{HF}(M)$. From this point of view, the effect of orientation reversal on bordered invariants corresponds to reflection in the line determined by the longitude. Figure 2 shows the invariant associated with the right-hand trefoil. We note that, in this context, calculation is extremely efficient:

Figure 1. The marked exterior of the figure eight knot, together with its bordered invariant as a pair of immersed curves.

Corollary 1.3. *In the case of trivial local systems, following the notation of Theorem 1.2, if no two components of γ_0 and γ_1 are parallel then $\dim \widehat{HF}(Y)$ is the minimal geometric intersection number between γ_0 and γ_1.*

For example, let Y_1 be the manifold obtained by splicing the complements of the left-hand trefoil K_L and the right-hand trefoil K_R; that is by taking $M_0 = S^3 \setminus \nu(K_L)$, $M_1 = S^3 \setminus \nu(K_R)$, and $h \colon \partial M_0 \to \partial M_1$ such that $h(\mu) = \lambda$ and $h(\lambda) = \mu$, where μ and λ are the preferred meridian and longitude of each trefoil in S^3. Similarly, let Y_2 be the manifold obtained by splicing two copies of the complement of the right-hand trefoil. Consulting Figure 3, we see that $\widehat{HF}(Y_1)$ has dimension 9, while $\widehat{HF}(Y_2)$ has dimension 7, as calculated by Hedden and Levine [24].

Figure 2. A curve for the right-hand trefoil: the horizontal direction corresponds to the preferred longitude λ, and the vertical direction to the standard meridian of the knot.
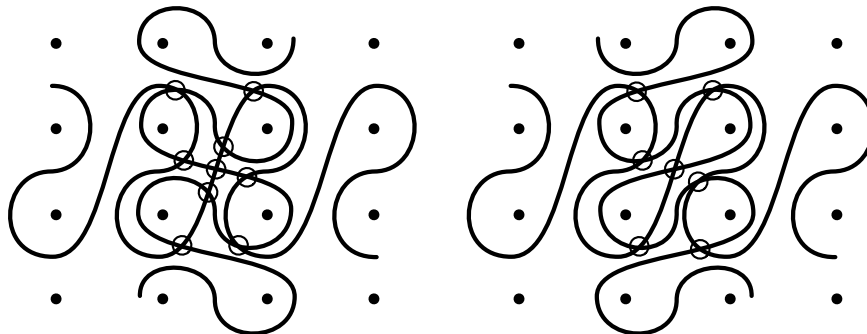
Figure 3. Splicing trefoils: The diagram on the left illustrates the intersection (carried out in $\widetilde{T}_M$) calculating $\dim \widehat{HF}(Y_1) = 9$ where Y_1 is the splice of a right-hand and left-hand trefoil, while the diagram on the right illustrates the intersection calculating $\dim \widehat{HF}(Y_2) = 7$ where Y_2 is the splice of two right-hand trefoils.

Manifolds for which the local systems are trivial are precisely the loop-type manifolds introduced by the first and third author [23]; the graphical formalism can thus be viewed as a geometric representation of the loop calculus. We remark that no explicit examples of three-manifolds are known for which the associated bordered invariant gives rise to a non-trivial local system. In practice—for instance for the gluing theorem needed in [21]—it is often enough to restrict attention to loop-type manifolds. As such, this class of manifolds seems interesting in its own right, and is discussed in further detail in a companion article [22].

1.3. **Train tracks and a structure theorem.** We now discuss the ideas behind the proof of Theorem 1.1. If M is as in the statement of the theorem, the bordered Floer homology $\widehat{CFD}(M, \alpha, \beta)$ is a type D structure (in the sense of Lipshitz, Ozsváth, and Thurston [39]) over an algebra $\mathcal{A}$, known as the *torus algebra*. In this case, a type D structure over $\mathcal{A}$ is simply a chain complex over $\mathcal{A}$ satisfying certain conditions. These are equivalent to projective differential modules over $\mathcal{A}$; see [39, Remark 2.25].

In this paper, we will restrict attention to a certain class of type D structures over $\mathcal{A}$, which we call *extendable*. The precise definition is given in section 3.1 but, briefly put, an *extended type D structure* is a curved differential module over a certain algebra $\widetilde{\mathcal{A}}$ that has $\mathcal{A}$ as a quotient; these have certain properties in common with matrix factorizations. A type D structure N is extendable if there is an extended type D structure $\widetilde{N}$ such that $N \cong \mathcal{A} \otimes_{\widetilde{\mathcal{A}}} \widetilde{N}$. The following theorem, which is essentially due to Lipshitz, Ozsváth and Thurston [39, Chapter 11], shows that this is not much of a restriction.

Theorem 1.4. *If M is a compact oriented manifold with torus boundary, $\widehat{CFD}(M, \alpha, \beta)$ is extendable.*

We introduce a new graphical calculus, based on immersed train tracks in the punctured torus, which describes extended type D structures. The graphical calculus provides an effective practical method for reducing a given extended type D structure to a collection of curves decorated with local systems. Using it, we prove the following structure theorem:

Theorem 1.5. *Every extendable type D structure over $\mathcal{A}$ can be represented by a collection of immersed curves in the punctured torus, decorated with local systems.*

The class of extendable type D structures known to arise as $\widehat{CFD}(M, \alpha, \beta)$ for some three-manifold M is considerably smaller than the set of all extendable type D structures. Indeed, as mentioned above, we do not have any explicit examples of a manifold M for which $\widehat{HF}(M)$ has a non-trivial local system (though we expect that such M should exist). Even in the case of trivial local systems there

are curve sets that do not correspond to three-manifold invariants. For instance, Proposition 7.13 shows that certain configurations of *loose* curves (introduced in Section 7.2) do not arise as invariants of three-manifolds. This improves on a result due to Gillespie [15], and should be compared with work of Alishahi and Lipshitz [2].

1.4. **Relation with the Fukaya category.** Theorem 1.1 can be interpreted as saying that if M is a manifold with torus boundary then $\widehat{HF}(M)$ is a compactly supported element of the Fukaya category $\mathcal{F}(T_M)$, while Theorem 1.2 says that the Floer homology of a closed manifold $Y = M_0 \cup_h M_1$ is given by the Hom pairing in the Fukaya category:

$$\widehat{HF}(Y) = \mathrm{Hom}\left(\widehat{HF}(M_0), \bar{h}(\widehat{HF}(M_1))\right)$$

The connection between the Heegaard Floer theory and the Fukaya category dates back to the introduction of Heegaard Floer homology, whose definition was motivated by a Seiberg-Witten analog of the Atiyah-Floer conjecture [42]. Following the introduction of bordered Floer homology [37, 39], Auroux [3, 4] and Lekili and Perutz [34] suggested that the Floer homology of a compact oriented M with $\partial M = \Sigma_g$ should be an object in $\mathcal{F}(\mathrm{Sym}^g(\Sigma_g \setminus z))$. In particular, Auroux showed that if $\mathcal{P}$ is a parametrization of ∂M (in other words, an identification of ∂M with a specific genus g surface F equipped with a handle body decomposition), the bordered Floer homology $\widehat{CFA}(M, \mathcal{P})$ defines an object $L(M, \mathcal{P})$ in the partially wrapped Fukaya category $\mathcal{F}(\mathrm{Sym}^g(F \setminus z))$, and that the pairing in bordered Floer homology is given by the Hom pairing. However it is unclear from this construction whether $L(M, \mathcal{P})$ should be compactly supported.

Auroux's construction is conceptually very useful, but for general g we don't have any way of making calculations in $\mathcal{F}(\mathrm{Sym}^g(F \setminus z))$ other than the one provided by bordered Floer homology. The one obvious exception is the case $g = 1$ where, naively, one might expect that objects in $\mathcal{F}(T^2 \setminus z)$ are given by curves in the punctured torus. In fact, the situation is somewhat more complicated. The Fukaya category is triangulated, so a typical object actually has the form of an iterated mapping cone built out of geometric curves. In [18], Haiden, Katzarkov and Kontsevich give an especially nice algebraic model for the partially wrapped Fukaya category of a punctured surface Σ; in the case of $T^2 \setminus z$, objects of $\mathcal{F}(T^2 \setminus z)$ correspond to chain complexes over the algebra $\mathcal{A}$ or, equivalently, with type D structures.

One of the main results of [18] is a structure theorem for objects of $\mathcal{F}(\Sigma)$, which says that any object can be expressed as a direct sum of immersed, possibly noncompact curves. Theorem 1.5 was motivated by this result but our proof is quite different. In particular, the condition that the type D structure is extendable implies that the curves are all compactly supported. More generally, if Σ is a punctured surface with stops on each boundary component, objects of $\mathcal{F}(\Sigma)$ correspond to type D structures over an appropriate algebra. Our proof of Theorem 1.5 generalizes to show that if the type D structure is extendable, it is isomorphic to a disjoint union of compactly supported curves equipped with local systems. This gives a new proof of the structure theorem of Haiden, Katzarkov, and Kontsevich in the compactly supported case (see Theorem 3.17), which is constructive and stays internal to the language of bordered Floer homology (namely, type D structures). This constructive approach has a key advantage: we obtain new information about Heegaard Floer theory and applications to 3-manifolds, as described below.

1.5. **Gradings.** There is a refined version of the invariant that takes spin^c structures and the absolute $\mathbb{Z}/2\mathbb{Z}$ grading on Heegaard Floer homology into account.

Definition 1.6. If M is as in Theorem 1.1, let $\overline{T}_M$ be the covering space of T_M whose fundamental group is the kernel of the composition

$$\pi_1(T_M) \to \pi_1(\partial M) \to H_1(\partial M) \to H_1(M)$$

Equivalently, if $\lambda \in H_1(\partial M; \mathbb{Z})$ generates the kernel of the inclusion $j_*: H_1(\partial M; \mathbb{Z}) \to H_1(M; \mathbb{Z})$ $\overline{T}_M$ is homeomorphic to the quotient $((H_1(\partial M; \mathbb{R}) \setminus H_1(\partial M; \mathbb{Z}))) / \langle \lambda \rangle$. We let $p: \overline{T}_M \to T_M$ be the covering map.

The set of spin^c structures on M can be identified with $H^2(M) \simeq H_1(M, \partial M) \simeq \operatorname{coker} j_*$. We have:

Theorem 1.7. *For each $\mathfrak{s} \in \operatorname{Spin}^c(M)$ there is an invariant $\widehat{HF}(M, \mathfrak{s})$, which is a collection of oriented immersed closed curves equipped with local systems in $\overline{T}_M$. Moreover, $\widehat{HF}(M, \mathfrak{s})$ is well-defined up to the action of translation by the deck group of p, and*

$$\widehat{HF}(M) = \bigcup_{\mathfrak{s} \in \operatorname{Spin}^c(M)} p(\widehat{HF}(M, \mathfrak{s})).$$

There is an analog of Theorem 1.2 recovering the spin^c decomposition of $\widehat{HF}(M_0 \cup_h M_1)$ from $\widehat{HF}(M_0, \mathfrak{s}_0)$ and $\widehat{HF}(M_1, \mathfrak{s}_1)$, where $\mathfrak{s}_i$ runs over spin^c structures on M_i. (See Proposition 6.7 for a precise statement.) The $\mathbb{Z}/2\mathbb{Z}$ grading on $\widehat{HF}(M_0 \cup_h M_1)$ is determined by the sign of the intersections between the oriented curves $\widehat{HF}(M_0)$ and $\widehat{HF}(M_1)$.

1.6. **Immediate consequences.** The geometric interpretation of bordered Floer invariants described above has several applications. For instance:

Theorem 1.8. *Let Y be a closed, orientable three-manifold. If Y contains a separating essential torus then $\dim \widehat{HF}(Y) \geq 5$.*

This follows quickly from Theorem 1.2. Indeed, a simple geometric argument shows that any two sufficiently non-trivial immersed curves in the punctured torus intersect in at least 5 points (a specific example realizing $\dim \widehat{HF}(Y) = 5$ is illustrated in Figure 60; see also Theorem 7.20). Theorem 1.8 gives rise to an interesting geometric statement:

Corollary 1.9. *If Y is a prime rational homology sphere with $\dim \widehat{HF}(Y) < 5$ then Y is geometric.*

Proof. Note that any essential torus in a rational homology sphere must be separating. Thus, if $\dim \widehat{HF}(Y) < 5$ then Y is atoroidal and we may appeal to Perelman's resolution of Thurston's geometrization conjecture to conclude that Y admits a geometric structure. □

Note that, more precisely, the geometric structure in question is either hyperbolic or it is one of 6 Seifert fibered geometries; see Scott [48], for example. As another immediate corollary of Theorem 1.8, we obtain a new proof of a theorem of Eftekhary [12]:

Corollary 1.10. *L-space integer homology spheres are atoroidal.* □

Recall that an L-space is a rational homology sphere for which $\dim \widehat{HF}(Y) = |H_1(Y)|$. Corollary 1.10 is a part of a conjecture of Ozsváth and Szabó; it implies that prime integer homology sphere L-spaces with infinite fundamental group—should examples exist—are hyperbolic.

We also obtain a statement about Khovanov homology.

Corollary 1.11. *If L is a link in S^3 containing an essential Conway sphere then $\dim \widetilde{Kh}(L) \geq 5$.*

Proof. Denoting by Σ_L the two-fold branched cover of S^3, branched along the link L, the presence of an essential Conway sphere in L is equivalent to the existence of an essential separating torus in Σ_L. Thus $\dim \widetilde{Kh}(L) \geq \widehat{HF}(\Sigma_L) \geq 5$, where the first inequality arises from the spectral sequence from the reduced Khovanov homology of (the mirror of) L to Heegaard Floer homology [44] and the second applies Theorem 1.8. □

It is worth comparing this result with the fact that the unknot [33] and the two trefoils [7] are characterized by their Khovanov homology. For statements about links, see [25, 26].

In another direction, we can establish strong behaviour for Heegaard Floer homology in the presence of certain degree one maps. Consider an integer homology sphere Y containing an essential torus, and write $Y = M_0 \cup_h M_1$ where the M_i are (necessarily) integer homology solid tori. Following our conventions above, $h(\lambda)$ must be a meridian in ∂M_0 when λ is the longitude of M_1. The Dehn filling $Y_0 = M_0(h(\lambda))$ is the result of replacing M_1 with a solid torus (called a *pinch*); there is a degree one map $Y \to Y_0$.

Theorem 1.12. *If Y is a toroidal integer homology sphere and Y_0 is the result of a pinch on Y then* $\dim \widehat{HF}(Y) \geq \dim \widehat{HF}(Y_0)$.

This follows from a more general statement about pinching along tori in rational homology spheres; see Theorem 7.15. In particular, to the best of the authors' knowledge, this is the only result concretely treating the following:

Question 1.13. *If $Y \to Y_0$ is a degree one map between closed, connected, orientable three-manifolds, is it the case that* $\dim \widehat{HF}(Y) \geq \dim \widehat{HF}(Y_0)$?

1.7. **The L-space gluing theorem.** The geometric invariants defined in this paper can also be applied to the classification of L-spaces; in particular, we give a complete characterization of when gluing along a torus produces an L-space. Again, an L-space is a rational homology sphere Y for which $\dim \widehat{HF}(Y) = |H_1(Y)|$. For a manifold with torus boundary M, the set of L-space slopes of M is given by

$$\mathcal{L}_M = \{\alpha \,|\, M(\alpha) \text{ is an L-space}\} \subset \mathbb{Q}P^1$$

where $M(\alpha)$ denotes Dehn filling along the slope α. Denote the interior of $\mathcal{L}_M$ by $\mathcal{L}_M^\circ$.

The set $\mathcal{L}_M$ is encoded by and easily extracted from the invariant $\widehat{HF}(M)$. In particular, a necessary condition for $|\mathcal{L}_M| > 1$ is that the associated local system be trivial. The complement of $\mathcal{L}_M$ is obtained by considering the minimal set of tangent lines to $\widehat{HF}(M)$; see Theorem 7.17 for a precise statement. This provides a satisfying solution to a problem posed by Boyer and Clay [9, Problem 1.9]. Toroidal L-spaces are then characterized as follows.

Theorem 1.14. *Let $Y = M_0 \cup_h M_1$ be a 3-manifold where M_i are boundary incompressible manifolds and $h \colon \partial M_1 \to \partial M_0$ is an orientation reversing homeomorphism between torus boundaries. Then Y is an L-space if and only if* $\mathcal{L}_{M_0}^\circ \cup h(\mathcal{L}_{M_1}^\circ) = \mathbb{Q}P^1$.

This confirms a conjecture due to the first author [20], and strengthens results in [19, 21, 23, 24, 46]. In addition to the applications discussed below, the L-space gluing theorem plays a key role in the program to understand L-spaces arising as cyclic branched covers of knots in the three-sphere; see Gordon and Lidman [16, 17] and Boileau, Boyer, and Gordon [8]. A weaker version of the L-space gluing theorem was the key ingredient, along with work of Boyer and Clay [9], in proving the L-space conjecture for graph manifolds [21] (an alternate, constructive, proof is given in [47]). We note that, in light of the L-space conjecture [10], the L-space gluing theorem makes some striking predictions about the behaviour of foliations on manifolds with torus boundary, as well as the behaviour of left-orders on the fundamental groups of these manifolds. Another notable application is given in the work of Némethi on links of rational singularities [41].

Theorem 1.14 allows us to refine Theorem 1.8 in the case that Y is a toroidal L-space. In addition to ruling out $|H_1(Y)| \leq 4$, we can easily enumerate all examples with $|H_1(Y)| < 7$; see Theorem 7.20. Note that this leads to another proof of Corollary 1.10. Another quick consequence of Theorem 1.14 is the following:

Theorem 1.15. *Suppose that K is a satellite knot in S^3. If K is an L-space knot (that is, K admits non-trivial L-space surgeries) then both the pattern knot and the companion knot are L-space knots as well.*

This confirms a conjecture of Hom, Lidman, and Vafaee [28, Conjecture 1.7]. It has subsequently been used by Baker and Motegi [6] to show that the pattern knot must be a braid in the solid torus (see also Hom [27]).

Organization. This paper is a substantially revised version of our earlier preprint of the same name, which dealt only with loop-type manifolds. A subsequent paper [22] will discuss further properties of the invariant and give some examples. The current paper is laid out as follows.

Section 2 summarizes the relevant background from bordered Floer homology (in the case of manifolds with torus boundary) and describes a geometric representation for these invariants in terms of immersed train tracks. The main new content in this section is a graphical interpretation of the box tensor product—the chain complex $\widehat{CFA}(M_0, \alpha_0, \beta_0) \boxtimes \widehat{CFD}(M_1, \alpha_1, \beta_1)$—in terms of intersection between train tracks; see Theorem 2.2. For train tracks that are immersed curves, this has a clear connection to Lagrangian intersection Floer theory. In general, however, the naive intersection of train tracks is not an invariant in any sense. Extracting invariants is the principal aim of the remaining sections.

Section 3 is devoted to algebraic issues concerning a well-behaved class of type D structures: we prove a structure theorem (Theorem 1.5) for extendable type D structures in two steps. We first reduce train tracks in this class to train tracks satisfying certain nice properties (Section 3.3, particularly Proposition 3.8), and then show that such train tracks can be manipulated and ultimately interpreted in terms of immersed curves and local systems (Section 3.6, particularly Proposition 3.13). Indeed, the class of train tracks considered ultimately gives a geometric interpretation of the relevant local systems. The section concludes with Theorem 3.17, which extends our techniques to surfaces of higher genus.

Section 4 establishes the equivalence between, and independence of choices made in the construction of, extendable type D structures and immersed curved with local systems. We note that, while only the existence of an extension is required for our purposes, it follows from the work in this section that extendable type D structures have essentially unique extensions; see Proposition 4.16.

Section 5 returns the focus to three-manifolds, completing the proof of Theorem 1.1 and Theorem 1.2, while Section 6 explains the modifications to the invariant needed to recover the spin^c and $\mathbb{Z}/2\mathbb{Z}$ gradings on $\widehat{HF}$, as described in Theorem 1.7.

Section 7 is devoted to the applications described above, and the paper concludes with a short Appendix containing the proof of Theorem 1.4.

the referee for many thoughtful comments and suggestions, which led to improvements at a number of key points. Part of this work was carried out while the third author was visiting Montréal as CIRGET research fellow, and part was carried out while the second and third authors were participants in the program *Homology Theories in Low Dimensions* at the Isaac Newton Institute.

2. Bordered invariants as train tracks

This section gives a geometric interpretation of bordered invariants in terms of immersed train tracks for manifolds with torus boundary. We begin with a brief review of the relevant notions from bordered Floer homology [39]. A less terse introduction, with essentially the same notation used here, is given in [23].

2.1. **Modules over the torus algebra.** Let M be an orientable three-manifold with torus boundary, and choose oriented essential simple closed curves α, β in ∂M with $\beta \cdot \alpha = 1$. The ordered triple (M, α, β) is called a bordered three-manifold; the pair (α, β) may be regarded as a parametrization of the torus boundary, in the sense that $\langle \alpha, \beta \rangle$ specifies the peripheral subgroup $\pi_1(\partial M)$.

We focus on two (equivalent) bordered invariants assigned to a bordered manifold (M, α, β). These will take the form of certain modules over the torus algebra $\mathcal{A}$. Various descriptions of this algebra are given by Lipshitz, Ozsváth and Thurston [39], but for our purposes, recall that $\mathcal{A}$ is generated, as an algebra over the two-element field $\mathbb{F}$, by elements ρ_1, ρ_2 and ρ_3 and idempotents ι_0 and ι_1. Multiplication in $\mathcal{A}$ is described by the quiver depicted in Figure 4 together with the relations $\rho_2\rho_1 = \rho_3\rho_2 = 0$. The shorthand $\rho_{12} = \rho_1\rho_2$, $\rho_{23} = \rho_2\rho_3$, $\rho_{123} = \rho_1\rho_2\rho_3$ is standard. Let $\mathcal{I}$ denote the subring of idempotents, and write $\mathbf{1} = \iota_0 + \iota_1$ for the unit in $\mathcal{A}$.
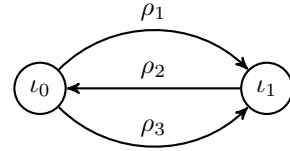
Figure 4. A quiver for the torus algebra, with relations $\rho_2\rho_1 = \rho_3\rho_2 = 0$.

The relevant bordered invariants are $\widehat{CFA}(M, \alpha, \beta)$ and $\widehat{CFD}(M, \alpha, \beta)$; both are invariants of the underlying bordered manifold up to homotopy [39]. The former is a type A structure, that is, a right A_∞-module over $\mathcal{A}$. The latter is a type D structure, which consists of (1) a vector space V (over $\mathbb{F}$) together with a splitting as a direct sum over a left action of the idempotents $V \cong \iota_0 V \oplus \iota_1 V$; and (2) a map $\delta^1 \colon V \to \mathcal{A} \otimes_{\mathcal{I}} V$ satisfying a compatibility condition ensuring that $\partial(a \otimes x) = a \cdot \delta(x)$ is a differential on $\mathcal{A} \otimes_{\mathcal{I}} V$ (with left $\mathcal{A}$-action specified by $a \cdot (b \otimes x) = ab \otimes x$). In particular, $\mathcal{A} \otimes_{\mathcal{I}} V$ has the structure of a left differential module over $\mathcal{A}$. We will sometimes confuse the type D structure $\widehat{CFD}(M, \alpha, \beta)$ (consisting of V and δ^1) and this associated differential module. Given any type D structure there is a collection of recursively defined maps $\delta^k \colon V \to \mathcal{A}^{\otimes k} \otimes V$ where $\delta^0 \colon V \to V$ is the identity and $\delta^k = (\mathrm{id}_{\mathcal{A}^{\otimes k-1}} \otimes \delta^1) \circ \delta^{k-1}$. The type D structure $\widehat{CFD}(M, \alpha, \beta)$ is bounded if δ^k vanishes for all sufficiently large integers k.

Given a bordered manifold (M, α, β) the associated type A and type D structures are related. In particular, according to [38, Corollary 1.1], $\widehat{CFA}(M, \alpha, \beta)$ is dual (in an appropriate sense) to $\widehat{CFD}(M, \alpha, \beta)$. However, the utility of the two structures, taken together, is a computable model for the A_∞ tensor product: Given bordered manifolds (M_0, α_0, β_0) and (M_1, α_1, β_1) consider the chain complex $\widehat{CFA}(M_0, \alpha_0, \beta_0) \boxtimes \widehat{CFD}(M_1, \alpha_1, \beta_1)$ obtained from $\widehat{CFA}(M_0, \alpha_0, \beta_0) \otimes_{\mathcal{I}} \widehat{CFD}(M_1, \alpha_1, \beta_1)$ with differential defined by

$$\partial^\boxtimes(x \otimes y) = \sum_{k=0}^{\infty} (m_{k+1} \otimes \mathrm{id})(x \otimes \delta^k(y))$$

and the requirement that $\widehat{CFD}(M_1, \alpha_1, \beta_1)$ is bounded (this ensures that the above sum is finite). Then the pairing theorem of Lipshitz, Ozsváth and Thurston asserts that the homology of

$\widehat{CFA}(M_0, \alpha_0, \beta_0) \boxtimes \widehat{CFD}(M_1, \alpha_1, \beta_1)$ is isomorphic to the Heegaard Floer homology $\widehat{HF}(M_0 \cup_h M_1)$ of the closed manifold obtained from the homeomorphism $h\colon \partial M_1 \to \partial M_0$ specified by $h(\alpha_1) = \beta_0$ and $h(\beta_1) = \alpha_0$ [39, Theorem 1.3].

2.2. **Conventions for bordered manifolds.** In the framework given by Lipshitz, Ozsváth and Thurston [39], the boundary parametrization of M is recorded by specifying a diffeomorphism $\phi\colon F \to \partial M$ where F is a torus with a fixed handle decomposition. The image of the cores of the one-handles in F correspond to the pair of α-arcs (α_1^a, α_2^a) in a bordered Heegaard diagram; completing this pair to curves in ∂M gives a pair corresponding to our parametrizing curves α and β. To relate the two notations, we must specify which curve corresponds to each α-arc in a bordered Heegaard diagram. We orient the arcs α_1^a and α_2^a so that, starting at the basepoint and following the boundary of the Heegaard surface, we pass the initial point of α_1^a, the initial point of α_2^a, the final point of α_1^a, and the final point of α_2^a. With this labelling, α in our notation corresponds to $-\alpha_1^a$ and β corresponds to α_2^a. Note that when gluing two bordered manifolds together by an orientation reversing diffeomorphism taking basepoint to basepoint, α_1^a must glue to $-\alpha_2^a$ and α_2^a must glue to $-\alpha_1^a$; thus in our notation, α glues to β and β glues to α. Finally, the arcs defined above in a Heegaard diagram seem to satisfy $(-\alpha_1) \cdot \alpha_2 = 1$, which would imply that $\alpha \cdot \beta = 1$. However, the orientation of the Heegaard surface agrees with the opposite of the orientation on ∂M. Said another way, the surface in a bordered Heegaard diagram for M should be interpreted as being viewed from *inside M*, but we will generally look at ∂M from *outside M* so that $\beta \cdot \alpha = 1$.

2.3. **Decorated graphs.** There is a convenient graph-theoretic shorthand for describing these bordered invariants (see, for instance, [23, 24]). Let Γ be a directed graph with vertex set $\mathcal{V}_\Gamma$ and edge set $\mathcal{E}_\Gamma$. Suppose that every $v \in \mathcal{V}_\Gamma$ is labelled with exactly one element from $\{\bullet, \circ\}$ and every $e \in \mathcal{E}_\Gamma$ is labelled with exactly one element from $\{\varnothing, 1, 2, 3, 12, 23, 123\}$. Moreover, suppose the edge labels are compatible with the vertex labels so that the only arrangements that occur are $\bullet \xrightarrow{\varnothing} \bullet$, $\circ \xrightarrow{\varnothing} \circ$, $\bullet \xrightarrow{1} \circ$, $\circ \xrightarrow{2} \bullet$, $\bullet \xrightarrow{3} \circ$, $\bullet \xrightarrow{12} \bullet$, $\circ \xrightarrow{23} \circ$, and $\bullet \xrightarrow{123} \circ$. Finally, we will require that for any pair of vertices v_1 and v_2 in a decorated graph and any element $I \in \{\varnothing, 1, 2, 3, 12, 23, 123\}$, there is an even number of length two directed paths from v_1 to v_2 for which the concatenation of the two edge labels is I. We will call a graph of this form an $\mathcal{A}$-*decorated graph*.

An $\mathcal{A}$-decorated graph Γ describes a type D structure over $\mathcal{A}$ as follows. The underlying vector space is generated by $\mathcal{V}_\Gamma$, with the idempotent splitting specified by $\bullet$ labels identifying the ι_0 generators and $\circ$ labels identifying the ι_1 generators. The map δ^1 is determined by the edge labels: Given $e \in \mathcal{E}_\Gamma$ with label $I \in \{\varnothing, 1, 2, 3, 12, 23, 123\}$ consider the source x and target y in $\mathcal{V}_\Gamma$. Then $\rho_I \otimes y$ is a summand of $\delta^1(x)$, with the interpretation that $\rho_\varnothing = \mathbf{1}$. A decorated graph (and its associated type D structure) is *reduced* if none of the edges is labelled by $\varnothing$. An example is shown in Figure 5, describing the bordered invariant $\widehat{CFD}(M, \mu, \lambda)$ when M is the complement of the right-hand trefoil, μ is the knot merid-

Figure 5. A decorated graph.

ian and λ is the Seifert longitude (see [39]). Notice that the higher δ^k are determined by directed paths in Γ; for the example shown, there exist generators x and y such that $\delta^3(x) = \rho_3 \otimes \rho_2 \otimes \rho_1 \otimes y$ corresponding to the unique directed path of length 3. Clearly, the associated type D structure is bounded if and only if the decorated graph contains no directed cycles. Note that the assumption that length two paths cancel mod 2 implies that $\partial^2 = 0$ in the corresponding differential module.
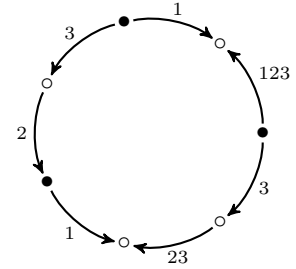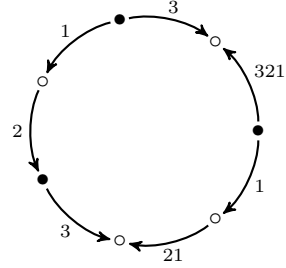
In the reduced case, the duality between type D and type A structures is encoded in the $\mathcal{A}$-decorated graphs following an algorithm described by Hedden and Levine [24]. This takes the same interpretation for the underlying vector space (that is, the idempotent splitting according to vertex labels) but requires a different interpretation of the edge labels. First, one re-writes/re-interprets the edge labels according to the bijection $1 \leftrightarrow 3$, $2 \leftrightarrow 2$, $3 \leftrightarrow 1$. Next, given any length n directed path in Γ with source vertex x and target vertex y we construct a sequence $I = I_1, \dots, I_k$ and assign a multiplication $m_{k+1}(x \otimes \rho_{I_1} \otimes \cdots \otimes \rho_{I_k}) = y$. The sequence I is constructed by forming a word in $\{1, 2, 3\}$ determined according to the labels of the directed edge read in order, and then regrouping to find the minimum k such that each I_j (for $1 \le j \le k$) is an element

Figure 6. Relabelling by $1 \leftrightarrow 3$, $2 \leftrightarrow 2$, $3 \leftrightarrow 1$ to extract a type A structure from a decorated graph.

of $\{1, 2, 3, 12, 23, 123\}$. Thus, in our example, to the length 3 directed path we assign the label sequence $I = \{123\}$ so that $m_2(x, \rho_{123}) = y$ while the edge label 321 (formerly 123 in the original decorated graph) gives rise to a sequence $I = \{3, 2, 1\}$ and the product $m_4(x', \rho_3, \rho_2, \rho_1) = y'$.

2.4. **Train tracks.** Given a reduced $\mathcal{A}$-decorated graph Γ as in the previous section, we can immerse Γ in the torus, as described below. First, we fix a specific model for the punctured torus.

Definition 2.1. The *marked torus T* is the oriented surface $T = \mathbb{R}^2/\mathbb{Z}^2$ punctured at $z = (1-\epsilon, 1-\epsilon)$. The images of the y and x-axes in T will be referred to as α and β respectively.

We embed the vertices of Γ into T so that the vertices corresponding to $V_0 = \iota_0 V$ generators (the • vertices) are distinct points on α and the vertices corresponding to $V_1 = \iota_1 V$ generators (the ○ vertices) are distinct points on β. We then embed each edge in T according to its label, as shown in Figure 8. While each edge is embedded, different edges may intersect so the result is an immersion of Γ. We will denote the image of this immersion by $\boldsymbol{\vartheta}_\Gamma$. Note that the order in which vertices are embedded along α and β is arbitrary, and the arcs in Figure 8 may be replaced by any homotopic path, so $\boldsymbol{\vartheta}_\Gamma$ is defined up to regular homotopy. A concrete example is shown in Figure 9.
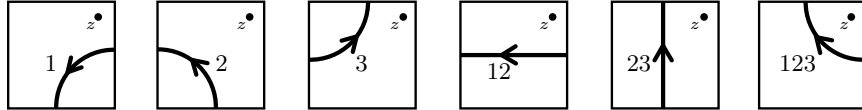
Figure 7. Assigning edges in a reduced decorated graph Γ to directed edges in the marked torus T. Notice that, by labelling corners of the (bordered) marked torus clockwise from the base point, the labels and orientations on edges may be suppressed without ambiguity.

.

We require that (1) all intersections between edges are transverse and away from the parametrizing curves α and β and (2) near a vertex, all edges are orthogonal to the curve (α or β) that the vertex lies on. This gives $\boldsymbol{\vartheta}_\Gamma$ the structure of an immersed train track (following Thurston's notion of a train track in a surface, see for example [40]). An immersed train track is an immersed graph for which all edges incident at a vertex share a common tangent line. Following the usual terminology (inspired by railroad junctions), the vertices of an immersed train track are referred to as *switches*. An important point is that, just as Γ need not be connected as a graph, the immersed train track $\boldsymbol{\vartheta}_\Gamma$, constituting equivalent data, may consist of a union of a collection of immersed train tracks.

A train track $\boldsymbol{\vartheta}_\Gamma$ constructed as above has a very particular form; for instance, it has no switches that are not on α or β. We will later have need to work with a more general class of immersed train tracks in T (see Definition 3.2), and we will call $\boldsymbol{\vartheta}_\Gamma$ as constructed here a *naive train track* representing Γ. We remark that the edge orientations may be dropped from $\boldsymbol{\vartheta}_\Gamma$ without losing information, since

they are determined by the rule that the basepoint lies on the right side of each arc, and thus $\boldsymbol{\vartheta}_\Gamma$ may be viewed as an unoriented train track. However, orientations will be meaningful for more general train tracks and we find it convenient to record them even in this case.
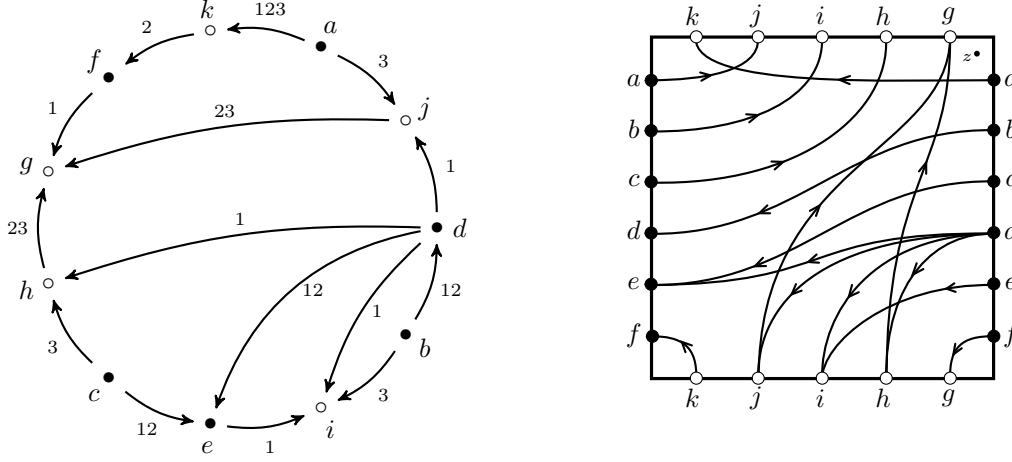
**Figure 8.** The train track $\boldsymbol{\vartheta}_\Gamma$ (right) associated with the $\mathcal{A}$-decorated graph Γ shown on the left. All train track switches occur at the vertices on the α and β curves; intersection points between edges of the graph are not (new) vertices of the graph. Note that the type D structure described is not known to come from a three-manifold.

There is an interesting special case in which $\boldsymbol{\vartheta}_\Gamma$ (ignoring edge orientations) is simply a collection of immersed curves in T. This corresponds to the case when Γ is a valence 2 graph and satisfies an extendability condition ensuring that the two arcs out of the switch of $\boldsymbol{\vartheta}_\Gamma$ associated with a given vertex leave the switch in opposite directions. The extendibility condition always holds for $\widehat{CFD}(M, \alpha, \beta)$ (see Theorem 1.4); in the case where $\widehat{CFD}(M, \alpha, \beta)$ admits a representative that may be described by a valence 2 decorated graph, M is called loop-type; compare [23]. As an example, the trefoil exterior is loop-type; see Figure 5. It is a surprising fact that a great many classes of manifolds with torus boundary are loop-type, including any manifold admitting multiple L-space fillings [21, 46]; this class of manifolds is considered in detail in a companion paper [22]. Indeed, we have no explicit example of a manifold that is not loop-type.

2.5. **Pairing train tracks.** Just as an $\mathcal{A}$-decorated graph encodes both a type D structure and a type A structure over $\mathcal{A}$, we now establish our conventions for train tracks with respect to this duality. Let $\boldsymbol{\vartheta}$ be a train track that is a naive train track representing some reduced $\mathcal{A}$-decorated graph. Fix a standard marked torus T and divide it into four quadrants. The *type A realization* $A(\boldsymbol{\vartheta})$ is obtained by including $\boldsymbol{\vartheta}$ (cut along α and β) into the first quadrant and adding unoriented horizontal edges in the second quadrant and unoriented vertical edges in the fourth quadrant connecting the two resulting copies of each switch; see Figure 10. Similarly, the *type D realization* $D(\boldsymbol{\vartheta})$ is obtained by reflecting $\boldsymbol{\vartheta}$ across the anti-diagonal $y = -x$, including into the third quadrant, and adding unoriented vertical (resp. horizontal) edges in the second (resp. fourth) quadrant. In other words, $D(\boldsymbol{\vartheta})$ is the reflection of $A(\boldsymbol{\vartheta})$ across $y = -x$. We will refer to the edges of $A(\boldsymbol{\vartheta})$ or $D(\boldsymbol{\vartheta})$ in the first or third quadrants as *corners*, since these are where paths carried by the train tracks can change direction, and the edges in the second and fourth quadrant are *horizontal edges* or *vertical edges*.

Now consider a pair of naive train tracks $\boldsymbol{\vartheta}_0$ and $\boldsymbol{\vartheta}_1$. We define $\mathcal{C}(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1)$ to be the vector space over $\mathbb{F}$ generated by the intersection points between $A(\boldsymbol{\vartheta}_0)$ and $D(\boldsymbol{\vartheta}_1)$. Note that these intersection points are confined to the second and fourth quadrants in T, by construction. We will define a linear map $d^{\boldsymbol{\vartheta}} \colon \mathcal{C}(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1) \to \mathcal{C}(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1)$ by counting bigons.
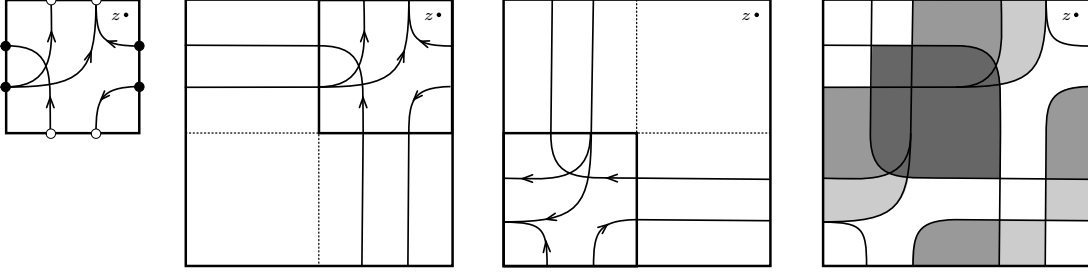
Figure 9. From left to right: a sample train track $\boldsymbol{\vartheta}$, its type A realization $A(\boldsymbol{\vartheta})$, its type D realization $D(\boldsymbol{\vartheta})$, and the pairing complex $\mathcal{C}(\boldsymbol{\vartheta}, \boldsymbol{\vartheta})$ between the two. The 5 bigons contributing to the map $d^{\boldsymbol{\vartheta}}$ in this case have been shaded.

If x and y are two intersection points between $A(\boldsymbol{\vartheta}_0)$ and $D(\boldsymbol{\vartheta}_1)$, we consider the set of continuous maps $u \colon D^2 \to T$ satisfying the following conditions (compare with [1, Definition 2.8]):

- u is an orientation-preserving immersion away from $-i$ and i;
- $u(-i) = x$ and $u(i) = y$ (viewing D^2 as the unit disk in $\mathbb{C}$);
- u maps the positive real part of ∂D^2 (oriented from $-i$ to i) to an oriented path in $A(\boldsymbol{\vartheta}_0)$;
- u maps the negative real part of ∂D^2 (oriented from $-i$ to i) to an oriented path in $D(\boldsymbol{\vartheta}_1)$;
- both x and y are convex corners of the disk u; and
- the basepoint $z \in T$ is not in the image of u.

We say two such maps u, u' are equivalent if there is an orientation preserving diffeomorphism $\phi \colon D^2 \to D^2$ with $u' = u \circ \phi$ and $\phi(\pm i) = \pm i$. Assuming the set of equivalence classes of such maps is finite, let $n(x, y)$ denote the number of equivalence classes modulo 2. (We will see below that by making a small modification of $\boldsymbol{\vartheta}_1$, we can arrange for the set of equivalence classes to be finite for all intersection points x, y.) We then define

$$d^{\boldsymbol{\vartheta}}(x) = \sum_{y \in A(\boldsymbol{\vartheta}_0) \cap D(\boldsymbol{\vartheta}_1)} n(x, y)\, y.$$

We remark that an oriented path in $A(\boldsymbol{\vartheta}_0)$ or $D(\boldsymbol{\vartheta}_1)$ may traverse the horizontal and vertical edges in either direction, since we view these as unoriented edges. In fact, just as we may drop the edge orientations on naive train tracks, we can also drop the requirement that paths from $-i$ to i in ∂D^2 map to oriented paths in $A(\boldsymbol{\vartheta}_0)$ and $D(\boldsymbol{\vartheta}_1)$ in the present setting that $\boldsymbol{\vartheta}_0$ and $\boldsymbol{\vartheta}_1$ are naive train tracks. Indeed, the fact that every edge in $\boldsymbol{\vartheta}_0$ or $\boldsymbol{\vartheta}_1$ has the basepoint on its right implies that for any immersed disk not covering z the paths determined by the boundary have the correct orientations. However, we state this requirement as it is relevant for generalizing this pairing to more general train tracks.

Under mild hypotheses, the vector space $\mathcal{C}(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1)$ with linear map $d^{\boldsymbol{\vartheta}}$ is, in fact, a chain complex: it may be identified with the box tensor product of the associated bordered invariants. To explain this, we need to introduce some variants of our train tracks (this will coincide with a notion of admissibility when we return our focus to three-manifold invariants later). Recall that, to this point, we have assumed our train tracks come from reduced $\mathcal{A}$-decorated graphs (i.e graphs where no edges are labelled with $\varnothing$). Since we require a bounded type D structure for the box tensor product to make sense, we will need to relax the reduced assumption slightly.

We will say an $\mathcal{A}$-deorated graph is *almost reduced* if the only edges labelled $\varnothing$ begin and end on valence 2 vertices and occur in one of precisely four configurations as shown in Figure 11. Note that each of these configurations can be replaced by a single arrow labelled 12, 23, 123, or 123, respectively, without changing the corresponding type D structure (up to homotopy). We can
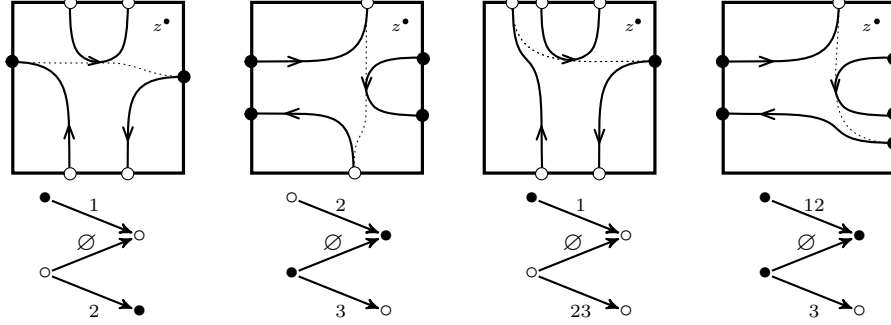
Figure 10. Modifications permitted in almost reduced train tracks. These local moves are useful for destroying oriented cycles and achieving admissibility in the proof of Theorem 2.2.

extend our naive train track construction to almost reduced $\mathcal{A}$-decorated graphs by embedding the edges labeled by $\varnothing$ as an edge starting and ending on the same side of the square $T \setminus (\alpha \cup \beta)$, as shown in the figure. Recall that an $\mathcal{A}$-decorated graph is bounded if it contains no oriented cycle. It is clear that we can replace any reduced graph by a bounded almost reduced graph representing a type D structure homotopy equivalent to the original by replacing some number of 12, 23, or 123 arrows by the corresponding zig-zag configuration. A train track will be called *special bounded* if the underlying $\mathcal{A}$-decorated graph is both bounded and almost reduced.

Theorem 2.2. *Fix train tracks $\boldsymbol{\vartheta}_0$ and $\boldsymbol{\vartheta}_1$, where $\boldsymbol{\vartheta}_0$ is reduced and $\boldsymbol{\vartheta}_1$ is special bounded. Let $N_{\mathcal{A}}^0$ denote the type A structure associated with $\boldsymbol{\vartheta}_0$ and let $^{\mathcal{A}}N^1$ denote the type D structure associated with $\boldsymbol{\vartheta}_1$. Then $d^{\boldsymbol{\vartheta}}$ may be identified with $\partial^{\boxtimes}$ and there is an isomorphism of chain complexes*

$$\mathcal{C}(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1) \cong N_{\mathcal{A}}^0 \boxtimes {}^{\mathcal{A}}N^1$$

Remark 2.3. With the exception of the proof that follows, we will typically opt for subscripts on the given type D structures and make use of the shorthand notation $N_{\mathcal{A}}^0 \boxtimes {}^{\mathcal{A}}N^1 = N_0^A \boxtimes N_1$.

Proof of Theorem 2.2. First observe that there is an identification of the generating sets for $\mathcal{C}(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1)$ and $N_{\mathcal{A}}^0 \boxtimes {}^{\mathcal{A}}N^1$, since each vertical (respectively horizontal) segment in $A(\boldsymbol{\vartheta}_0)$ intersects each horizontal (respectively vertical) segment of $D(\boldsymbol{\vartheta}_1)$ and there are no other intersection points. Note that the horizontal segments of $A(\boldsymbol{\vartheta}_0)$ correspond directly to the ι_0 generators of $N_{\mathcal{A}}^0$, while the vertical segments correspond to the ι_1 generators. Similarly, the vertical segments of $D(\boldsymbol{\vartheta}_1)$ correspond directly to the ι_0 generators of $^{\mathcal{A}}N^1$, while the horizontal segments correspond to the ι_1 generators.

Since $\boldsymbol{\vartheta}_1$ is bounded, the associated type D structure $^{\mathcal{A}}N^1$ is bounded and hence $\partial^{\boxtimes}$ gives a well-defined differential on $N_{\mathcal{A}}^0 \boxtimes {}^{\mathcal{A}}N^1$. We will identify the linear map $d^{\boldsymbol{\vartheta}}$ with the differential $\partial^{\boxtimes}$ by showing that that each immersed bigon contributing to $d^{\boldsymbol{\vartheta}}$ is uniquely associated with a pairing between type D and type A operations contributing to $\partial^{\boxtimes}$.

We first aim to classify the immersed bigons that contribute to $d^{\boldsymbol{\vartheta}}$ by decomposing them into pieces. Observe that by cutting the torus T along the curves α and β, any immersed bigon connecting intersection points p to q decomposes into embedded pieces. We will show that the only possible pieces are those shown in Figure 12. We find it helpful to keep track of the type of corners on each boundary of the bigon (our convention for doing so will be explained shortly), and the dotted boxes in the top right and lower left quadrants in the figures are meant to highlight these corners. The top right quadrant will be called the corner region for $A(\boldsymbol{\vartheta}_0)$ and the lower left quadrant is the corner region for $D(\boldsymbol{\vartheta}_1)$.

The first row of the figure shows pieces arising from bigons which involve a $\varnothing$-labelled corner in $D(\boldsymbol{\vartheta}_1)$ (these are corners that start and end on the same side of the corner region, corresponding to $\varnothing$-labelled edges in the $\mathcal{A}$-decorated graph associated to $\boldsymbol{\vartheta}_1$); we will call such bigons *small bigons*. Small bigons are always cut into exactly two pieces; we explain this for $\varnothing$-labelled corners connecting

Pieces of small bigons:
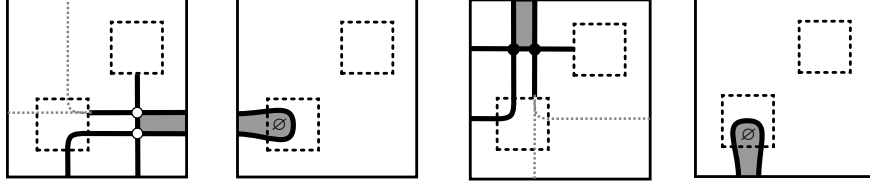
Starting pieces:
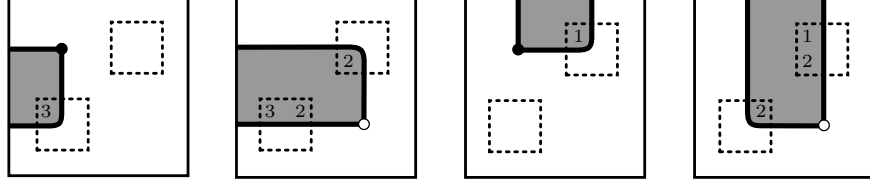
Ending pieces:
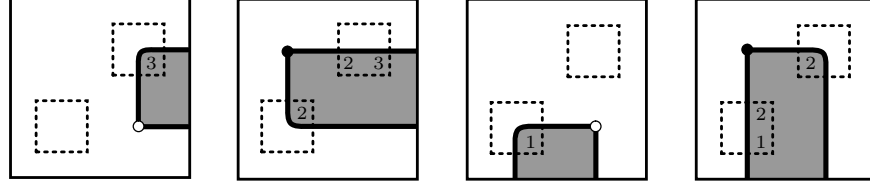
Connecting pieces:
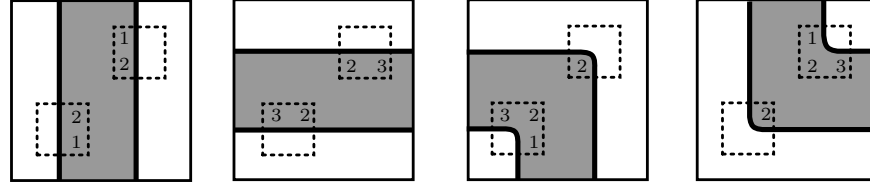
Single piece bigon:
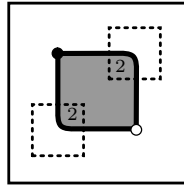
Figure 11. Decomposing a bigon into tiles: Any small bigon carried by $(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1)$ consists of exactly two pieces of either the first two types or the second two types in the first row. Any other bigon either is disjoint from $\alpha \cup \beta$ and has the form shown in the bottom row, or it decomposes into one starting piece, some number of connecting pieces, and one ending piece.

two horizontal edges in $D(\boldsymbol{\vartheta}_1)$, and the case of a $\varnothing$ corner connecting two vertical edges is similar. The $\varnothing$ corner in question connects the right ends of two horizontal segments corresponding to two ∘-generators in the type D structure corresponding to $\boldsymbol{\vartheta}_1$, and because $\boldsymbol{\vartheta}_1$ is almost reduced these generators appear in either the first or third zig-zag arrangement in Figure 11. At the left end of these horizontal segments, the lower one has a corner turning downward (corresponding to the 1 labelled arrow in the decorated graph) and the upper one either continues leftward or turns upward (corresponding to a 23 or 2 labelled arrow). It is clear that the bigon can not extend past the left

end of these horizontal edges, since it would then necessarily cover the basepoint, and so it must stop in the lower right quadrant of T at some vertical edge of $A(\boldsymbol{\vartheta}_0)$.

For the remaining bigons, we will assume they do not contain any corners of type $\varnothing$. Consider a bigon from an intersection point p to an intersection point q. The only bigons that do not intersect α and β and thus do not decompose into multiple pieces have the form shown in the last row of Figure 12; they begin at p in the bottom right quadrant of T and end at q in the top left quadrant, and each side of the boundary has a single corner labelled by 2. Apart from bigons of this type, the piece of the bigon containing the initial generator p must look like one of the *starting pieces* in the second row of Figure 12. For example, if p is in the lower quadrant then the bigon must open up and left from p, since a bigon opening down and right would cover the basepoint. The $A(\boldsymbol{\vartheta}_0)$ boundary must either pass straight through its corner region (the top right quadrant) or turn left, since turning right would again lead to the basepoint being covered. If the $A(\boldsymbol{\vartheta}_0)$ continues upward, then the $D(\boldsymbol{\vartheta}_1)$ boundary must turn upward in its corner region to avoid covering the basepoint, leading to the fourth starting piece in Figure 12, and if the $A(\boldsymbol{\vartheta}_0)$ boundary turns left then the $D(\boldsymbol{\vartheta}_1)$ boundary must continue leftward through its corner region, leading to the second starting piece shown. If p lies in the top left quadrant then the bigon can either open down and left or up and right from p; in the former case the $D(\boldsymbol{\vartheta}_1)$ boundary must turn leftward and in the latter case the $A(\boldsymbol{\vartheta}_0)$ boundary must turn upwards, producing starting pieces of the first and third types. Note that the type D and type A boundaries can only change direction in their respective corner regions, and backtracking is not possible since $\boldsymbol{\vartheta}_0$ is reduced and we assume the $\varnothing$-labelled edges of $D(\boldsymbol{\vartheta}_1)$ are not involved. Similar consideration shows that the piece of a bigon containing the terminal generator q takes the form of one of the *ending pieces* in Figure 12, and all other pieces must be one of the *connecting pieces* shown in the figure.

In order to identify $d^{\boldsymbol{\vartheta}}$ with $\partial^{\boxtimes}$, we will relate sequences of bigon pieces to type A operations and sequences of type D operations which pair in the box tensor product. Toward that end, we now describe our convention for labeling corners of bigon pieces. Recall that each corner in $A(\boldsymbol{\vartheta}_0)$ or $D(\boldsymbol{\vartheta}_0)$ represents an arrow in the corresponding $\mathcal{A}$-decorated graph. Referring to Figure 8 (and remembering that $\boldsymbol{\vartheta}_1$ is reflected across the antidiagonal when constructing $D(\boldsymbol{\vartheta}_1)$), we label the corners of $D(\boldsymbol{\vartheta}_1)$ the same as the corresponding edge. The label on a corner of $A(\boldsymbol{\vartheta}_0)$ is the subword of 321 obtained from the arrow labels in the corresponding $\mathcal{A}$-decorated graph by interchanging 3's and 1's. An equivalent graphical way of describing these labels is to label the corners of each corner region 0, 1, 2, and 3, ordered counterclockwise starting from the top right corner for the corner region of $A(\boldsymbol{\vartheta}_0)$ and from the bottom left corner for the corner region of $D(\boldsymbol{\vartheta}_1)$, and then label each corner edge in the boundary of a bigon by the sequence of these labels which are covered (in the order obtained by following the boundary from p to q). Note that the path from p to q in $A(\boldsymbol{\vartheta}_0)$ follows the boundary orientation of the bigon and thus has the basepoint to its right at each corner, while the path from p to q in $D(\boldsymbol{\vartheta}_1)$ opposes the boundary orientation and has the basepoint to its left at each corner. It follows that traversing either path from p to q agrees with the orientation of the edges in the corresponding $\mathcal{A}$-decorated graphs, and thus each boundary path from p to q determines a directed path in the $\mathcal{A}$-decorated graphs. Finally, recall that a directed path in an $\mathcal{A}$-decorated graph defines either a δ^k map in the associated type D structure or an m_k map in associated $\mathcal{A}_\infty$ module.

For a given bigon from p to q, let x denote the horizontal or vertical edge of $A(\boldsymbol{\vartheta}_0)$ containing p and let x' denote the horizontal or vertical edge containing q; similarly, let y and y' denote the edges of $D(\boldsymbol{\vartheta}_1)$ containing p and q. Let $\tilde{x}$, $\tilde{x}'$, $\tilde{y}$, and $\tilde{y}'$ denote the corresponding generators of $N_{\mathcal{A}}^0$ or $^{\mathcal{A}}N^1$, so that p corresponds to $\tilde{x} \otimes \tilde{y}$ and q corresponds to $\tilde{x}' \otimes \tilde{y}'$. Let $I_1, \ldots, I_k$ is the sequence of corner labels along the $D(\boldsymbol{\vartheta}_1)$ part of the boundary of the bigon, viewed as a path from y to y'. This path encodes an operation

$$\delta^k(\tilde{y}) = \rho_{I_1} \otimes \cdots \otimes \rho_{I_k} \otimes \tilde{y}'$$

in $^{\mathcal{A}}N^1$. Let $J_1, \ldots, J_\ell$ be the sequence of corner labels along the $A(\boldsymbol{\vartheta}_0)$ part of the boundary, viewed as a path from x to x'. To determine the corresponding $\mathcal{A}_\infty$ operation, note that 1's and 3's are

already swapped relative to labels in the $\mathcal{A}$-decorated graph, so by the duality described in Section 2.3 it just remains to concatenate the corner labels and divide into maximal increasing subwords. That is, let $I_1', \ldots, I_{k'}'$ be the sequence such that $I_1' \cdots I_{k'}' = J_1 \cdots J_\ell$, each I_i' is a subword of 123, and k' is minimal with respect to these conditions. Then the path from x to x' in $A(\boldsymbol{\vartheta}_0)$ encodes an operation

$$\mu^{k'+1}(\tilde{x}, \rho_{I_1'}, \cdots, \rho_{I_{k'}'}) = \tilde{x}'$$

in $N_{\mathcal{A}}^0$.

A key observation is that for any bigon, the sequences of corner labels from each half of the boundary label have the same concatenation—that is, $I_1 \cdots I_k = J_1 \cdots J_\ell$. To see this, observe that for each piece in Figure 12, the label on the $D(\boldsymbol{\vartheta}_1)$ corner can be obtained from the label on the $A(\boldsymbol{\vartheta}_0)$ corner by removing a 3 at the beginning if the piece has boundary on the right side of the square, adding a 3 at the end if the piece has boundary on the left side of the square, removing a 1 at the end if the piece has boundary on the top side of the square, and adding a 1 at the beginning if the piece has boundary on the bottom of the square. Note that for any bigon formed from these pieces, excluding small bigons, both boundary arcs move only upward and leftward when traveling from x to y, so a piece with a boundary on the left side of the square is always followed by a piece with boundary on the right side of the square, and a piece with boundary on the top is always followed by a piece with boundary on the bottom. It follows that the changes made to obtain the $D(\boldsymbol{\vartheta}_1)$ labels from the $A(\boldsymbol{\vartheta}_0)$ labels cancel out when concatenating all labels along the boundary paths from x to y. For example, for the bigon in Figure 13 the concatenation of the corner labels along either boundary path is the word 123. Since $I_1 \cdots I_k = J_1 \cdots J_\ell$, and $\{I_1, \cdots, I_k\}$ is a sequence of subwords of 123 with the first letter of I_i less than the last letter of I_{i-1}, we have immediately that the sequences $\{I_1, \ldots, I_k\}$ and $\{I_1', \ldots, I_{k'}'\}$ agree. From this it follows that the operations discussed in the preceding paragraph pair in the box tensor product to produce a $\tilde{x}' \otimes \tilde{y}'$ term in $\partial^{\boxtimes}(\tilde{x} \otimes \tilde{y})$.

We have shown that every non-small bigon encodes a term in $\partial^{\boxtimes}$. It is easy to check the same is true for small bigons. In this case, p and q lie on the same horizontal or vertical edge of $A(\boldsymbol{\vartheta}_0)$, so $x = x'$, and the $\varnothing$ corner on $D(\boldsymbol{\vartheta}_1)$ encodes an operation $\delta^1(\tilde{y}) = \mathbf{1} \otimes \tilde{y}'$ in $^{\mathcal{A}}N^1$. Such a term in δ^1 contributes $\tilde{x} \otimes \tilde{y}'$ to $\partial^{\boxtimes}(\tilde{x} \otimes \tilde{y})$, as desired.

It remains to show that every term in $\partial^{\boxtimes}$ corresponds to a bigon in the way described above. First consider terms in $\partial^{\boxtimes}(x \otimes y)$ arising from differentials of the form $\delta^1(y) = \mathbf{1} \otimes y'$ in $^{\mathcal{A}}N^1$. For each such term, $x \otimes y'$ appears in $\partial^{\boxtimes}(x \otimes y)$ for each generator x of $N_{\mathcal{A}}^0$ with the appropriate idempotent. It is clear that each of these terms correspond to a small bigon, since the horizontal (or vertical) edges of $D(\boldsymbol{\vartheta}_1)$ corresponding to y and y' intersect each vertical (or horizontal) edge of $A(\boldsymbol{\vartheta}_0)$ and are connected by a $\varnothing$-labelled corner on one end. All other terms in $\partial^{\boxtimes}(x \otimes y)$ arise from pairing two operations of the form

$$\mu^{k+1}(\tilde{x}, \rho_{I_1}, \cdots, \rho_{I_k}) = \tilde{x}' \qquad \text{and} \qquad \delta^k(\tilde{y}) = \rho_{I_1} \otimes \cdots \otimes \rho_{I_k} \otimes \tilde{y}'$$

for some sequence of algebra elements $\rho_{I_1}, \cdots, \rho_{I_k}$. The presence of this δ^k map in $^{\mathcal{A}}N_1$ implies the existence of a path in $D(\boldsymbol{\vartheta}_1)$ from y to y' with k corners of types $I_1, \ldots, I_k$. Similarly, the presence of this m^{k+1} operation in $N_{\mathcal{A}}^0$ implies the existence of a path in $A(\boldsymbol{\vartheta}_0)$ from x to x' with sequence of corner labels $\{J_1, \ldots, J_\ell\}$ obtained by dividing $I_1 \cdots I_k$ into a minimal number of subwords of 321. We need to show that these two paths must bound a bigon. It is clear that there is some path consisting of horizontal edges in the top left quadrant, vertical edges in the bottom right quadrant, and corners in the top right quadrant that would form a bigon with the given path in $D(\boldsymbol{\vartheta}_1)$—such a path could be obtained by applying a homotopy to the path in $D(\boldsymbol{\vartheta}_1)$ avoiding the basepoint to achieve the desired form. By the discussion above, this path must have sequence of corner labels $\{J_1, \ldots, J_\ell\}$, and any path in $A(\boldsymbol{\vartheta}_0)$ with the same sequence of corners would also bound a bigon with the path in $D(\boldsymbol{\vartheta}_1)$. $\square$
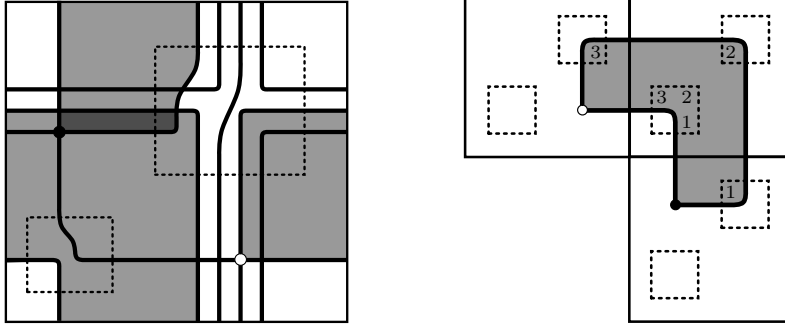
Figure 12. The chain complex arising from -1 surgery on the right-hand trefoil. The shaded region highlights a bigon with edges corresponding to $m_2(x_0, \rho_{123}) = y_0$ and $\delta^1(x_1) = \rho_{123} \otimes y_1$ so that $y_0 \otimes y_1$ is a summand of $\partial^{\boxtimes}(x_0 \otimes x_1)$. The reader can also find an example of a single-piece bigon; compare Figure 12.

3. CLASSIFYING EXTENDABLE TYPE D STRUCTURES

This section proves Theorem 1.5, our structure theorem for type D structures under the hypothesis that the module in question is extendable. While this is a purely algebraic result, it is of interest in the present context owing to the fact that the three-manifold invariants that we wish to study all satisfy this additional hypothesis; compare Theorem 1.4. As such, the ultimate outcome of this section will be the existence part of Theorem 1.1.

Broadly, after defining extendable type D structures in Section 3.1, we establish the structure theorem in two steps. First, we give a partial characterization in terms of matrices over the ring $\mathbb{F}[U]/U^2$, and second, we give a combinatorial algorithm that simplifies the train tracks resulting from this partial characterization.

3.1. **Extended type D structures.** The torus algebra $\mathcal{A}$ admits an extension, denoted by $\widetilde{\mathcal{A}}$, which plays a key role.

The algebra $\widetilde{\mathcal{A}}$ is a path algebra over the quiver shown in Figure 14 with relations $\rho_i \rho_{i-1} = 0$, where the indices are to be interpreted mod 4, along with the relation $\rho_0 \rho_1 \rho_2 \rho_3 \rho_0 = 0$. As a vector space over $\mathbb{F}$ it is spanned by the idempotents ι_0 and ι_1, along with elements $\rho_I = \rho_{i_1} \rho_{i_2} \cdots \rho_{i_k}$ for each sequence $I = i_1 \ldots i_k$ of indices in $\mathbb{Z}/4\mathbb{Z}$ containing at most one 0 such that $i_{j+1} \equiv i_j + 1$ for each j. The torus algebra may be recovered from $\widetilde{\mathcal{A}}$ by setting $\rho_0 = 0$. More precisely, if $\mathcal{J}$ is the ideal generated by ρ_0 then $\mathcal{A} = \widetilde{\mathcal{A}}/\mathcal{J}$. As with $\mathcal{A}$, write $\mathbf{1} = \iota_0 + \iota_1$ and use the convention $\rho_\varnothing = \mathbf{1}$. We distinguish the element $U = \rho_{1230} + \rho_{2301} + \rho_{3012} + \rho_{0123}$ and note that U is central in $\widetilde{\mathcal{A}}$.
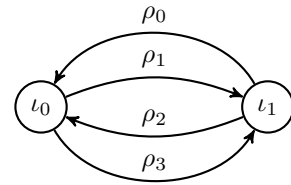
Figure 13. A quiver for the extended torus algebra.

Definition 3.1. An *extended type D structure* over $\widetilde{\mathcal{A}}$ is an $\mathcal{I}$ module $V = \iota_0 V \oplus \iota_1 V$ equipped with a map $\widetilde{\delta}^1 : V \to \widetilde{\mathcal{A}} \otimes_{\mathcal{I}} V$ such that the map $\widetilde{\partial} : \widetilde{\mathcal{A}} \otimes_{\mathcal{I}} V \to \widetilde{\mathcal{A}} \otimes_{\mathcal{I}} V$ defined by $\widetilde{\partial}(a \otimes x) = a \cdot \widetilde{\delta}^1 x$ satisfies $\widetilde{\partial}^2(x) = Ux$ for all $x \in V$. The module $\widetilde{N} = \widetilde{\mathcal{A}} \otimes_{\mathcal{I}} V$ is the *extended type D module* associated with V and $\widetilde{\delta}^1$.

An extended type D structure over $\widetilde{\mathcal{A}}$ can be specified by the pair of vector spaces $\iota_0 V$ and $\iota_1 V$ along with a collection of *coefficient maps* $D_I : \iota_{i_1-1} V \to \iota_{i_k} V$ where $I = i_1 \ldots i_k$ is a (possibly empty) sequence of indices in $\mathbb{Z}/4\mathbb{Z}$ containing at most one 0 such that $i_{j+1} \equiv i_j + 1$ for each j. In the case of the empty sequence we have two maps $D_\varnothing : \iota_i V \to \iota_i V$ for $i \in \{0, 1\}$. The coefficient

maps encode $\widetilde{\delta}^1$ by

$$\widetilde{\delta}^1 = \sum_I \rho_I \otimes D_I.$$

The condition that $\widetilde{\partial}^2(x) = Ux$ translates to the condition that for each I, the sum $\sum_{J \cup K = I} D_K \circ D_J$ is zero unless I is 0123, 1230, 2301, or 3012, in which case the sum is the identity on the appropriate summand of V. A type D structure over $\mathcal{A}$ admits a similar description in terms of coefficient maps D_I where $I \in \{\varnothing, 1, 2, 3, 12, 23, 123\}$.

Any extended type D structure over $\widetilde{\mathcal{A}}$ determines a type D structure over $\mathcal{A}$ by ignoring all coefficient maps for which I contains 0 or, equivalently, ignoring all terms in $\widetilde{\delta}_1$ involving a multiple of ρ_0; this corresponds to taking the quotient by the ideal $\mathcal{J}$. In this case we say that the initial extended type D structure is an *extension* of the quotient type D structure. Equivalently, an extended type D module $\widetilde{N}$ is an extension of N if $\mathcal{A} \otimes_{\widetilde{\mathcal{A}}} \widetilde{N}$ is isomorphic to N as a differential $\mathcal{A}$ module. We say that a type D module N is *extendable* if it has an extension $\widetilde{N}$, and we say that $N = \mathcal{A} \otimes_{\widetilde{\mathcal{A}}} \widetilde{N}$ is the underlying type D module of N.

We remark that, though it is not clear *a priori*, Proposition 4.16 shows that any two extensions of a type D structure N must be homotopy equivalent. In other words, an extended type D structure contains no more information than its underlying type D structure. Thus the chief advantage to working with extended type D structures is not the extension that we choose, but simply the fact that an extension exists. Being extendable is a rather strong algebraic constraint and, by restricting to extendable modules (as we will do for the rest of the paper), the geometric description of type D modules over $\mathcal{A}$ can be simplified considerably.

Just as a type D structure (along with a choice of basis) can be represented by an $\mathcal{A}$-decorated graph, an extended type D structure with a specified basis can be represented by an $\widetilde{\mathcal{A}}$-decorated graph, where the set of possible edge labels is expanded to include all sequences $I = i_1 \ldots i_k$ of elements of $\mathbb{Z}/4\mathbb{Z}$ with $i_\ell \equiv i_{\ell-1} + 1$ and with at most one $i_\ell = 0$. The method for constructing an extended type D structure from a graph is the same: vertices labeled by $\bullet$ and $\circ$ correspond to ι_0-generators and ι_1-generators, respectively, and an arrow labeled by I from x to y contributes a $\rho_I \otimes y$ term to $\widetilde{\delta}^1(x)$. Analogous to $\mathcal{A}$-decorated graphs, $\widetilde{\mathcal{A}}$-decorated graphs satisfy a condition on the counts of length two paths of arrows, which ensures the result is in fact an extended type D structure.

In Section 2 we realized the $\mathcal{A}$-decorated graph Γ representing a reduced or almost reduced type D structure geometrically by immersing it in the parametrized torus T. This construction extends immediately to decorated graphs associated with extensions, where the arcs corresponding to the additional edge labels are determined by Figure 15; the result is a naive train track representing the extended type D structure. Note that all edges of this train track are oriented, but a pair of vertices may be connected by a pair of parallel but oppositely oriented edges; in this case, for convenience, we adopt the convention that the two opposing oriented edges are replaced with a single unoriented edge. For example, consider the type D module in Figure 9; a choice of extension is illustrated in Figure 16 together with the corresponding train track.

3.2. **Train tracks and matrices.** We have so far considered only naive train tracks constructed from $\mathcal{A}$-decorated or $\widetilde{\mathcal{A}}$-decorated graphs, but we will have need to work with a larger class of train tracks in the marked torus, which we now formalize.
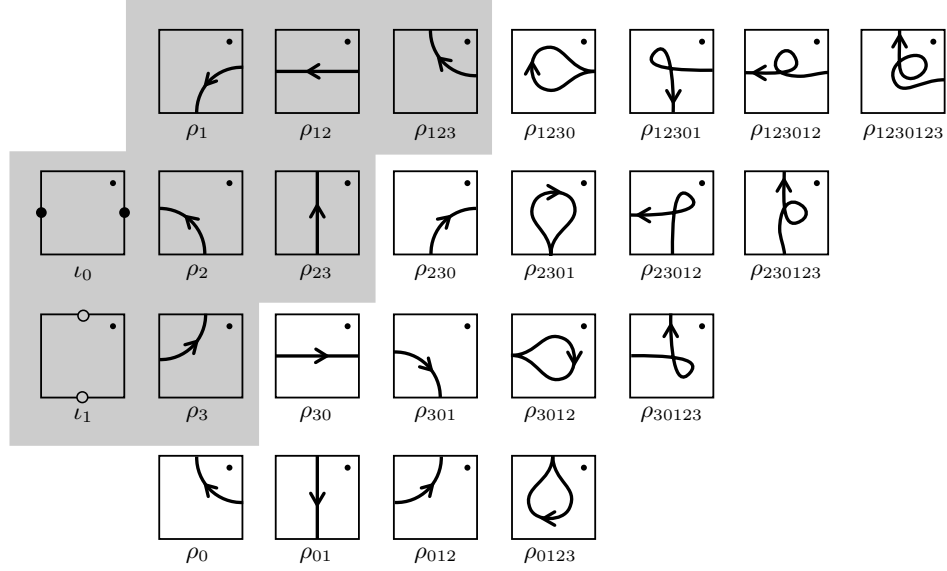
Figure 14. Train track edges corresponding to elements of the extended algebra $\widetilde{\mathcal{A}}$. These can also be interpeted as homotopy classes of clockwise moving paths (or constant paths, in the case of ι_0 and ι_1) connecting sides of a square and passing the top right corner at most once, and $\widetilde{\mathcal{A}}$ is generated as an $\mathbb{F}$-vector space by these homotopy classes; algebra multiplication is concatenation of paths, where the product is zero if the concatenation does not exist or passes the top right corner twice. The shaded elements generate $\mathcal{A}$.
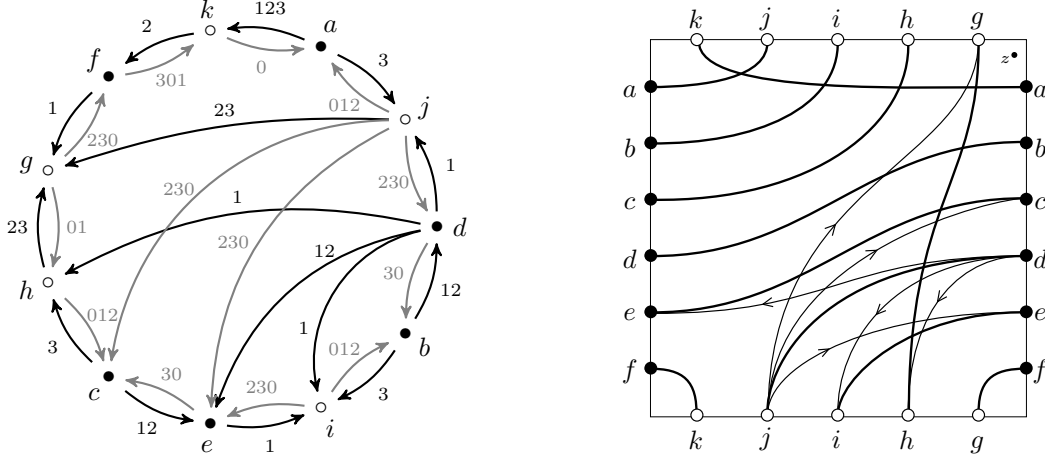
Figure 15. The train track (right) associated with an extended type D structure (left). This extended type D structure is an extension of the type D structure in Figure 9; the extended coefficient maps are shown in gray.

Definition 3.2. A *train track in the marked torus T* will refer to an immersed train track $\boldsymbol{\vartheta}$ in T with the following properties: (1) $\boldsymbol{\vartheta}$ is disjoint from the basepoint in T and perpendicular to α and β; (2) every intersection of $\boldsymbol{\vartheta}$ with $\alpha \cup \beta$ is a switch and called a *primary switch* of $\boldsymbol{\vartheta}$; and (3) no oriented path carried by $\boldsymbol{\vartheta}$ contains a counterclockwise fishtail (see Figure 17).
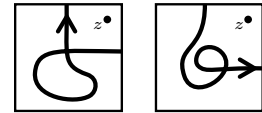
Figure 16. Examples of counterclockwise fishtails. These never appear in a toroidal train track.

Note that all the switches in a naive train track are primary, but
this is not required by the definition. As we will see in the next subsection, one of the key advantages
of the train track notation is that it allows us to push some of the switches in a train track into
the interior of the fundamental domain for T. In addition, the definition allows for train tracks
where a section of the track simply terminates in a monovalent vertex. Although tracks of this form
cannot appear as the train track associated with an extendable type D structure, pairing with them
provides a useful way of distinguishing between two such tracks (*c.f.* Section 4.3).

While we will often simply use the term *train track*, all train tracks we will discuss are in the marked
torus (with the exception of those in Section 3.8) and should be understood to satisfy the conditions
above. Within this broader class of objects, we will be able to manipulate our naive train tracks
geometrically to find much simpler representatives. Importantly, the discussion of pairing above will
carry over directly to this more general class of train tracks.

Given a train track $\boldsymbol{\vartheta}$ in T with n primary switches, we can construct a $2n \times 2n$ matrix $M_{\boldsymbol{\vartheta}}$
encoding the essential information of $\boldsymbol{\vartheta}$. We cut T open along α and β to form a square, with our
usual convention that β is horizontal (oriented rightward), α is vertical (oriented upward), and the
basepoint is in the top right corner of the square. The primary switches give $2n$ marked points on
the edges of the square, which we label $v_1, \ldots, v_{2n}$ clockwise, starting from the top right corner. We
let k denote the number of primary switches of $\boldsymbol{\vartheta}$ on α, so that v_i is on the right side of the square
if $1 \leq i \leq k$, the bottom side if $k+1 \leq i \leq n$, the left side if $n+1 \leq i \leq n+k$, and the top side if
$n+k+1 \leq i \leq 2n$. The (i,j) entry of $M_{\boldsymbol{\vartheta}}$ counts immersed oriented paths in $\boldsymbol{\vartheta}$ through the square
from v_i to v_j. Such a path can be given a label recording how many times it passes the base point
on its left. More precisely, any immersed path through the square from a primary switch x to a
primary switch y can be projected to a path in the boundary of the square by sending each point in
the interior of the path to the boundary following the path's leftward pointing normal vector. The
resulting path connects the corner after x (moving clockwise around the boundary of the square) to
the corner before y. We precompose this path with the clockwise moving path from x to the corner
after x and postcompose with the path from the corner before y to y, resulting in a path from x to
y in the boundary of the square. The assumption that $\boldsymbol{\vartheta}$ has no counterclockwise fishtails ensures
that this projected path will have net clockwise movement around the boundary. We weight each
path by U^m, where m is the number of times (counted with sign) that the projected path passes
the top right corner of the square. We set $U^2 = 0$ and count paths modulo 2, so $M_{\boldsymbol{\vartheta}}$ has coefficients
in $\mathbb{F}[U]/U^2$.

Definition 3.3. We will say that train tracks $\boldsymbol{\vartheta}_1$ and $\boldsymbol{\vartheta}_2$ in T are *equivalent* if $M_{\boldsymbol{\vartheta}_1} = M_{\boldsymbol{\vartheta}_2}$. A train
track $\boldsymbol{\vartheta}$ is said to be *reduced* if no immersed oriented paths through the square $T \setminus (\alpha \cup \beta)$ begin
and end on the same side of the square unless they are weighted by a positive power of U. A train
track $\boldsymbol{\vartheta}$ is *valid* if $M_{\boldsymbol{\vartheta}}^2 = UI_{2n}$, where I_{2n} is the $2n \times 2n$ identity matrix.

The significance of these restrictions on train tracks is found in the following proposition:

Proposition 3.4. *The following are equivalent:*

(i) *Reduced extended type D structures with a chosen (ordered) set of generators.*
(ii) *Equivalence classes of valid, reduced train tracks in T.*
(iii) *$2n \times 2n$ matrices M over $\mathbb{F}[U]/U^2$, together with an integer $0 \leq k \leq n$, such that $M^2 = UI_{2n}$
and M is strictly block upper triangular modulo U with respect to blocks of size k, $n-k$, k,
and $n-k$.*

Before proving this proposition it is worth pausing to consider a simple example. The pair of train tracks in Figure 18 are equivalent; the reader should construct the associated four-by-four matrix in each case and observe that these are equal, as required. The key feature here is that the additional path in the train track on the right carries the label U^2, which is set to zero in the matrix. On the other hand, an inequivalent train track is obtained by considering only the horizontal two-way tracks of this

Figure 17. Equivalent valid train tracks in T.

example. Although this train track is inequivalent to the ones given, it represents an isomorphic module. For generators x and y such that $D_{123012}(x) = y$ there is an isomorphism of modules over $\widetilde{\mathcal{A}}$ given by $x' = x + \rho_{1230} \otimes y$ and $y' = y$, so that $D_{123012}(x') = 0$. Note that this is an isomorphism between two different choices of extension (of a fixed type D structure) as $\widetilde{\mathcal{A}}$ modules but not an isomorphism as $\mathbb{F}$ vector spaces. That is, this train track is weakly equivalent to those of Figure 18 in the sense of Definition 3.5 below.
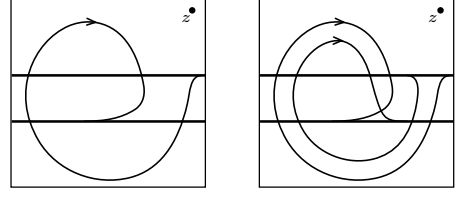
Proof of Proposition 3.4. Any extended type D structure $\widetilde{N}$, along with a specified basis B, determines a naive train track $\boldsymbol{\vartheta}(\widetilde{N}, B)$ as described above, and it is immediate from the construction that if $\widetilde{N}$ is reduced then so is $\boldsymbol{\vartheta}(\widetilde{N}, B)$. This particular train track determines an equivalence class of train tracks. We must check that $\boldsymbol{\vartheta}(\widetilde{N}, B)$ is valid, but postpone this until after the discussion of matrices below.

Any valid reduced train track $\boldsymbol{\vartheta}$ determines a matrix $M_{\boldsymbol{\vartheta}}$ as described above. This is a $2n \times 2n$ matrix, where each row or column corresponds to primary switches on the boundary of the square $T \setminus (\alpha \cup \beta)$. These may be split into blocks according to which side each primary switch is on. The fact that paths in $\boldsymbol{\vartheta}$ have net clockwise rotation ensures that any path labelled by 1 contributes to an entry above the diagonal in $M_{\boldsymbol{\vartheta}}$; the fact that $\boldsymbol{\vartheta}$ is reduced further ensures that no 1 appears in the blocks along the diagonal of $M_{\boldsymbol{\vartheta}}$. By definition, the fact that $\boldsymbol{\vartheta}$ is valid implies $M_{\boldsymbol{\vartheta}}^2 = UI_{2n}$.

Finally, a reduced extendable type D structure can be extracted from a matrix M and an integer k as in (iii). The vector space $\iota_0 V$ has generators $x_1, \ldots, x_k$ and $\iota_1 V$ has generators $x_{k+1}, \ldots, x_n$. Each generator x_i corresponds to a pair of indices i and $t(i)$, where

$$t(i) = \begin{cases} n + k + 1 - i & \text{if } i \leq k \\ 2n + k - i + 1 & \text{if } i > k. \end{cases}$$

An entry in the j or $t(j)$ column and the i or $t(i)$ row of M contributes an x_j term to the coefficient map $D_I(x_i)$, where I is determined by the position in the matrix with respect to the block decomposition of M. More precisely, write $M = M_1 + M_U$ where M_1 has entries that are 1 or 0 while M_U has entries that are U or 0. Then the block decompositions of M_1 and M_U with respect to blocks of size k, $n - k$, k, and $n - k$ recover the coefficient maps for the corresponding extended type D structure as follows:

(1) $\quad M_1 = \begin{bmatrix} 0 & D_1 & D_{12} & D_{123} \\ 0 & 0 & D_2 & D_{23} \\ 0 & 0 & 0 & D_3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad M_U = U \begin{bmatrix} D_{1230} & D_{12301} & D_{123012} & D_{1230123} \\ D_{230} & D_{2301} & D_{23012} & D_{230123} \\ D_{30} & D_{301} & D_{3012} & D_{30123} \\ D_0 & D_{01} & D_{012} & D_{0123} \end{bmatrix}$

Here the blocks are the matrices representing D_I, acting by *right* multiplication, with respect to the basis $\{x_1, \ldots, x_k\}$ of $\iota_0 V$, the basis $\{x_{k+1}, \ldots, x_n\}$ of $\iota_1 V$, or these bases in reversed order, as appropriate. In the case that $M = M_{\boldsymbol{\vartheta}(\widetilde{N}, B)}$ is the matrix associated with a pair $(\widetilde{N}, B)$, the extended type D structure extracted from M in this way exactly recovers $\widetilde{N}$, as there is a one to one

correspondence between nonzero terms in M_1 or M_U, paths in the train track $\boldsymbol{\vartheta}(\widetilde{N}, B)$, and nonzero terms in the coefficient maps for $\widetilde{N}$.

It remains to show that $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}(\widetilde{N}, B)$ is valid, or equivalently that the corresponding matrix $M_{\boldsymbol{\vartheta}(\widetilde{N}, B)}$ squares to UI_{2n}, and conversely that $\widetilde{\partial}^2(x) = Ux$ in the extended type D structure determined by any matrix M with $M^2 = UI_{2n}$. To see this, we look at the block decomposition of M^2, which can be computed from the block decomposition of M_1 and M_U above. Let $\bar{D}_I$ denote the sum $\sum_{J \cup K = I} D_J D_K$ where J and K are nonempty sequences of indices in $\mathbb{Z}/4\mathbb{Z}$. Using this shorthand, we have:

$$
M^2 = \left[\begin{array}{c|c|c|c} 0 & 0 & \bar{D}_{12} & \bar{D}_{123} \\ \hline 0 & 0 & 0 & \bar{D}_{23} \\ \hline 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right] + U \left[\begin{array}{c|c|c|c} \bar{D}_{1230} & \bar{D}_{12301} & \bar{D}_{123012} & \bar{D}_{1230123} \\ \hline \bar{D}_{230} & \bar{D}_{2301} & \bar{D}_{23012} & \bar{D}_{230123} \\ \hline \bar{D}_{30} & \bar{D}_{301} & \bar{D}_{3012} & \bar{D}_{30123} \\ \hline \bar{D}_0 & \bar{D}_{01} & \bar{D}_{012} & \bar{D}_{0123} \end{array}\right]
$$

Note that, by slight abuse of notation, we are using D_I to denote both a coefficient map and a matrix representing it, and the matrix acts by right multiplication. Thus the matrix $D_J D_K$ is the matrix representing the map $D_K \circ D_J$. It is now clear that the condition that $M^2 = UI_{2n}$ is precisely the condition on composition of coefficient maps for the corresponding reduced extended type D structure. □

To restrict an extended type D module $\widetilde{N}$ to its underlying type D module N, we simply set $\rho_0 = 0$ in the algebra, and thus ignore all coefficient maps D_I for which I contains 0. It is clear that the analogous operation for the corresponding matrix is setting $U = 0$, resulting in a matrix with coefficients in $\mathbb{F}$. This motivates the following definitions:

Definition 3.5. A train track $\boldsymbol{\vartheta}$ in T is *weakly valid* if the corresponding matrix $M_{\boldsymbol{\vartheta}}$ squares to 0 modulo U. Two train tracks $\boldsymbol{\vartheta}_1$ and $\boldsymbol{\vartheta}_2$ are *weakly equivalent* if the matrices $M_{\boldsymbol{\vartheta}_1}$ and $M_{\boldsymbol{\vartheta}_2}$ agree modulo U.

The following is immediate from Proposition 3.4.

Proposition 3.6. *The following are equivalent:*

  (i) *Reduced type D structures over $\mathcal{A}$, up to isomorphism as $\mathbb{F}$ vector spaces, with a chosen (ordered) set of generators.*
  (ii) *Weak equivalence classes of weakly valid, reduced train tracks in T.*
  (ii) *$2n \times 2n$ matrices M over $\mathbb{F}$, together with an integer $0 \leq k \leq n$, such that $M^2 = 0$ and M is strictly block upper triangular with respect to blocks of size k, $n - k$, k, and $n - k$.* □

Although we are mainly interested in reduced train tracks, we observe that valid almost reduced train tracks can be defined as those corresponding to almost reduced type D structures over $\mathcal{A}$.

We pause briefly to discuss pairing of train tracks in T, which is defined just as it was for naive train tracks in Section 2.5. A pair $(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1)$ is *admissible* if there are no immersed annuli cobounded by $A(\boldsymbol{\vartheta}_0)$ and $D(\boldsymbol{\vartheta}_1)$, where $A(\boldsymbol{\vartheta})$ denotes the result of including $\boldsymbol{\vartheta}$ into the first quadrant and extending vertically/horizontally and $D(\boldsymbol{\vartheta})$ is the reflection of $A(\boldsymbol{\vartheta})$ across the anti-diagonal. Then $\mathcal{C}(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1)$ is the vector space generated by intersection points of $A(\boldsymbol{\vartheta}_0)$ and $D(\boldsymbol{\vartheta}_1)$, equipped with a linear map $d^{\boldsymbol{\vartheta}}$ that counts bigons not covering the basepoint with boundary carried by the pair $(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1)$, with the requirement that both sides of the bigon are oriented paths from the initial intersection point to the terminal intersection point. Note that, unlike for naive train tracks, the orientation requirement is meaningful for an arbitrary train track in T. The key observation is that defining $\mathcal{C}(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1)$ does not require all the information in $\boldsymbol{\vartheta}_1$ and $\boldsymbol{\vartheta}_2$, it is in fact determined only by their weak equivalence classes.

Proposition 3.7. *Let $\boldsymbol{\vartheta}_0$, $\boldsymbol{\vartheta}_0'$ be weakly valid reduced train tracks and $\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_1'$ be weakly valid almost reduced train tracks such that the pairs $(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1)$ and $(\boldsymbol{\vartheta}_0', \boldsymbol{\vartheta}_1')$ are admissible. If $\boldsymbol{\vartheta}_0$ is weakly equivalent to $\boldsymbol{\vartheta}_0'$ and $\boldsymbol{\vartheta}_1$ is weakly equivalent to $\boldsymbol{\vartheta}_1'$, then $\mathcal{C}(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1) = \mathcal{C}(\boldsymbol{\vartheta}_0', \boldsymbol{\vartheta}_1')$ as chain complexes.*

Proof. Note that to count bigons connecting two intersection points, one need only know the count of paths connecting each pair of primary vertices in each train track. The orientation requirement implies that we can ignore any path in $\boldsymbol{\vartheta}_0$ or $\boldsymbol{\vartheta}_1$ that has the puncture on the left side of the path; these are precisely the paths weighted by U in $M_{\boldsymbol{\vartheta}_0}$ and $M_{\boldsymbol{\vartheta}_1}$. Thus the map $d^{\boldsymbol{\vartheta}}$ is determined by the weak equivalence class of $\boldsymbol{\vartheta}_0$ and $\boldsymbol{\vartheta}_1$. □

Note that under the hypotheses above $d^{\boldsymbol{\vartheta}}$ is a differential, making $\mathcal{C}(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1)$ a chain complex. Indeed, if $\boldsymbol{\vartheta}_0$ and $\boldsymbol{\vartheta}_1$ correspond to type D structures N_0 and N_1, respectively, then they are weakly equivalent to the naive train tracks representing N_0 and N_1, and it follows from Theorem 2.2 that $\mathcal{C}(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1)$ is isomorphic as a chain complex to $N_0^A \boxtimes N_1$.

3.3. **Simplifying valid train tracks.** Consider our running example and the train track in Figure 16. We observe that this train track has two special properties. First, there are exactly two unoriented edges, one from each direction, incident to each switch. It follows that if the oriented edges are ignored, the unoriented edges form a collection of immersed curves. Second, the extra oriented edges can be grouped in pairs, where each pair connects the same two sections of the immersed curve, oriented the same way but turning opposite directions at each end and crossing once. We can slide the endpoints of each pair of oriented segments near each other along the unoriented segments, without changing the equivalence class of the train track, as shown in Figure 19(a). (In doing so, we have passed from a naive train track in which every switch is primary, to one in which some of the switches lie in the interior of the fundatmental domain for T.) The resulting X shaped pairs of oriented arrows will be called *crossover arrows*, and to simplify diagrams these pairs will be replaced by a bold arrow (Figure 19(b)).

(a)   (b)
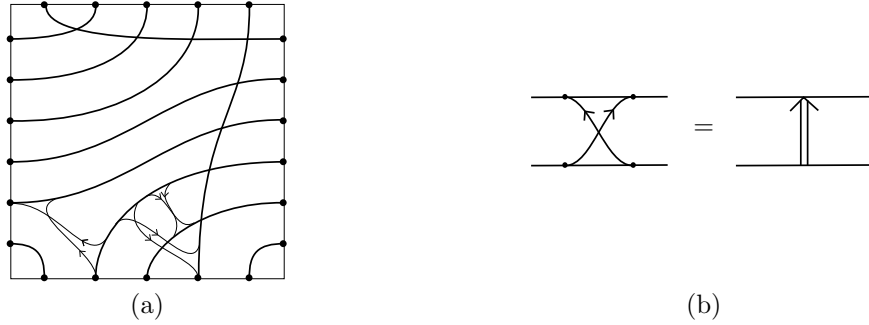
Figure 18. (a) The train track from Figure 16 homotoped so that the oriented edges form crossover arrows. (b) Diagrammatic shorthand for crossover arrows.

The goal of the present subsection is to prove that these features are not unique to this example, but indeed hold for (some representative of) every equivalence class of valid reduced train tracks.

Proposition 3.8. *Any valid reduced train track in T is equivalent to a train track $\boldsymbol{\vartheta}$ with the following form:*

- *Every switch of $\boldsymbol{\vartheta}$ has exactly one unoriented edge on each side, so that $\boldsymbol{\vartheta}$ is a collection of (unoriented) immersed curves along with additional oriented edges connecting points on the immersed curves;*
- *the immersed curves restrict to horizontal and vertical segments outside of the square $[\frac{1}{4}, \frac{3}{4}] \times [\frac{1}{4}, \frac{3}{4}]$, which we call the corner box since it is where the immersed curves change direction;*
- *within the corner box, the immersed curves restrict to embedded arcs connecting different sides of the corner box (we call these arcs corners);*

- *the oriented edges appear in pairs that form crossover arrows (as in Figure 19);*
- *the crossover arrows appear outside the corner box, and they move clockwise around the corner box.*

Conversely, any train track of this form is a valid, reduced train track in T.

See Figure 21(c) for an example of a train track of this form.

Before proving Proposition 3.8, we review some linear algebra of matrices over $\mathbb{F}[U]/U^2$. Any invertible matrix over $\mathbb{F}[U]/U^2$ can be written as the product of elementary matrices of the following two types:

$$
A_{i,j} = \quad
\begin{matrix} & & & & j & \\ \end{matrix}
\begin{bmatrix}
1 & & & & & & \\
& \ddots & & & & & \\
& & 1 & & 1 & & \\
& & & \ddots & & & \\
& & 0 & & 1 & & \\
& & & & & \ddots & \\
& & & & & & 1
\end{bmatrix}
\qquad
A_{i,j}^U = \quad
\begin{bmatrix}
1 & & & & & & \\
& \ddots & & & & & \\
& & 1 & & U & & \\
& & & \ddots & & & \\
& & 0 & & 1 & & \\
& & & & & \ddots & \\
& & & & & & 1
\end{bmatrix}
$$

For $A_{i,j}$ we require $i \neq j$; in $A_{i,j}^U$ the (i,i) entry is $1 + U$. Each of these elementary matrices is its own inverse. We remark that there is another type of elementary matrix, transposition matrices, but over $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$ these can be obtained by products of matrices of type $A_{i,j}$.

Lemma 3.9. *Let M be a $2n \times 2n$ matrix over the base ring $\mathbb{F}[U]/U^2$ that is upper triangular modulo U such that $M^2 = UI_{2n}$. Then M can be written as $P = P\bar{M}P^{-1}$ where P is composition of elementary matrices of the form $A_{i,j}^U$ or of the form $A_{i,j}$ with $i < j$, and $\bar{M}$ has the following form:*

- *Each row and each column of $\bar{M}$ has exactly one nonzero entry;*
- *The diagonal entries of $\bar{M}$ are zero;*
- *If $1 \leq i < j \leq 2n$, either the (i,j) entry and the (j,i) entry of $\bar{M}$ are both zero or the (i,j) entry is 1 and the (j,i) entry is U.*

Proof. Since the elementary matrices are their own inverses, it is enough to show that M can be reduced to a matrix $\bar{M}$ of the desired form by conjugating with elementary matrices of the form $A_{i,j}^U$ or of the form $A_{i,j}$ with $i < j$. Note that conjugating by these matrices preserves the fact that M is upper triangular mod U and that $M^2 = UI_{2n}$.

We proceed by induction on n. For $n = 1$, there are finitely many 2×2 matrices over $\mathbb{F}[U]/U^2$. Only four of these square to UI_2 and are upper triangular mod U:

$$
\begin{bmatrix} 0 & 1 \\ U & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1+U \\ U & 0 \end{bmatrix}, \begin{bmatrix} U & 1 \\ U & U \end{bmatrix}, \text{ and } \begin{bmatrix} U & 1+U \\ U & U \end{bmatrix}.
$$

The first has the desired form already, and it is routine to check that the remaining three reduce to the first after conjugating with $A_{1,1}^U$, $A_{2,1}^U$, and $A_{1,1}^U A_{2,1}^U$, respectively.

If $n > 1$, choose i and j with $j - i$ as small as possible so that the (i,j) entry of M is 1 or $1 + U$. Conjugating by $A_{j,k}$ has the effect of adding the jth column to the kth column and adding the kth row to the jth row. After conjugating by $A_{j,k}$ if necessary for each $k > j$, we can ensure that the (i,k) entry of M is 0 or U. After conjugating with matrices of the form $A_{j,k}^U$ for any k, we can assume that the ith row has a 1 in the jth entry and zeros elsewhere. The fact that $M^2 = UI_{2n}$ then implies that the jth row has a U in the ith entry and zeroes elsewhere. Similarly, after conjugating with $A_{k,i}$ with $k < i$ or with $A_{k,i}^U$ for any k we can ensure that the only nonzero entry in the jth column of M is a 1 in the ith entry. It follows from the fact that $M^2 = UI_{2n}$ that the only nonzero entry in the ith column is a U in the jth entry.

After the conjugations described above, the matrix M has the form desired in the ith and jth rows and columns. We need to show that rest of the matrix can be put in the desired form using conjugations with $A_{k,\ell}$ and $A_{k,\ell}^U$ with k and ℓ not in $\{i,j\}$. This is equivalent to showing that the the $2(n-1) \times 2(n-1)$ matrix obtained from this one by deleting the ith and jth rows and columns can be reduced to the desired form, which we take as the inductive hypothesis. □

Lemma 3.9 lets us relate M to a matrix of a simple form through conjugation by elementary matrices; this in turn has a nice geometric interpretation in terms of train tracks. Let $\boldsymbol{\vartheta}$ be a train track that has only horizontal or vertical segments outside the corner box $[\frac{1}{4}, \frac{3}{4}] \times [\frac{1}{4}, \frac{3}{4}]$, and let $\boldsymbol{\vartheta}'$ be the train track obtained from $\boldsymbol{\vartheta}$ by adding a clockwise moving crossover arrow outside the corner box from the segment with endpoint x_i to the segment with endpoint x_j. We claim that $M_{\boldsymbol{\vartheta}'}$ is obtained from $M_{\boldsymbol{\vartheta}}$ by conjugating with an elementary matrix. The counts of smooth paths are identical except that there are some new paths in $\boldsymbol{\vartheta}'$ involving the crossover arrow. For each path into x_i in $\boldsymbol{\vartheta}$, $\boldsymbol{\vartheta}'$ will have an additional path into x_j, and for each path out of x_j in $\boldsymbol{\vartheta}$, $\boldsymbol{\vartheta}'$ will have an additional path out of x_i; if the crossover arrow passes the basepoint in the top right corner of the square m times, the new paths will be weighted by an


Figure 19. After adding a crossover arrow from x_i to x_j, the path from y to x_i gives rise to a path from y to x_j (shaded). There is also a new path from x_i to y', which carries a weight of U.

additional factor U^m. An example of this is shown in Figure 20. It follows that $M_{\boldsymbol{\vartheta}'}$ is obtained from M by adding U^m times the jth row to the ith row and U^m times the ith column to the jth column. In other words, we conjugate by $A_{i,j}$ if the crossover arrow does not pass the base point or by $A_{i,j}^U$ if the crossover arrow passes the basepoint once. (If the basepoint is passed more than once then $M_{\boldsymbol{\vartheta}'} \equiv M_{\boldsymbol{\vartheta}}$ modulo U^2).

Proof of Proposition 3.8. Let $\boldsymbol{\vartheta}'$ be a valid reduced train track in T and let $M = M_{\boldsymbol{\vartheta}'}$ be the corresponding matrix. Set $M = P\bar{M}P^{-1}$ as in Lemma 3.9; the matrix $\bar{M}$ corresponds to an extended type D structure, from which we can construct a train track $\bar{\boldsymbol{\vartheta}}$. To construct $\bar{\boldsymbol{\vartheta}}$, let $\bar{\boldsymbol{\vartheta}}$ have the same primary vertices as $\boldsymbol{\vartheta}'$, labelled $x_1, \ldots x_{2n}$ as described above, and add oriented corner edges from x_i to x_j if the (i,j) entry of $\bar{M}$ is nonzero. The form of $\bar{M}$ implies that there will be exactly one incoming and one outgoing corner edge at each primary vertex, and that for each corner edge from x_i to x_j there is an oppositely oriented edge from x_j to x_i. Replacing the pairs of oppositely oriented corner edges with single unoriented corner edges as usual, observe that there is exactly one unoriented corner edge at each primary vertex. It follows that $\bar{\boldsymbol{\vartheta}}$ is an immersed multicurve.

We observed that M is obtained from $\bar{M}$ by conjugating with a sequence of elementary matrices of the form $A_{i,j}$ with $i < j$ or $A_{i,j}^U$. Each conjugation corresponds to adding a crossover arrow to $\bar{\boldsymbol{\vartheta}}$ near the boundary of the square. That is, adding the appropriate sequence of crossover arrows to $\bar{\boldsymbol{\vartheta}}$ produces a train track $\boldsymbol{\vartheta}$ whose corresponding matrix is M. Since we only use type $A_{i,j}$ matrices if $i < j$, all the crossover arrows added move clockwise. Up to regular homotopy, we can assume that $\bar{\boldsymbol{\vartheta}}$ is horizontal or vertical outside of the corner box $[\frac{1}{4}, \frac{3}{4}] \times [\frac{1}{4}, \frac{3}{4}]$ and that the crossover arrows added lie outside the corner box. Since $\boldsymbol{\vartheta}$ is reduced, the arcs within the corner box connect different sides of the square. Thus $\boldsymbol{\vartheta}$ has the form claimed and $\boldsymbol{\vartheta}$ is equivalent to $\boldsymbol{\vartheta}'$ since the matrix associated to each is M.

Conversely, given any train track $\boldsymbol{\vartheta}$ of this form, there is a reduced train track $\bar{\boldsymbol{\vartheta}}$ obtained by removing the crossover arrows. $\bar{\boldsymbol{\vartheta}}$ is clearly valid and $M_{\boldsymbol{\vartheta}}$ is obtained from $M_{\bar{\boldsymbol{\vartheta}}}$ by conjugating with elementary matrices, which preserves the property that the matrix squares to UI_{2n}. It follows that $\boldsymbol{\vartheta}$ is valid as well as reduced. □
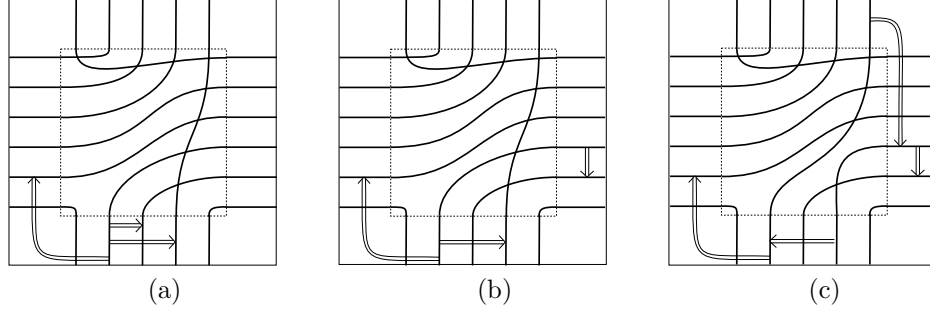
Figure 20. (a) A train track equivalent to the one in in Figure 19, after a slight homotopy. This has the form specified in Proposition 3.8 except that two crossover arrows move counterclockwise. (b) The innermost counterclockwise arrow can be replaced by a clockwise arrow by pushing it through the corner region. (c) The remaining counterclockwise arrow can be replaced with two clockwise arrows by pushing through the corner region and resolving a crossing. Note that this train track is in the form guaranteed by Proposition 3.8.

We remark that in many cases, a naive train track can be transformed into a train track of the form predicted by Proposition 3.8 geometrically, without using the linear algebra above. For example, in the running example in Figure 19(a), pairs of oriented edges have been homotoped together to form crossover arrows; it is clear that this homotopy preserves the equivalence class of train tracks. A further homotopy ensures that the crossover arrows lie outside of a corner box, as in Figure 21(a). This train track has the desired form except that two of the crossover arrows run counterclockwise. This can be fixed by sliding the two counterclockwise arrows through the corner box and using some local moves that preserve the equivalence class of train track, which will be discussed in Section 3.4 (see Figure 22); in this case the local moves result in resolving a crossing of the immersed curve and adding an additional crossover arrow. In small examples, this geometric approach is often faster than the linear algebra used above. However, it would difficult to make this approach systematic for all cases. By passing from train tracks to matrices and back, the proof of Proposition 3.8 takes advantage of the interplay between linear algebra and our geometric arguments.

3.4. **Crossover arrow calculus.** By Proposition 3.8, we can always represent a reduced extendable type D structure with a train track consisting of an immersed multicurve with crossover arrows. Our next aim is to simplify these train train tracks using certain geometric moves involving sliding crossover arrows along the immersed curves. First, we are allowed to slide arrows along the multicurve within the square $T \setminus (\alpha \cup \beta)$ without changing the equivalence class of the train track, as long as ends of arrows do not slide past each other. Moreover, the local moves shown in Figure 22 produce equivalent train tracks. To see this, note that the mod 2 count of oriented paths between any pair of endpoints is the same on both sides of each replacement. Arrows may slide past each other within the square $T \setminus (\alpha \cup \beta)$ with the caveat that if the head of one arrow passes the tail of another arrow, we must add a new arrow that is homotopic to the composition of the two arrows. By composition, we mean the the arrows are stacked head to tail, where we assume the tail of the second arrow and the head of the first arrow lie on opposite sides of an immersed curve at the same point; to arrange this we may first need to add U-turns at the beginning and end of one arrow, which may be viewed as a regular homotopy of train tracks. Finally, a pair of opposing arrows connecting the same two segments of immersed curve can be replaced with a crossing and one arrow.

In addition to the local moves in Figure 22, we will make use of two additional moves to modify the train tracks:

(M1) add or remove a clockwise crossover arrow that is in a neighborhood of the boundary of the square $T \setminus (\alpha \cup \beta)$ and which passes at least one corner of the square;
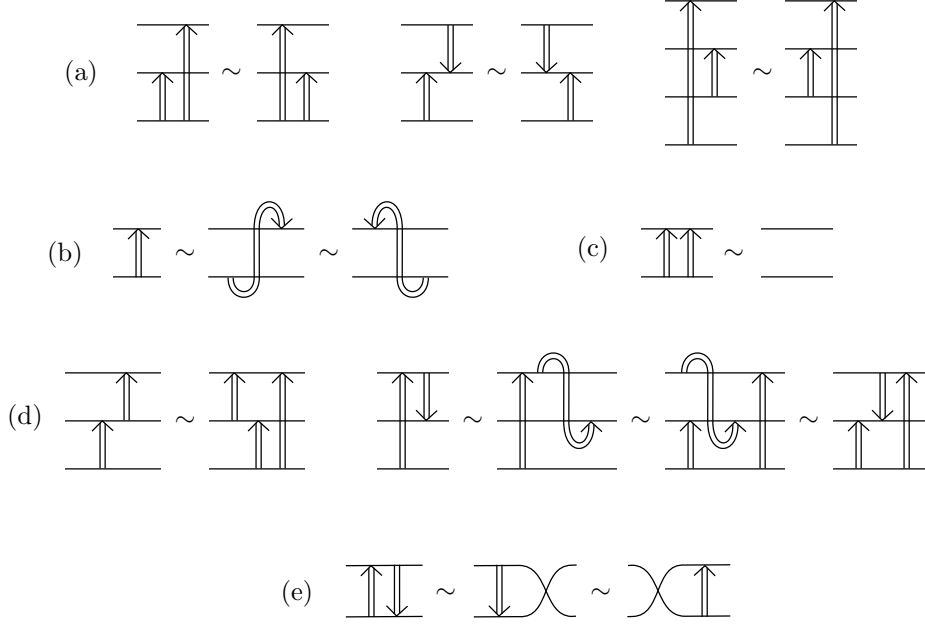
Figure 21. Local moves for crossover arrows producing equivalent train tracks: (a) beginnings or ends of arrows slide past each other freely (three such instances are pictured, though this is not exhaustive; for instance, similar configurations can be obtained by reversing the direction of all arrows), (b) U-turns can be added at the start and end of an arrow if they turn opposite directions; (c) a pair of adjacent parallel crossover arrows connecting the same two strands can be added or removed; (d) the start of one arrow may slide past the end of another arrow at the expense of adding a new arrow that is the composition of the two (two example configurations are shown); (e) opposing arrows can be replaced with one arrow and a crossing.
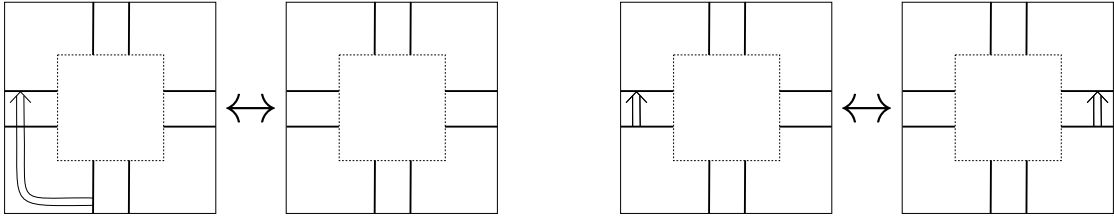
Figure 22. Examples of the moves (M1), left, and (M2), right.

(M2) move a crossover arrow which is in a neighborhood of one side of the square $T \setminus (\alpha \cup \beta)$ to the opposite side of the square; in other words, slide an arrow across either α or β in the torus T.

Examples of these moves are shown in Figure 23. In each move, we assume that there are no other crossover arrows between the arrow being (re)moved and the boundary of the square. More precisely, there is a neighborhood of the boundary of the square containing the crossover arrow in question but no other crossover arrows, and in which the immersed curve consists of horizontal and vertical segments. In contrast with the local moves described in Figure 22, the moves (M1) and (M2) change the equivalence class of the train track. However, these moves induce isomorphisms between the associated extended type D structures.

Proposition 3.10. *Suppose ϑ' is obtained from ϑ by an application of a sequence of the moves (M1) and (M2) together with the local moves depicted in Figure 22. If $\widetilde{N}'$ and $\widetilde{N}$ are the extended type D structures associated with ϑ' and ϑ, respectively, then there is an isomorphism $\widetilde{N}' \cong \widetilde{N}$ as $\widetilde{\mathcal{A}}$ modules.*

Proof. As observed, all local moves produce equivalent train tracks, and hence equality on the associated extended type D structures. To prove the proposition then, we need to establish changes of basis for each of the moves (M1) and (M2) that induce the desired isomorphisms.

Suppose first that $\boldsymbol{\vartheta}'$ is obtained from $\boldsymbol{\vartheta}$ by an application of (M1), adding a crossover arrow from horizontal or vertical edge representing a generator x of $\widetilde{N}$ to an edge representing a generator y. We claim that $\widetilde{N}'$ is related to $\widetilde{N}$ by the change of basis replacing x with $x + \rho_I \otimes y$, where I is the sequence of corners passed by the crossover arrow being added.

Consider for concreteness the case where the sequence of corners I associated with the arrow being added is 1 (other cases of clockwise arrows are similar). Thus x has idempotent ι_0, y has idempotent ι_1, and we remove a crossover arrow from the left end of the horizontal segment corresponding to x, which we will call v_x, to the top end of the vertical segment corresponding to y, which we will call v_y. Adding the crossover arrow adds a new corner path starting at v_x for each corner path starting at v_y, and a new corner path ending at v_y for each one ending at v_x. The corner paths starting at v_y correspond to $\rho_J \otimes z$ terms in $\tilde{\partial}(y)$ where the sequence J begins with 2. The new corner paths starting from v_x pass one extra corner, so they correspond to terms $(\rho_1 \rho_J) \otimes z$ in $\tilde{\partial}(x)$. In other words, we add $\rho_1 \tilde{\partial}(y)$ to $\tilde{\partial}(x)$ (note that multiplying $\tilde{\partial}(y)$ on the left by ρ_1 picks out exactly the terms for which the Reeb chord interval starts with 2). Similarly, each corner path ending at v_x corresponds to a term $\rho_I \otimes x$ in $\tilde{\partial}(z)$ for some z and some I ending in 0, and for each the new corner path added corresponds to a $(\rho_I \rho_1) \otimes y$ term in $\tilde{\partial}(z)$. This is precisely the effect of the change of basis (over $\widetilde{\mathcal{A}}$) replacing x with $x + \rho_1 \otimes y$.

Now suppose that $\boldsymbol{\vartheta}'$ is obtained from $\boldsymbol{\vartheta}$ by an application of (M2), where the crossover arrows slides between two horizontal or two vertical edges representing generators x and y of $\widetilde{N}$ and the crossover arrow is oriented from the edge representing x to the edge representing y. In this setting, we claim that $\widetilde{N}'$ is obtained from $\widetilde{N}$ by the change of basis that replaces x with $x + y$.

Consider a train track $\boldsymbol{\vartheta}$ in T with associated extended type D structure $\widetilde{N}$. Let x and y be generators of $\widetilde{N}$, and suppose that x and y both have idempotent ι_0 (the case where x and y both have idempotent ι_1 is identical). Let v_{i_x} and v_{j_x} be the marked points on the right and left boundary, respectively, of the square $T \setminus (\alpha \cup \beta)$ associated to x, and let v_{i_y} and v_{j_y} be marked points on the right and left boundary of the square associated to y. Suppose $\boldsymbol{\vartheta}'$ is obtained by connecting the horizontal segments by crossover arrows at each end, one from v_{i_x} to v_{i_y} and one from v_{j_x} to v_{j_y}; note that this is equivalent to move (M2), since we could equivalently add two cancelling arrows at one end of the horizontal strands and then slide one to the other end. The effect of this move on $M_{\boldsymbol{\vartheta}}$ is conjugation by $A_{i_x, i_y} A_{j_x, j_y}$. For each corner path out of v_{i_y} or v_{j_y} we add the corresponding corner path out of v_{i_x} or v_{j_x}, and for each corner path into v_{i_x} or v_{j_x} we add the corresponding corner path into v_{i_y} or v_{j_y}. Each corner path out of v_{i_y} or v_{j_y} corresponds to a term z in $D_I(y)$ for some I, and the new corner path added corresponds to the same term in $D_I(x)$. Thus we add each term in $\partial(y)$ to $\partial(x)$. Each corner path into v_{i_x} or v_{j_x} corresponds to an x term in $D_I(z)$ for some generator in z and some I, so for each such term we add the term y to $D_I(z)$. This is exactly the effect of the change of basis on $\widetilde{N}$ that replaces x with $x+y$. This shows that move (M2) corresponds to a change of basis in the extended type D structure $\widetilde{N}$, as claimed. $\square$

3.5. **Prospectus for simplification.** With these observations in place, our goal is to use the local moves in Figure 22 and moves (M1) and (M2) to simplify a given train track by systematically removing crossover arrows. Below, we describe our general strategy for doing this. In individual examples, it is often easy to simplify the train track by inspection, using the strategy as a guide. In the next section, we will give a formal algorithm showing that the strategy can always be applied.

One might hope to remove all crossover arrows from the train track, resulting in simply a collection of immersed curves. This, it turns out, is not always possible; instead, we will remove all arrows

which do not connect parallel immersed curves. To make sense of this, we must first define what we mean by parallelism. Associated with a crossover arrow, we have a pair of pointed curves (γ, p) and (γ', p'), where $p \in \gamma$ is the tail of the arrow, and $p' \in \gamma'$ is its head. We identify $\pi_1(T, p)$ with $\pi_1(T, p')$ by choosing a path from p to p' which is disjoint from $\alpha \cup \beta$, so after choosing orientations we can view (γ, p) and (γ', p') as defining $\gamma, \gamma' \in \pi_1(T, p)$.

Definition 3.11. Pointed closed curves $(\gamma, p), (\gamma, p')$ are parallel if there is some $\delta \in \pi_1(T, p)$ and $k, k' \in \mathbb{Z}$ such that $\gamma = \delta^k$, $\gamma' = \delta^{k'}$.

Since $k, k' \in \mathbb{Z}$, this definition does not depend on the choice of orientations. Note that the choice of basepoints is very important; for example, if p and p' lie on different components of $\gamma - (\alpha \cup \beta)$, (γ, p) is usually not parallel to (γ, p').

Equivalently, assume that γ is in minimal position with respect to α and β. Following the orientation on γ starting at p determines a (periodic) infinite sequence of signed intersections with α or β, which we can interpret as an infinite periodic word in two variables and their inverses; two pointed curves are parallel if there is some choice of orientations for which they determine the same word.

The definition of parallel pointed curves above suggests a local notion of parallelism. This can be made more precise with the notions of colors and weights that will be introduced in Section 3.6, but for now we simply say two pointed curves are *locally parallel* moving in a given direction (either with the orientations or against the orientations) if the first signed intersection with $\alpha \cup \beta$ on each curve is of the same type. If pointed curves are not locally parallel we say they *diverge* in the given direction. When two curves diverge, one curve diverges to the left and one diverges to the right. In general, curves will be locally parallel for some finite number of intersections and then diverge. Curves which never diverge are parallel in the sense of Definition 3.11; we will sometimes say such curves are *globally parallel* to avoid confusion.
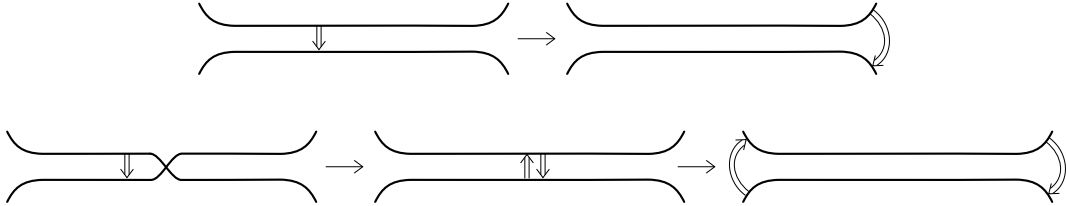

Figure 23. A schematic for removing a single crossover arrow that does not connect globally parallel curves. In the first case we slide the arrow until the curves diverge and the arrow moves left to right, and then remove the arrow. In the second case we resolve a crossing and then remove the two resulting arrows as in the first case.

Suppose a train track consists of a collection of immersed curves and a single crossover arrow. If the crossover arrow does not connect (globally) parallel curves, then it can be removed using the moves (M1), (M2), and the local moves from Figure 22. Since the pointed curves connected by the arrow are not globally parallel, they must eventually diverge in both directions. We begin by sliding the crossover arrow along the curves in the direction of the orientation. If the curves are locally parallel in this direction, we slide the arrow across α or β using (M2); repeating this as necessary we slide the arrow until the two curves diverge. If the arrow goes from the left curve to the right curve when the curves diverge, it can be realized as a clockwise moving crossover arrow around the corner box that passes at least one corner of the square; in this case the arrow can be removed using (M1). If instead the arrow goes from the right curve to the left curve, then we slide the arrow along the curves in the opposite direction until they diverge. Once again, if the arrow goes from the left curve to the right curve at this point of divergence, it can be removed using (M1). If the arrow moves from right to left at both ends, then the curves must cross at some point in between. In this case we slide the arrow to the intersection and resolve the crossing using the local move (e) from Figure 22. The resulting train track has two crossover arrows, but when these slide along the curves in opposite

directions they will be left-to-right, and thus removable, at both ends. This process is summarized in Figure 24. Note that in the second case case we have removed the original arrow and arrived at a collection of immersed curves with no crossover arrows, but the immersed curves are not the ones we started with because a crossing has been resolved.
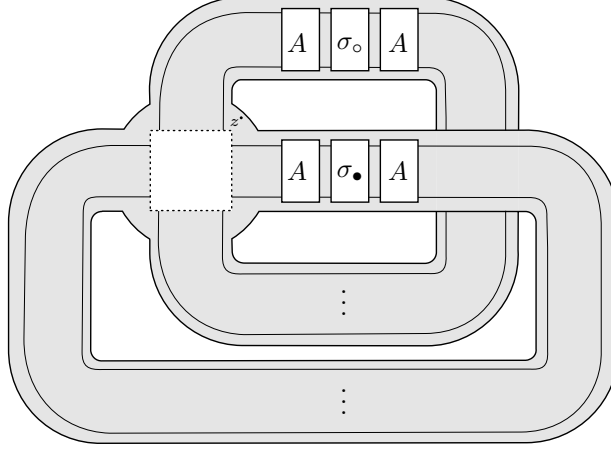
If a train track contains many crossover arrows, any one of them can be removed by the procedure above, provided it does not connect parallel curves. It is plausible, then, that we can remove all crossover arrows that do not connect parallel curves by repeatedly removing one at a time. The key potential issue here is that when multiple arrows are present, removing one of them may involve sliding it past others and this may create more arrows. We need to control the number of new arrows added to ensure that the process of removing arrows one at a time eventually terminates. To deal with this, we will introduce a bookkeeping tool in the form of a weight system for arrows, which measure how far an arrow needs to be pushed before the strands diverge. The strategy is to deal with the easiest to remove arrows (i.e. those with smallest weights) first. These weights will be infinite for crossover arrows connecting globally parallel curves, and we will show that there is an inductive algorithm that, at each step, increases the smallest weight (taken over all arrows) by one while controling the number of arrows added in the process. By compactness of the underlying curves, after a finite number of steps the only remaining arrows must have infinite weight. Figure 25 shows the outcome of this process applied to the train track in Figure 21(c). Note that when performing this procedure by hand, it may not be necessary to explicitly specify this weight system.
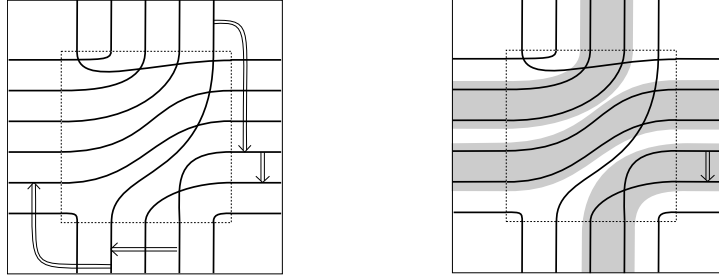

Figure 24. The running example: On the left is the train track from Figure 21(c). The arrow passing the top right corner of the square can be removed by an application of (M1); since it is not the outermost arrow it must first be moved closer to the boundary of the square, which introduces a new arrow, but this can also be removed by (M1). The arrow passing the bottom left corner of the square can also be removed by (M1). The arrow parallel to the bottom side of the square is slid downward, applying (M2) twice before being removed with (M1). The remaining crossover arrow, which connects two points on an immersed curve at which the curve is self parallel, cannot be removed.

3.6. **Arrow sliding algorithm.** We now formalize the strategy suggested above. By Proposition 3.8, we may replace any train track with a configuration of curves and crossover arrows; it remains to show that all crossover arrows can be eliminated unless they connect parallel (pointed) curves.

It will now be useful to make explicit the handle decomposition associated with the punctured torus T, in which the corner box corresponds to a 0-handle; see Figure 26. The corner box contains an arbitrary configuration of embedded arcs (possibly intersecting one another). Because all crossover arrows in the output of Proposition 3.8 move clockwise, any arrows passing a corner of the corner box can be immediately removed by move (M1) and we may assume all other arrows lie in one of the two 1-handles. We previously assumed that the immersed curves consist of purely horizontal or vertical segments outside the corner box (that is, in the 1-handles), but it will be convenient to allow these segments to cross each other so that the endpoints of these segments can be ordered independently on each end of the 1-handle. We collect all crossings into *permutation boxes* near the middle of the handle, which we label $\sigma_\bullet$ and $\sigma_\circ$; they contain a braid-like immersed curve diagram. The boxes labeled A are *arrow boxes*; they contain parallel curve segments joined by an arbitrary configuration of crossover arrows. All of the crossover arrows are contained in these arrow boxes.

Figure 25. The anatomy of a train track in T, according to Proposition 3.8: After applying (M1) moves to remove each of the (clockwise) crossover arrows covering a corner, all of the crossover arrows can be moved into the 1-handles (these are contained in the four boxes labeled A). Each one handle contains a permutation box (labelled $\sigma_\bullet$ and $\sigma_\circ$ to agree with the notation for idempotents) containing an immersed collection of curve segments encoding the permutation. Finally, following previous sections, the dashed box in the 0-handle corresponds to the corner box.
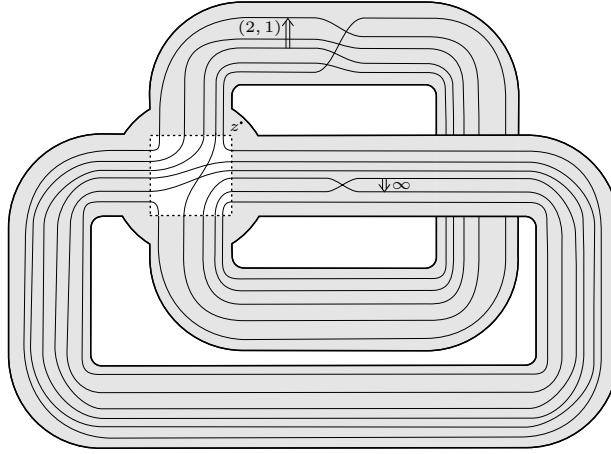
Figure 26. An equivalent view of the train track from Figure 21(c), where the two crossover arrows that passed corners have been removed by a change of basis using (M1) moves. Notice that, in each handle, the arrows and permutations satisfy Lemma 3.15. The crossover arrows have been labelled by their respective weights; comparing Figure 25 shows that the arrow with finite depth $(\check{w}, \hat{w}) = (2, 1)$ can be removed by applications of the moves (M1) and (M2), while the infinite depth arrow persists.

Each strand in an arrow box has two directions: towards the corner box and away from the corner box (toward the permutation box). A specific example is shown in Figure 27.

Weights. Every crossover arrow in an arrow box can be assigned a weight, which is a value in the set $(\mathbb{Z} \setminus \{0\})^2 \cup \{\infty\}$. If the strands on which the arrow starts and ends are globally parallel, the weight is ∞. Otherwise, the weight is a pair $(\check{w}, \hat{w})$, where the integer $\check{w}$ (w-to) is defined as follows. Follow the strands on which the arrow begins and ends, starting by moving towards the corner box; at some point, the two strands will diverge (that is, leave the corner box through different edges). The non-zero integer $\check{w}$ indicates that the arrow enters the corner box $|\check{w}|$ times before the strands diverge; the sign of $\check{w}$ records whether the arrow runs clockwise ($+$) or counterclockwise ($-$) around

the corner box when this happens. The non-zero integer $\hat{w}$ (w-from) is defined similarly by following the strands in the opposite direction, initially traveling away from the corner box, until they diverge.
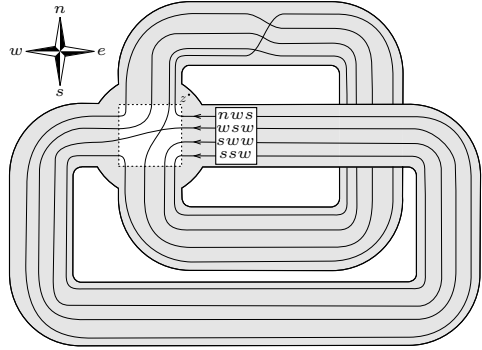
Definition 3.12. The *depth* of an arrow is defined to be $\min\{|\check{w}|, |\hat{w}|\}$, or ∞ if the weight is ∞. We will refer to a collection of immersed curves along with a collection of crossover arrows as a *curve configuration*; the depth of a curve configuration is the minimum depth of all its crossover arrows, or ∞ if there are no crossover arrows.

By compactness of the curves, if an arrow has finite depth, then there is an absolute bound on that depth depending only on the number of strands in the configuration. Our aim is to prove:

Proposition 3.13. *A curve configuration of depth m is equivalent to another curve configuration with the same number of strands and depth no less than $m + 1$.*

By induction, it will follow that any curve configuration is equivalent to an infinite depth curve configuration, that is, one in which the only crossover arrows run between parallel strands. Some additional structures and preliminary lemmas are required before proving Proposition 3.13.

Colors and orders. We now define a partial order on the set of strands in an arrow box. To do this, observe that any strand a in the box lifts to a quasi-geodesic in the universal cover of T. By following the lift of a in the direction that initially moves towards the corner box, we get a ray γ_a in the cover. More explicitly, the universal cover of T is a ribbon graph, which we can identify with a regular neighborhood of a regular 4-valent tree properly embedded in the hyperbolic plane $\mathbb{H}^2$. The path γ_a takes in this tree can be described by an infinite string in the letters n, e, s, w recording which edge of the corner box the strand exits through each time it passes through the corner box, as illustrated in Figure 28. By compactness of the curves in the configuration, this string is clearly periodic. Such a string associated with a is called the coloring of a. Said another way, retracting the universal cover of T to the Cayley graph

Figure 27. A depth 3 coloring in an arrow box; these strands are (lexicographically) ordered top-to-bottom since $n < w < s$ for the first letter and, for the two strands which begin with s, $w < s < e$ for the second letter.

for $\pi_1(T) \cong \langle n, e \rangle$, these colors corresponds to reduced words in the free group where $s = n^{-1}$ and $w = e^{-1}$.

Informally, the partial ordering on strands is given by $a < b$ if the strand a is on the left relative to b whenever the strands diverge. Equivalently, $a < b$ means that if we take a crossover arrow from a to b and push it in the positive direction until the strands diverge, the arrow will point in a counterclockwise direction when the strands diverge. This can be made precise in two ways. First, the Gromov boundary of the 4-valent tree in $\mathbb{H}^2$ can be naturally identified with a Cantor set $C \subset S^1 = \partial \mathbb{H}^2$. Since γ_a does not pass through the initial branch of the tree corresponding to the arrowbox it starts in, the set of possible endpoints of γ_a is the intersection of C with an interval $I \subset S^1$ of length $3\pi/2$. Thus there is a natural order on the set of possible endpoints defined by declaring $x < y$ if we can go counterclockwise from x to y in I. We define $a < b$ if $\gamma_a(\infty) < \gamma_b(\infty)$. This ordering can also be described as a lexicographic order on colorings of strands, with the understanding that the order on the letters differs in each arrow box and thus in each position in the word. Each time the ray γ_a passes through an arrow box and into the corner box, there are only three ways it can leave the corner box since backtracking along the 1-handle it came from is not allowed. At each pass through the corner box, we order the choices from rightmost to leftmost. In terms of letters, no backtracking means that, for instance, the letter n is never followed by s. There are always three possible choices for the $(k + 1)$st letter of a coloring, depending on kth letter, and these three choices inherit a linear ordering from the cyclic ordering (n, w, s, e) by

removing the disallowed letter. We then define $a < b$ if the color of a is less than the color of b in the lexicographical order—that is, if the first k letters of the colors agree for some k and the $(k+1)$st letter for a is less than the $(k+1)$st letter for b using the ordering obtained by disallowing the opposite of the kth letter.
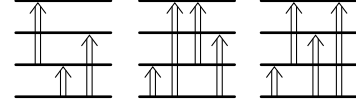
Note that $<$ is a partial order, but a and b have an ordering relationship unless their colors agree; this happens precisely when the pointed curves starting with the strands a and b are globally parallel in the sense of Definition 3.11. We can define a sequence of weaker partial orders $<_m$ on the set of strands in an arrow box using the truncation of the coloring of a to the first m letters, which we call the *depth m coloring of a*. We say that $a <_m b$ if the depth m color of a is less than the depth m color of m using the lexicographical ordering described above. The partial order $<_m$ does not distinguish between strands for which the rays γ_a and γ_b are locally parallel and do not diverge before passing through the corner box at least m times. It is easy to see that $a <_m b$ implies $a <_{m+1} b$ and that $<_m$ agrees with $<$ for sufficiently large m. By a slight abuse of notation, we write $a \leq_m b$ to mean that $a < b$ or the depth m colors of a and b agree, noting that $a \leq_m b$ does not imply $a <_m b$ or $a = b$.

Ordering crossover arrows in an arrowbox. In an arrowbox, where all strands are parallel, crossover arrows have a well-defined length, measured by how many strands they cross. A collection of arrows that point in the same direction can be sorted by length, as follows:

Lemma 3.14. *A configuration of crossover arrows that all point up is equivalent to a new configuration in which all the arrows point up and are sorted by length, in the sense that shorter arrows lie to the left of longer arrows.*

Proof. Let $n+1$ be the total number of strands in the arrowbox. We say a configuration is *k-sorted* if all arrows of length n appear on the right, followed by all arrows of length $n-1$, then all arrows of length $n-2$, *etc.* up to arrows of length $n-k$. The length $< n-k$ part of a configuration is the configuration obtained by deleting all arrows of length greater or equal to $n-k$.

We prove by induction on k that any configuration is equivalent to a k-sorted configuration whose length $< n-k$ part is the same as the length $< n-k$ part of the original configuration. The base case $k=0$ is clear, since an arrow of length n slides freely past any other arrow. For the general case, given a configuration C_0, we first apply the induction hypothesis to find a k-sorted configuration C_1 whose length $< n-k$ part agrees with C_0. If there is no arrow of length

Figure 28. Ordering arrows from left to right, following the steps used in the proof of Lemma 3.14.

$n-k-1$, then the configuration is $(k+1)$-sorted, and we are done. Otherwise choose the length $n-k-1$ arrow which is farthest to the right among those arrows of C_1 which are out of order, and slide it to the right, past the first arrow it encounters, to obtain a new configuration C_2.

If this slide did not create a new arrow (as in Figure 22(a)) C_2 is still k sorted and has the same length $< n-k-1$ part as C_1. If the slide created a new arrow (as in Figure 22(d)) C_2 has the same length $< n-k-1$ part as C_1 but is not k-sorted. In this case, we apply the induction hypothesis to find a configuration C_3 which is k-sorted and has the same length $< n-k$ part as C_2. (In particular, the position of the length $n-k-1$ arrow has been switched, and we have not added any arrows of length $n-k-1$). Repeating this process, we eventually arrive at a $(k+1)$-sorted configuration whose length $< n-k-1$ part agrees with that of C_0. $\square$

On a one-handle, the curve configuration consists of two arrowboxes and one permutation box, as shown in Figure 26. We introduce the notational shorthand $[A_1, \sigma, A_2]$ to specify this data. We say that $[A_1, \sigma, A_2]$ and $[A'_1, \sigma', A'_2]$ are equivalent if they represent equivalent train tracks in the 1-handle, in the sense that the (mod 2) counts of paths from any endpoint at one end of the 1-handle to any endpoint at the other end agree. Note that when the local moves in Figure 22 are applied

within a 1-handle the resulting configurations are equivalent. In addition to sliding arrows past each other within each arrow box, as in Lemma 3.14, we can also slide arrows through the permutation box from one arrow box to the other, and we can resolve a crossing in the permutation box using the move in Figure 22(e), which changes the permutation σ as well as the arrow boxes.

Our next goal is to show that we can replace any configuration $[A_1, \sigma, A_2]$ with an equivalent configuration $[A'_1, \sigma', A'_2]$ in which the crossover arrows are nicely arranged with respect to a chosen ordering of the strands on each end. Fix an ordering $<$ on the strands in each arrowbox. We say that the arrow boxes are sorted with respect to $<$ if $b < a$ whenever there is a crossover arrow from a to b.

Lemma 3.15. *Any configuration $[A_1, \sigma, A_2]$ in a 1-handle is equivalent to a configuration $[A'_1, \sigma', A'_2]$, where the arrow boxes A'_1 and A'_2 are sorted with respect to the fixed ordering $<$. Moreover, if the original configuration has depth m, then the depth m colorings on the old and new configurations are the same, and the new configuration has depth $\geq m$.*

We remark that the configuration predicted by Lemma 3.15 is not unique, as can be seen for the two-strand bundle with a pair of opposite arrows pictured in Figure 30. In general, when applying Lemma 3.15 we choose the ordering $<$ to be the color ordering (or, more precisely, any total ordering on strands which is consistent with the partial order coming from colors). Note that in the result-ing configuration, the order may no longer coincide with the color

Figure 29. Three equivalent con-figurations; the pair on the right both have the form guaranteed by Lemma 3.15.

ordering, since resolving crossings in the permutation box changes the colors of the strands. How-ever, the last part of the lemma implies that the depth m partial ordering $<_m$ is the same for both configurations.

Proof of Lemma 3.15. We begin by making some simplifications. First, if the original configuration has depth m, we divide the strands into equivalence classes (bundles). Two strands belong to the same equivalence class if they run parallel for $m-1$ steps in both directions. The hypothesis that the configuration has depth m means that any two strands joined by a crossover arrow are in the same equivalence class. Crossover arrows from one bundle do not interact with stands from a different bundle, so we can operate on each bundle separately; thus we assume without loss of generality that we are operating on a single bundle of strands.

Next, we claim that if the statement holds for one ordering $<$ on the strands in each arrow box, it holds for any other ordering $<'$. To see this, note that $<'$ is related to $<$ by some permutation τ of the strands. We define A^τ to be the result of permuting the strands according to τ and carrying the crossover arrows along, so that an arrow from strand i to strand j in A becomes an arrow from strand $\tau(i)$ to strand $\tau(j)$ in A^τ; we refer to A^τ as the conjugation of the arrow box by τ. By applying a homotopy to the strands in A while fixing their endpoints, we can see that the arrow box A is equivalent to the arrow box A^τ with a collection of crossings added on either end encoding the permutation τ on the left and τ^{-1} on the right. We can absorb the crossings on the inside of A^τ (that is, toward the permutation box) into the permutation box, ignore the crossings on the outside of A^τ (toward the end of the 1-handle), and apply the Lemma with $<$ to the remaining configuration. Undoing the homotopy, we see that this configuration being ordered with respect to $<$ is equivalent to the original configuration being ordered with respect to $<'$.

From now on, we will assume that the strands are ordered from top to bottom in A_1, and bottom to top in A_2, so that we are aiming to have all arrows in A'_1 pointing up, and all arrows in A'_2 pointing down. We begin by arranging the bundle so that all crossings between strands are on the right and all crossover arrows are on the left. It is enough to show that the right-most down arrow can be moved past every up arrow, and to the other side of the crossings (at the expense of possibly altering the permutation between strands). An example is shown in Figure 31. Using Lemma 3.14, arrange

the collection of up arrows so that the shortest arrows appear first. We will induct on the length of the down arrow immediately to the left of the collection of up arrows.
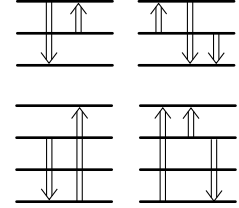
Suppose that the right-most down arrow has length 1. This arrow slides past all length 1 up arrows, with the exception of the possibility illustrated in Figure 30, namely, the length 1 down arrow meets a length 1 up arrow between the same two strands. If this occurs, we replace the pair with a single up arrow and a new crossing between the strands. Effectively, the down arrow is replaced by a crossing and this new crossing slides to the right to compose with (and alter) the permutation. Notice that this will, in general, have a non-trivial effect on the remaining up arrows encountered, however this can only increase their length (if a starting or ending point is on one of the strands in question) and not, in particular, switch their direction.
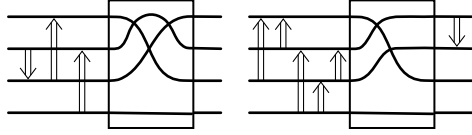
Figure 30. Two equivalent views of the same bundle. The configuration on the left is in the form desired at the beginning of the algorithm; the configuration on the right shows the result of an application of the base case in the induction.

If the length 1 down arrow does not meet a length 1 up arrow between the same strands, it slides freely past all remaining length $k > 1$ arrows. New arrows may be produced in the process by composition, but these will have length $k - 1$ and be (additional) up arrows; compare Figure 31. Finally, we slide the arrow through the crossings specified by the permutation σ. There are two cases: Either the arrow remains a down arrow on the other side of these crossings and we are done, or the arrow switches to an up arrow. In the latter case there must be a crossing between the strands that caused the swicth, which we resolve (see, for example, Figure 31). The result alters the permutation but produces an up arrow on the left and a down arrow on the right. This completes the base case for induction.

Now suppose that the right-most down arrow has length n; our induction hypothesis is that arrows of length less than n can be moved to the right of the up arrows. As before, the up arrows collected to the right are ordered by length. There are three groups in this collection: the short arrows (those of length less than n), the length n arrows, and the long arrows (those of length greater than n). Sliding the down arrow past the short arrows may produce compositions (see Figure 32), but these new down arrows have length less than n and are dealt with by the induction hypothesis. To pass the length n arrows, we proceed as in the base case. That is, the arrow passes freely unless it meets an up arrow joining the

Figure 31. Compositions with short (above) and long (below) arrows.

same strands. In this case the arrow is replaced by a crossing, which slides to the right; no new down arrows are created as the remaining arrows are all long. Finally, if the arrow passes to the long arrows the arrow proceeds to the permutation (again without producing new down arrows, compare Figure 32) and we resolve a crossing in the permutation if needed.

Finally, we consider the effect of the operations we have performed on the weights of crossover arrows and orderings of strands. In the course of the proof, we have added transpositions between strands belonging to the same bundle. Strands belonging to the same bundle have the same depth $m - 1$ coloring, and any strand must pass at least once through the corner box before it goes over one of the new transpositions. Thus the depth m coloring of all strands is unchanged. From this, it follows that any arrow with depth $\geq m$ in the old configuration has depth $\geq m$ in the new configuration as well. In the course of the proof, we may have created many new crossover arrows, but these all run between strands in the same bundle, so they have depth $\geq m$ as well. $\square$

Proof of Proposition 3.13. The argument is divided into four steps.

Step 1: Remove arrows with $\check{w} = -m$. Order to depth m by applying Lemma 3.15 to each one-handle, with respect to any extension of the partial ordering $\leq_m$ to a total ordering. Since the original configuration has depth m, the ordering $\leq_m$ is unaffected by this operation, and it does not matter which handle we order first. The new configuration has depth m and is sorted with respect to $\leq_m$. This means that if we take an arrow with $|\check{w}| = m$ and push it m steps in the direction towards the corner box (to where its endpoints diverge), it will remain positively oriented, and thus will run clockwise around the corner box. Thus any such arrow actually has $\check{w} = m$, and there are no arrows with $\check{w} = -m$ in the new configuration.
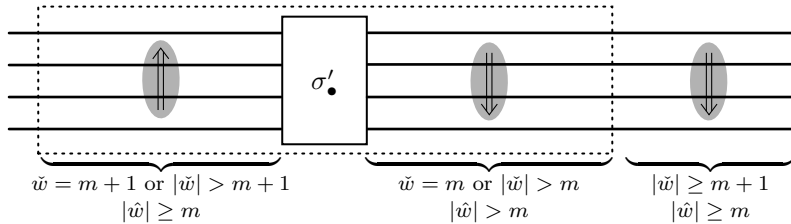
Step 2: Remove arrows with $\check{w} = m$. Fix an arrow box. If it contains any arrows with $\check{w} = m$, choose the one among them which is closest to the corner box. We can remove this arrow by pushing it in the positive direction until its ends diverge. As we do so, we may encounter other arrows with ends on one of the strands which our arrow runs between. We claim that these arrows can be pushed along in front of our chosen arrow (like a pile of snow accumulating in front of a snowplow) without their ends diverging until the chosen arrow reaches the point where its ends diverge. To see this, note that every arrow in the arrow box which lies between the corner box and our chosen arrow has $|\check{w}| > m$, so it can be pushed along m steps without its ends diverging. Similarly, arrows which we encounter in any subsequent arrow box have depth at least m, so they can be pushed $m - 1$ steps in any direction without diverging. Other arrows which these arrows would have to pass can likewise be pushed along (they need to move at most $m - 1$ steps). At the end of the process, we can remove the chosen arrow, and then just push all other arrows back to where they started. The net result is that we have removed the chosen arrow without making any other changes to the curve configuration. Repeating this process removes all arrows with $\check{w} = m$.

Step 3: Remove arrows with $\hat{w} = -m$. For a given arrow box A, we will change the configuration so as to eliminate all arrows with $\hat{w} = -m$ from A. To do this, we first apply Lemma 3.15 to the region consisting of the arrow box and its adjacent permutation with ordering $\leq_{m+1}$. This region is shown in Figure 33.

By Lemma 3.15, the depth m colorings in the old and new configurations agree. Moreover, the depth $m + 1$ colorings on A are the same in both configurations. To see this, consider the number of steps a strand leaving the arrowbox towards the corner box takes to return to it. If it returns after one step through the corner box, it must re-enter through the right-hand side of the figure and exit through the left. Since all crossover arrows in the original arrowbox have $|\check{w}| \geq m+1$ (Steps 1 and 2), any new transpositions we added in the course of applying the lemma will not change the next m labels in the coloring. Thus the first $m + 1$ labels in the original coloring will remain unchanged. A similar argument applies if the strand takes two or more steps to return to the arrowbox.
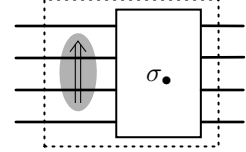
Figure 32. Lemma 3.15 applied to the dashed region; arrows have $|\check{w}| \geq m+1$ and $|\hat{w}| \geq m$.

$$\underbrace{}_{\substack{\check{w} = m + 1 \text{ or } |\check{w}| > m + 1 \\ |\hat{w}| \geq m}} \qquad \underbrace{}_{\substack{\check{w} = m \text{ or } |\check{w}| > m \\ |\hat{w}| > m}} \qquad \underbrace{}_{\substack{|\check{w}| \geq m + 1 \\ |\hat{w}| \geq m}}$$

Figure 33. A summary of weights after Lemma 3.15 is applied to the region of the horizontal strands in the dashed box. Weights outside the dashed box are inherited from the previous step.

At this point, the configuration is as shown in Figure 34. It is sorted to depth $m + 1$ on the left, and to depth m on the right. It follows from Lemma 3.15 that every arrow in the dashed box has depth greater or equal to m. In fact, every arrow in the dashed box has $\check{w} \geq m + 1$ on the left of the permutation, and $\hat{w} \geq m + 1$ on the right. To see this, note that when we apply Lemma 3.15, the strands in each bundle used in the proof run parallel for m steps to the left. Finally, since the configuration is sorted, we see that $\check{w} \neq -(m + 1)$ for arrows on the left, and $\check{w} \neq -m$ for arrows on the right. In summary, the weights in the new configuration are as shown in the figure.

In the process of applying Lemma 3.15, we may have created some new arrows with $\check{w} = m$ in the right-hand side of the dashed region. Since the right-hand arrowbox in the 1-handle is sorted to depth m, these can now be removed just as in Step 2.

Consider the arrow in A which is closest to the corner box. This arrow has $|\check{w}| \geq m + 1 \geq 2$, so we can slide it through the corner box to obtain a new configuration. In its old position, the arrow had $|\hat{w}| \geq m$, so the new arrow will have $|\check{w}| \geq m + 1$. Similarly, the old arrow had $\check{w} = m + 1$ or $|\check{w}| > m + 1$, so the new arrow will have $\hat{w} = m$ or $|\hat{w}| > m$. Repeating this operation for each arrow in A, we eventually arrive at a depth m configuration in which we have slid every arrow out of A.

Summarising our progress to this point: we have removed all arrows with $\check{w} = -m$ or $\hat{w} = -m$ from our chosen arrowbox A without adding any new arrows with $\check{w} = -m$ or $\hat{w} = -m$ in any of the other arrowboxes. Therefore, by repeating this sequence of steps on each arrowbox, we arrive at a depth m configuration of curves in which there are no crossover arrows with $\check{w} = -m$ or $\hat{w} = -m$.

Step 4: Remove arrows with $\hat{w} = m$. There may remain arrows for which $\hat{w} = m$; the last step in the proof of Proposition 3.13 is to show that we can eliminate all arrows of depth m without changing the rest of the configuration. This proceeds in exactly the same way as in Step 2. Fix a 1-handle, choose a direction $\rightarrow$ on it, and consider all crossover arrows which have weight $w_\rightarrow = m$ when sliding in this direction. All the other arrows in the 1-handle have $|w_\rightarrow| \geq m + 1$, so we can slide the arrows with $w_\rightarrow = m$ off one at a time without changing the weights of any other arrow, just as we did in Step 2. Repeating this operation for each 1-handle and each direction, we arrive at a configuration in which every arrow has depth greater or equal to $m + 1$. $\square$

3.7. **The proof of Theorem 1.5.** We are now positioned to prove the theorem that is the focus of the section.

Fix a reduced, extendable type D structure and express it as an immersed train track. By Proposition 3.8 this may be expressed as a curve configuration, a schematic for which is given in Figure 26. Then the algorithm detailed in Section 3.6 ensures that this curve configuration can be simplified to one in which the only crossover arrows connect parallel curves in the marked torus. Indeed, the original type D structure is isomorphic as an $\mathcal{A}$-module to the type D structure determined by this simplified curve configuration. It is convenient to introduce terminology for a train track of this form.

Definition 3.16. A *curve-like train track* is an $\mathcal{A}$-train track that has the form of immersed curves together with crossover arrows connecting parallel curves.

Figure 34. Equivalent configurations.

It remains to associate a local system with any collection of parallel curves that arise. Consider such a collection of n parallel curves, and notice that we can replace, without loss of generality, any crossings between two curves with a triple of alternating crossover arrows as in Figure 35; compare Figure 22(e). Moreover, by applying the move (M2) we may assume that all of the arrows are contained in the 0-handle and run between curves whose endpoints are on two different sides of the corner box.

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

Figure 35. Extracting a local system from a bundle of curves, where the generators run from top to bottom.

Interpreting this configuration as (part of) a train track, we take the segments on each side of the collection of arrows as generators and form a $n \times n$ matrix (x_{ij}) where the entry x_{ij} is the mod 2 count of oriented paths from x_i to x_j. The result is a local system of dimension n over the immersed curve γ carrying the bundle of curves. Note that to define this local system we must 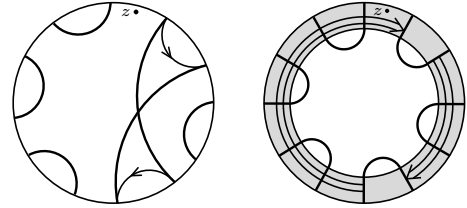choose an orientation on γ (for the example in Figure 36 we count paths oriented left to right); choosing the opposite orientation corresponds to inverting the matrix. □

3.8. **An aside on type D structures associated with general surfaces.** Thus far we have focused on type D structures associated to the torus algebra $\mathcal{A}$, since these are the objects relevant for bordered Heegaard Floer homology of manifolds with torus boundary. However, the proof above applies more broadly to classify extendable type D structures associated with more general surfaces in terms of immersed curves in those surfaces.

We begin by defining algebras associated to a disk with a single marked point. For a given positive integer m we can associate a quiver with the marked disk: there are m vertices $\{0, \ldots, m-1\}$ placed on the boundary of the disk and m edges $\{\rho_0, \ldots, \rho_{m-1}\}$ oriented clockwise along the boundary. We assume that the marked point is arbitrarily close to the boundary of the disk and is next to the edge ρ_0. Fix the target of ρ_i to be the point i; below it will be helpful to distinguish between the vertex i in the quiver and the associated idempotent ι_i in the path algebra of the quiver. We follow the usual convention of writing ρ_I for the product associated with a cyclically-ordered string I from $\{1, 2 \ldots, m-1, 0\}$. The path algebra of this quiver determines two algebras that are of interest to us: $\widetilde{\mathcal{B}}$ is the result of quotienting by $\rho_I = 0$ for any string I containing more than one 0, while $\mathcal{B}$ is the result of quotienting by $\rho_I = 0$ for any string I containing at least one 0. The algebra $\widetilde{\mathcal{B}}$ contains a distinguished central element $U = \sum_{|I|=m} \rho_I$.

Now a type D structure over $\mathcal{B}$ may be defined as in Section 2.1. In the same way, we can define the notion of an extension, which will be a module over $\widetilde{\mathcal{B}}$. These (extended) type D structures can be encoded by train tracks in the marked disk, where generators with idempotent ι_i give rise to primary switches on the boundary of the disk in a neighborhood of the vertex i and terms in the differential are encoded by paths through the disk. An example is shown in Figure 37 (where, as usual, two-way track segments are recorded as unoriented edges). As such, in this setting, the type D structure and a choice of extension for it can be represented by a matrix M over $\mathbb{F}[U]/U^2$ satis-

Figure 36. Equivalent train tracks in the marked disk, according to Lemma 3.9. Here $m = 12$ and there is one primary vertex for each idempotent.

fying the hypotheses of Lemma 3.9, and hence may be expressed as $M = P\bar{M}P^{-1}$ where P is a composition of elementary matrices of the form $A_{i,j}^U$ or of the form $A_{i,j}$ with $i < j$, and M is as in Proposition 3.8. Following the proof of Proposition 3.8, this new form may be interpreted in terms of the associated train tracks in the disk as follows: up to equivalence, all train tracks for which $M^2 = UI_{2n}$ can be represented by (1) a collection of unoriented track segments connecting primary switches in the boundary of the disk; and (2) a collection of clockwise-moving crossover arrows confined to a small collar neighborhood of the boundary of the disk. For the example in Figure 37, the reader can check that $P = A_{6,1}^U A_{2,5}$ (where determining M and $\bar{M}$ is left as an exercise), leading to the train track on the right of the figure.

For $m = 4g$ and a specified handle attachment of $2g$ 1-handles, we can define an algebra associated with a punctured genus g surface. We denote this algebra by $\mathcal{A}_F$, where F is the surface along with the specified handle decomposition. Let ι_i and ι_j be idempotents of $\mathcal{B}$ corresponding to points $i < j$ in the boundary of the disk. Adding a 1-handle $[-\epsilon, \epsilon] \times I$ to the disk so that $(0,0)$ is identified with i and $(0,1)$ is identified with j gives rise to a modified quiver: the points i and j are identified to form a quotient graph, but new paths are disallowed by setting $\rho_i \rho_{j+1} = 0$, and $\rho_j \rho_{i+1} = 0$. The new algebra $\mathcal{A}_F$ is a subalgebra of $\mathcal{B}$ obtained by replacing the idempotents ι_i and ι_j with the single idempotent $\iota_i + \iota_j$. In other words: the points of the relevant idempotents are identified, while any and all possible *new* paths created are set to zero. The extended algebra $\widetilde{\mathcal{A}}_F$ is defined analogously as a quotient of $\widetilde{\mathcal{B}}$. (The reader may recognize the geometric significance of the idempotent subring of $\mathcal{A}_F$ in general: the Grothendieck group of the appropriate associated category of modules is completely determined by this ring; see [29].)

Note that when $m = 4$ and F is the punctured torus, with 1-handles identifying idempotents ι_0 with ι_2 and ι_1 with ι_3, the algebra $\mathcal{A}_F$ defined above coincides with the torus algebra $\mathcal{A}$ as defined in Section 2.1, and $\widetilde{\mathcal{A}}_F$ agrees with $\widetilde{\mathcal{A}}$. In this case we identify $\iota_\bullet = \iota_0 + \iota_2$ and $\iota_\circ = \iota_1 + \iota_3$, and the relations from the 1-handle attachments account for the familiar $\rho_3 \rho_2 = \rho_2 \rho_1 = 0$ in the definition of $\mathcal{A}$. When $m = 4g$ and F is a punctured genus g surface, the algebra $\mathcal{A}_F$ is the one-moving-strand part of the strand algebra associated to F in [39]. More specifically, the algebra $\mathcal{B}$ is the algebra $\mathcal{A}(m, 1)$ defined in [39, Section 3.1] and $\mathcal{A}_F$ coincides with $\mathcal{A}(\mathcal{Z}, -g+1)$ defined in [39, Section 3.2], where $\mathcal{Z}$ is a pointed matched circle representing the given handle decomposition of F.

In fact, we need not restrict to surfaces with one zero handle and one boundary component: we can construct algebras $\mathcal{A}_F$ and $\widetilde{\mathcal{A}}_F$ for other surfaces F admitting a prescribed (finite) 0- and 1-handle decomposition, whenever each 0-handle carries a marked point (i.e. there is a distinguished ρ_0 for each 0-handle). We begin with the algebra $\oplus_{i=1}^n \mathcal{B}^i$ associated with the collection of 0-handles, that is, the path algebra associated with a disjoint union of n cycles. Each 1-handle attachment identifies some point i_a (corresponding to $\iota_i^a \in \mathcal{B}^a$) with i_b (corresponding to $\iota_i^b \in \mathcal{B}^b$) and gives rise to a quotient graph; this quiver has new relations imposed by $\rho_i^a \rho_{j+1}^b = 0$ and $\rho_j^b \rho_i^a = 0$ (just as in the case of a single 0-handle). Having attached all of the handles in the description of F, and appropriately quotienting the associated quiver at each stage, we arrive at an algebra $\mathcal{A}_F$ (see Figure 38 for an explicit example). A similar construction gives $\widetilde{\mathcal{A}}_F$. There is are close ties here with work of Zarev; see in particular the examples considered in [52, Section 9].

To describe an extendable type D structure over $\mathcal{A}_F$ for general surface F with n 0-handles, we can first describe a type D structure over $\oplus_{i=1}^n \mathcal{B}^i$: the underlying vector space $\oplus_{i=1}^n V^i$ has left-action by the idempotents $\mathcal{I}(\oplus_{i=1}^n \mathcal{B}^i) = \oplus_{i=1}^n \mathcal{I}(\mathcal{B}^i)$ so that V^i is a type D structure associated with the ith 0-handle. The essential observation is that any quotient identifying i_a with i_b in the quiver is compatible with a left-action of the new idempotent subring where ι_i^a and ι_i^b are replaced by $\iota_i^a + \iota_i^b$, and hence gives a well-defined type D structure over the (subalgebra) associated with the quotient quiver. Notice that the matrices M^i describing this type D structure are data that remains associated with the 0-handles in the decomposition of F. In particular, while the type D structures change (the underlying vector space changes, as well as the algebra), the M_i are unchanged at every stage. We emphasise that, in the course of this construction, the relevant type D structures are *not* obtained by a sequence of quotients as differential modules, rather, the vector spaces are quotiented in a way compatible with a change to the idempotent subring. Moreover, such quotients are only possible when $\dim(\iota_i^a V^a)$ and $\dim(\iota_i^b V^b)$ agree. Geometrically, the train tracks representing the type D-structures in each 0-handle extend over the 1-handles as parallel lines connecting the relevant generators in the quotient; see Figure 38. This is the key observation required in order to establish Theorem 3.17, below, treating type D structures associated with general surfaces.

We note that the move (M1), along with the local moves for train tracks, apply in precisely the same way to train tracks in a marked disk, where the role of the corners of the square are played by points on the boundary of the disk between two adjacent quiver vertices (and outside of the neighborhoods
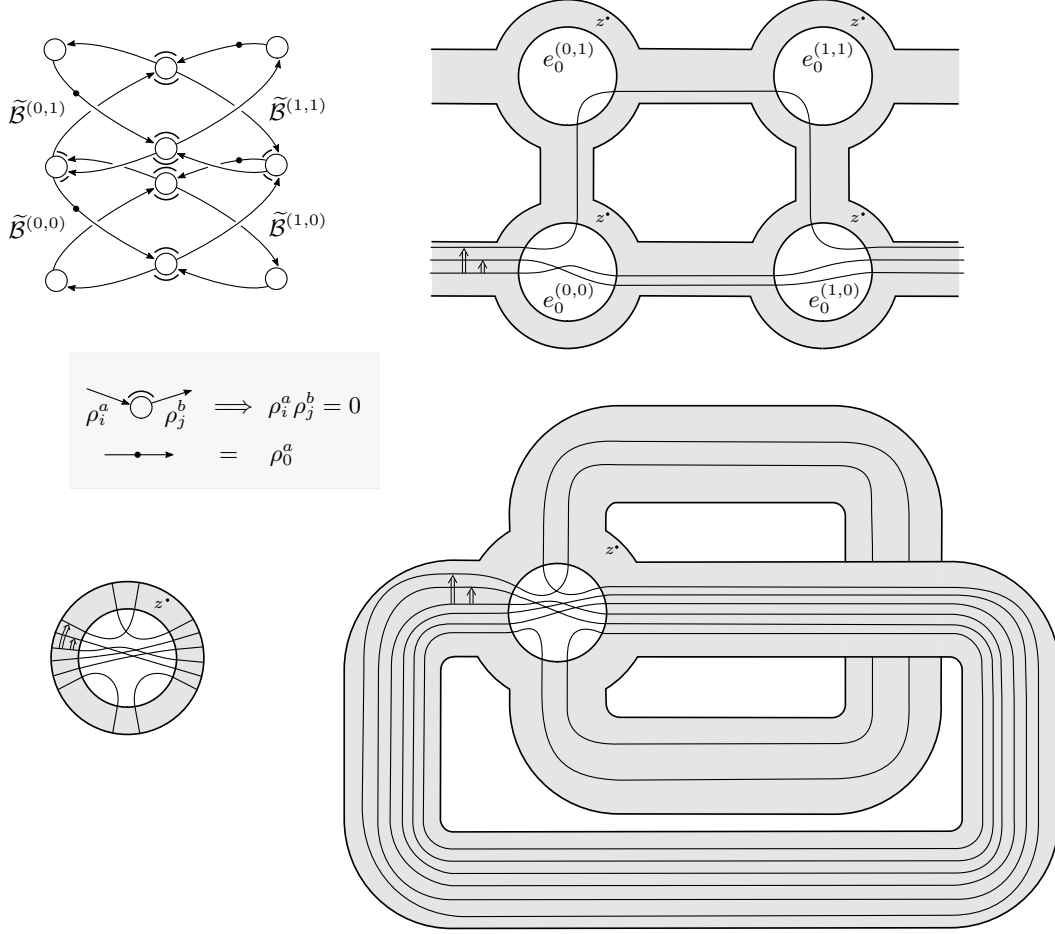
Figure 37. Three (related) extended type D structures: over $\widetilde{\mathcal{B}}$ (lower left) where $m = 4$ has been omitted from the notation; over $\widetilde{\mathcal{A}}$ (lower right); and over the algebra described by the quiver in the top left (upper right). The associated surface in this last example is made up of four 0-handles and six 1-handles (it should be viewed in the cylinder $S^1 \times \mathbb{R}$), and each 1-handle attachment gives rise to a quotient of the quiver, with relations as described by the key in the shaded box. The end result is a subalgebra of $\widetilde{\mathcal{B}}^{(0,0)} \oplus \widetilde{\mathcal{B}}^{(1,0)} \oplus \widetilde{\mathcal{B}}^{(0,1)} \oplus \widetilde{\mathcal{B}}^{(1,1)}$. This last example covers the example associated with the torus. The reader can check that the matrices underlying the first two examples are identical, while the last example requires 4 matrices (one for each 0-handle).

of these vertices in which all primary switches lie). Thus by Proposition 3.8 the train tracks within each 0-handle can be represented by arcs with clockwise moving crossover arrows, and any crossover arrows passing a corner can be immediately removed up to a change of basis in the corresponding module. These steps make sense for modules over $\mathcal{B}$ before passing to the quotient $\mathcal{A}_F$ (i.e. before attaching 1-handles). In contrast, the move (M2) does not make sense without 1-handles, but once we pass to train tracks representing type D structures over $\mathcal{A}_F$ the obvious analog of (M2) is allowed: up to an appropriate change of basis, we may slide a crossover arrow near the boundary of a 0-handle across a 1-handle into the 0-handle on the other end.

We conclude by observing that the arrow sliding algorithm described in Section 3.6 applies to train tracks in arbitrary surfaces. After removing crossover arrows which pass a corner of a 0-handle, all crossover arrows can be pushed into a 1-handle. The arrow sliding algorithm takes place primarily in one 1-handle at a time, where we rearrange a depth m configuration of arrows in that 1-handle to remove all depth m arrows. The inductive step follows from repeating this procedure in each 1-handle. The only difference for general surfaces is that the colors of strands can not be written as words in the letters $\{n, e, s, w\}$, but rather colors encode a path through the Cayley graph for

$\pi_1(F)$. An ordering on colors, and thus a partial ordering on strands, can still be defined as before. So, while we are primarily interested in the torus algebra for this work, we have in fact shown:

Theorem 3.17. *Let $\mathcal{A}_F$ be the algebra associated with a marked surface F equipped with a decomposition into finitely many 0- and 1-handles. Then any extendable type D structure over $\mathcal{A}_F$ may be realised as a collection of immersed curves in F decorated with local systems.*

This recovers a result due to Haiden, Katzarkov, and Kontsevitch [18] via rather different techniques.

4. Decorated immersed curves as type D structures

The previous section shows how an extended type D structure over $\widetilde{\mathcal{A}}$ can be interpreted as a collection of immersed curves decorated by local systems in the marked torus T. The goal of this section is to show that this in fact gives a bijection between the set of homotopy equivalence classes of extendable type D structures and the set of decorated immersed multicurves in the marked torus. Moreover, we show that the pairing on extendable type D structures coming from the box tensor product is given by the intersection Floer homology of the corresponding curves.

4.1. **Intersection Floer homology.** In [1], Abouzaid gave a combinatorial definition of the intersection Floer homology of immersed curves in a symplectic surface (*cf.* [11] for a detailed discussion in the embedded case, and section 2 of [5] for a recent summary). We will use a slight variation on this construction: rather than working with coefficients in a Novikov ring, as is standard in symplectic geometry, we use the notion of an admissible diagram, which was introduced by Ozsváth and Szabó [43] in the context of Heegaard Floer homology (see also Kotelskiy [31]).

Let S be a (possibly noncompact) orientable surface with $\chi(S) < 0$, so the universal cover $\widetilde{S}$ is homeomorphic to $\mathbb{R}^2$. A *curve* γ in S is either an immersed closed curve $\gamma\colon S^1 \to S$ that represents a nontrivial element of $\pi_1(S)$ (a *loop*), or a properly immersed arc $\gamma\colon \mathbb{R} \to S$ (an *arc*). If γ is an arc it lifts to an arc $\widetilde{\gamma}$ in $\widetilde{S}$. If γ is a loop, we consider the composition $\gamma \circ p\colon \mathbb{R} \to S$, where $p\colon \mathbb{R} \to S^1$ is the covering map. This map lifts to an arc in $\widetilde{S}$, which we again denote by $\widetilde{\gamma}$. Of course, $\widetilde{\gamma}$ is not unique in either case, but any two such lifts are related by the action of the group of deck transformations on $\widetilde{S}$.

Definition 4.1. If $\widetilde{\gamma}$ is embedded, we say that γ is unobstructed.

By [5, Lemma 2.2] this definition is equivalent to γ having no immersed fishtail. More generally, we define a multicurve $\boldsymbol{\gamma}$, to be a finite union of curves γ_i, and say that $\boldsymbol{\gamma}$ is unobstructed if each γ_i is.

If γ is a loop, we fix a basepoint $* \in S$ which in the image of γ but not a double point, and consider the covering space $S_\gamma \to S$ defined by $\pi_1(S_\gamma, *') = \langle \gamma \rangle \subset \pi_1(S, *)$. Let $\overline{\gamma}$ be the lift of γ to S_γ that passes though the basepoint $*'$. It is easy to see that $S_\gamma \cong S^1 \times \mathbb{R}$, and that γ is unobstructed if and only if $\overline{\gamma}$ is embedded.

Definition 4.2. Loops $\gamma, \gamma' \subset S$ are *commensurable* if there is an immersed loop $\delta \subset S$ and integers n, n' such that γ is freely homotopic to δ^n and γ' is freely homotopic to $\delta^{n'}$.

For curves in the marked torus T, the condition that two curves are commensurable is equivalent to them being globally parallel, in the sense of Definition 3.11. Group theoretically, this condition can be described as follows. As above, we assume our basepoint $*$ lies on γ, so γ determines an element $\gamma \in \pi_1(S)$. The free homotopy class of γ' determines a conjugacy class $\{x\gamma'x^{-1}\} \subset \pi_1(S)$. Then γ and γ' are commensurable if and only if there is a cyclic subgroup $C \subset \pi_1(S)$ containing both γ and $x\gamma'x^{-1}$ for some x.

Lemma 4.3. *Suppose S is noncompact so that $\pi_1(S)$ is free. If γ, γ' are not commensurable, any lift of γ' to S_γ is an arc. If they are commensurable, there is precisely one lift of γ' to S_γ that is a loop.*

Proof. The different lifts of γ' to S_γ correspond to conjugates $x\gamma'x^{-1}$ of γ' in the free group $\pi_1(S)$. The element $x\gamma'x^{-1}$ lifts to a loop if and only if some power of $x\gamma'x^{-1}$ is contained in $\langle\gamma\rangle$, *i.e.* $x(\gamma')^n x^{-1} = \gamma^m$ for some n, m. Consider the subgroup $G = \langle x\gamma'x^{-1}, \gamma\rangle \subset \pi_1(S)$. By hypothesis, $\pi_1(S)$ is free, so G is free as well; G contains a nontrivial central element (namely $x(\gamma')^n x^{-1} = \gamma^m$) so it must be free of rank 1. Hence γ and γ' are commensurable if and only if some lift of γ' is a loop.

Now suppose that γ' and γ are commensurable. After replacing γ' with $x\gamma'x^{-1}$, we may assume that γ' and γ belong to the same cyclic subgroup of $\pi_1(S)$, so there is some $\delta \in \pi_1(S)$ with $\gamma = \delta^r$, $\gamma' = \delta^s$. Saying that γ' has a different lift that is a loop means there is some $w \in \pi_1(S)$ with $w\gamma'w^{-1} \neq \gamma'$ and $w(\gamma')^n w^{-1} = \gamma^m$, or equivalently, $w\delta^{sn}w^{-1} = \delta^{rm}$.

Suppose this is the case, and consider the groups $H = \langle w, \delta\rangle \subset \pi_1(S)$, and F_2, the free group generated by w and δ. H is free, and there is an obvious surjection $\pi\colon F_2 \to H$. By Grushko's theorem, the rank of H is ≤ 2, and if it is equal to 2, π is an isomorphism. Now w and δ satisfy the nontrivial relation $w\delta^{sn}w^{-1} = \delta^{rm}$ in H, so π cannot be an isomorphism. Thus H is free of rank 1, so it is abelian. But in this case w commutes with δ, and thus also with γ'. This contradicts our assumption that $w\gamma'w^{-1} \neq \gamma'$. We conclude that no other lift of γ' is a loop. $\square$

Suppose that γ, γ' are unobstructed curves and that $x, y \in \gamma \cap \gamma'$. Let $\pi_2(x, y)$ be the set of homotopy classes of bigons from x to y.

Lemma 4.4. *If γ and γ' are incommensurable, $\pi_2(x, y)$ is either empty or a single point. If γ and γ' are commensurable, $\pi_2(x, y)$ is either empty or a $\mathbb{Z}$-torsor.*

Proof. The set $\pi_2(x, y)$ is either empty or a torsor over $\pi_2(x, x)$. The latter group can be computed from the long exact sequence

$$\pi_2(S) \to \pi_2(x, x) \to \pi_1(\gamma) \oplus \pi_1(\gamma') \xrightarrow{\alpha} \pi_1(S),$$

where $\alpha(m, n) = \gamma^m(\gamma')^n$. (Note that this is only a map of sets, not a homomorphism.) We have $\pi_2(S) = 0$, and γ, γ' represent nontrivial elements of $\pi_1(S)$, so $\alpha^{-1}(1)$ is either the identity element in $\pi_1(\gamma) \oplus \pi_1(\gamma')$ (if γ and γ' are incommensurable), or isomorphic to $\mathbb{Z}$. $\square$

Now we are ready to discuss the Floer chain complex. Suppose that $\boldsymbol{\gamma} = \cup\gamma_i$ and $\boldsymbol{\gamma}' = \cup\gamma_j'$ are unobstructed transversally intersecting multicurves and that every component of $\boldsymbol{\gamma}$ is a loop. As above, if $x, y \in \boldsymbol{\gamma} \cap \boldsymbol{\gamma}'$, we let $\pi_2(x, y)$ be the set of homotopy classes of bigons from x to y, where a bigon from x to y is a continuous map $\psi\colon D^2 \to S$ such that $\psi(-i) = x, \psi(i) = y$, the restriction of ψ to the right half of S^1 factors through some γ_i, and its restriction to the left half of S^1 factors through some γ_j'. It is immediate from the definition that $\pi_2(x, y) = \emptyset$ unless x and y belong to the same γ_i and γ_j'.

Roughly speaking, we want the Floer chain complex $CF(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$ to be the $\mathbb{F}$ vector space generated by $\boldsymbol{\gamma} \cap \boldsymbol{\gamma}'$, with differential given by $dx = \sum_y n(x, y)y$, where $n(x, y)$ is the count of immersed bigons from x to y, just as in Section 2.5. Now $\pi_2(x, y) = \emptyset$ unless x and y both belong to the same γ_i and the same γ_j', so

$$CF(\boldsymbol{\gamma}, \boldsymbol{\gamma}') \cong \bigoplus_{i,j} CF(\gamma_i, \gamma_j').$$

Hence in what follows we can restrict attention to the case where $\boldsymbol{\gamma} = \gamma$ and $\boldsymbol{\gamma}' = \gamma'$ are individual curves.

The fact that $d^2 = 0$ when γ and γ' are unobstructed follows from a combinatorial count of disks exactly as in [1, Lemma 2.11]; Figure 39 illustrates the idea of the proof.
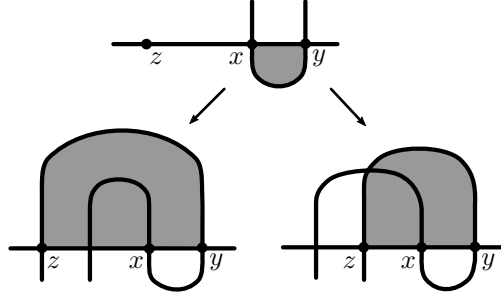
Figure 38. Local pictures illustrating why $d^2 = 0$: Either $d^2(x) = 0$ owing to the fact that 4 bigons must be present (as illustrated on the left), or there must be a cusp in one of the γ_i (as illustrated on the right).

There are a few subtleties to consider. The first is the issue of local systems. Abstractly, a local system on a curve γ assigns a vector space V_x to each point in the domain of γ, together with an isomorphism $\phi_\alpha \colon V_x \to V_y$ for each homotopy class of paths α from x to y, with the property that $\phi_{\alpha \circ \alpha'} = \phi_\alpha \circ \phi_{\alpha'}$. If γ is an arc, two local systems on γ are isomorphic if and only if they have the same dimension. If γ is an oriented loop, a local system on γ is determined up to isomorphism by $k = \dim V_x$ and a matrix $A \in GL_k(\mathbb{F})$ (well defined up to conjugacy) that represents the monodromy $\phi_\alpha \colon V_x \to V_x$ where α is a generator of $\pi_1(S^1, x)$. We represent γ equipped with a local system as above by the triple $\boldsymbol{\gamma} = (\gamma, k, A)$, where γ carries an orientation given by our choice of generator for π_1. Changing the orientation on γ has the effect of replacing A with A^{-1}.

If $\boldsymbol{\gamma}, \boldsymbol{\gamma}'$ are curves as above equipped with local systems (V, ϕ) and (V', ϕ'), then we define

$$CF(\boldsymbol{\gamma}, \boldsymbol{\gamma}') = \bigoplus_{x \in \gamma \cap \gamma'} V_x \otimes V_x'.$$

An immersed bigon $\psi \in \pi_2(x, y)$ induces a map $d_\psi \colon V_x \otimes V_x' \to V_y \otimes V_y'$ given by $d_\psi = \phi_{\psi_1} \otimes \phi_{\psi_2}$ where ψ_1 and ψ_2 are the sides of ψ running along γ and γ' respectively. Both ψ_1 and ψ_2 are oriented to point from x to y. The differential on $CF(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$ is given by $d = \sum_\psi d_\psi$, where the sum runs over all immersed bigons in S. (A more common convention is to use $CF(\boldsymbol{\gamma}, \boldsymbol{\gamma}') = \bigoplus_{x \in \gamma \cap \gamma'} \mathrm{Hom}(V_x, V_x')$. The two are related by replacing $\boldsymbol{\gamma}' = (\gamma', k, A)$ with $(\gamma', k, (A^{-1})^T)$.)

The second issue we must address is the finiteness of this sum. The definition we have given differs from the standard one in symplectic geometry [1], which uses coefficients in a Novikov ring and keeps track of the symplectic area of the disks. Since we are working with coefficients in $\mathbb{F}$ rather than the Novikov ring, we must ensure that there are only finitely many immersed bigons in order for the sum in the definition to make sense. If γ and γ' are incommensurable, this is automatic: there are finitely many $x, y \in \gamma \cap \gamma'$; $\pi_2(x, y)$ contains either 0 or 1 representative, and each homotopy class contains at most one immersed bigon.

If γ and γ' are commensurable, we must analyze the situation more closely. Since the curves are commensurable, $\alpha = \gamma^n$ is homotopic to $\alpha' = \gamma'^m$ for some $n, m \in \mathbb{Z}$. Since they are unobstructed, α and α' lift to homotopic simple closed curves $\widetilde{\alpha}$ and $\widetilde{\alpha}'$ in the covering space S_α.

The following lemma is well-known, but since the proof is short, we give it here.

Lemma 4.5. *Suppose that γ is a loop, that $p \colon \overline{S} \to S$ is a covering map, and that $\overline{\gamma} \colon S^1 \to \overline{S}$ is a lift of γ to $\overline{S}$. Then $CF(\gamma, \gamma') \cong CF(\overline{\gamma}, p^{-1}(\gamma'))$.*

Proof. Since $\overline{\gamma}$ is a lift of γ, p gives a bijection between $\overline{\gamma} \cap p^{-1}(\gamma')$ and $\gamma \cap \gamma'$. Hence the two complexes have the same generators.

Suppose $\psi \colon D^2 \to S$ is an immersed disk from x to y, and let $\overline{x}$ and $\overline{y}$ be the preimage of x and y in $\overline{\gamma}$. Since D^2 is simply connected, ψ lifts to an immersion $\overline{\psi} \colon D^2 \to \overline{S}$ with $\overline{\psi}(-i) = \overline{x}$. The side of ψ running along γ lifts to $\overline{\gamma}$, so $\psi(i) = \overline{y}$, and $\overline{\psi}$ is an immersed disk from $\overline{x}$ to $\overline{y}$. Conversely, if $\overline{\psi}$ is an immersed disk from $\overline{x}$ to $\overline{y}$, $p \circ \psi$ is an immersed disk from x to y. It follows that the two complexes have the same differentials as well. $\square$

Corollary 4.6. *Suppose γ and γ' are commensurable, and α, α' are as above. Then the sum appearing in the boundary operator for $CF(\gamma, \gamma')$ is finite if and only if the sum for $CF(\widetilde{\alpha}, \widetilde{\alpha}')$ is.*

Proof. We apply the lemma to the cover $p_\gamma \colon S_\gamma \to S$. By Lemma 4.3, $p_\gamma^{-1}(\gamma')$ contains finitely many arcs $\beta'_1, \ldots, \beta'_k$ intersecting $\overline{\gamma}$, along with a single loop β'_0 that is commensurable with $\overline{\gamma}$, so

$$CF(\gamma, \gamma') \cong \bigoplus_{i=0}^{k} CF(\overline{\gamma}, \beta'_i).$$

If $\gamma^n = \gamma'^m$ then $\beta_0 = \overline{\gamma}^n$ in $\pi_1(S_\gamma) \cong \mathbb{Z}$. Applying the lemma again, this time to the n-fold-cyclic cover $p \colon S_\alpha \to S_\gamma$ we see that $CF(\overline{\gamma}, \beta_0) \cong CF(p^{-1}(\overline{\gamma}), \overline{\beta_0}) = CF(\widetilde{\alpha}, \widetilde{\alpha}')$. For $i > 0$, β_i is an arc, so the differential in the complex $CF(\overline{\gamma}, \beta_i)$ is a finite sum. Hence the differential on $CF(\gamma, \gamma')$ is well-defined if and only if the differential on $CF(\widetilde{\alpha}, \widetilde{\alpha}')$ is. $\square$

We define $\mathcal{P}_{\alpha, \alpha'}$ to be the *periodic domain* bounded by α and α', that is, the domain of the compactly supported 2-chain in S_α which is bounded by $\widetilde{\alpha} - \widetilde{\alpha}'$. If $z \in S_\alpha - \widetilde{\alpha} - \widetilde{\alpha}'$, we let $n_z(\mathcal{P}_{\alpha, \alpha'})$ be the multiplicity of $\mathcal{P}_{\alpha, \alpha'}$ at z.

Definition 4.7. We say that γ and γ' are in *admissible position* if $\mathcal{P}_{\alpha, \alpha'}$ has both positive and negative multiplicities.

Since $\widetilde{\alpha}$ and $\widetilde{\alpha}'$ are transverse embedded curves in $S^1 \times \mathbb{R}$, γ and γ' are in admissible position if and only if α and α' intersect.

Suppose that γ and γ' are in admissible position, that $x, y \in \alpha \cap \alpha'$, and that $\psi_1, \psi_2 \in \pi_2(x, y)$. If $\mathcal{D}(\psi_i)$ is the domain associated with ψ, then $\mathcal{D}(\psi_1) = \mathcal{D}(\psi_2) + k\mathcal{P}_{\alpha, \alpha'}$ for some $k \in \mathbb{Z}$. Since $\mathcal{P}_{\alpha, \alpha'}$ has both positive and negative multiplicities, there are only finitely many values of k for which all the multiplicities of $\mathcal{D}(\psi_1)$ are positive, and hence only finitely many terms in the sum for the differential on $CF(\alpha, \alpha')$. Hence if γ and γ' are in admissible position, $CF(\gamma, \gamma')$ is well-defined.

More generally, if $\boldsymbol{\gamma} = \cup \boldsymbol{\gamma}_i$ and $\boldsymbol{\gamma}' = \cup \boldsymbol{\gamma}'_j$ are unobstructed immersed curves equipped with local systems, we say $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$ are in admissible position if γ_i and γ'_j are in admissible position for each i and j. If this is the case, the Floer chain complex $CF(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$ is well defined. We will now explain how to compute its homology $HF(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$. We start with an elementary lemma, whose proof is left to the reader.

Lemma 4.8. *If $\boldsymbol{\gamma}_1 = (\gamma, k_1, A_1)$ and $\boldsymbol{\gamma}_2 = (\gamma, k_2, A_2)$, then $CF(\boldsymbol{\gamma}_1 \cup \boldsymbol{\gamma}_2, \boldsymbol{\gamma}') = CF(\boldsymbol{\gamma}_3, \boldsymbol{\gamma}')$, where $\boldsymbol{\gamma}_3 = (\gamma, k_1 + k_2, A_1 \oplus A_2)$. Similarly, if $\boldsymbol{\gamma}_1 = (\gamma_1, k, A)$, where γ_1 is the n-fold cover of γ, then $CF(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}') = CF(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$, where $\boldsymbol{\gamma} = (\gamma, nk, A')$ and A' is a block matrix of the form*

(2)
$$\begin{bmatrix} 0 & 0 & 0 & A \\ \hline I_k & 0 & 0 & 0 \\ \hline 0 & I_k & 0 & 0 \\ \hline 0 & 0 & I_k & 0 \end{bmatrix}$$

Next, we consider the effect of varying a loop γ by a homotopy. Fix some γ'; we say that the homotopy γ_t is *good* if for each $t \in \{0,1\}$, γ_t is unobstructed and the pair (γ_t, γ') is admissible. If (V_0, ϕ_0) is a local system on γ_0, the pullback to $S^1 \times [0,1]$ induces a local system (V_1, ϕ_1) on γ_1.

Theorem 4.9. *Suppose that $\boldsymbol{\gamma}_0, \boldsymbol{\gamma}'$ are unobstructed curves equipped with local systems. If γ_t is a good homotopy from γ_0 to γ_1, then $HF(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}') \cong HF(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}')$.*

Proof. This is a standard result in the symplectic context, where we work with area preserving isotopies and coefficients in the Novikov ring. For a proof in this setting, see [1, Section 4] or [5, Lemma 2.12]. We sketch the necessary changes to make the proof work in the context of admissible diagrams. As in the proof of Proposition 4.1 in [1], we first pass to the cover S_{γ_0}, and thus reduce to the case where γ_0 and γ_1 are embedded. After splitting the preimage of γ' into separate components and passing to a further cover, as in the proof of Corollary 4.6, we can further assume that γ' is either (a) incommensurable with γ_0 or (b) embedded and isotopic to γ_0. Case (b) is actually a special case of the invariance of sutured Floer homology [30], since the sutured Heegaard diagrams $(S^1 \times [0,1], \gamma_0, \gamma')$ and $(S^1 \times [0,1], \gamma_1, \gamma')$ are admissible sutured Heegaard diagrams representing the same sutured manifold. Henceforth we assume we are in case (a).

By breaking the isotopy γ_t into a sequence of smaller isotopies, as in [1, Lemma 4.2], we may assume that γ_0 and γ_1 are parallel curves intersecting in two points θ_1 and θ_2. Then a (combinatorially defined) count of immersed triangles corresponding to composition with θ_1 and θ_2 gives chain maps $f \colon CF(\gamma_0, \gamma') \to CF(\gamma_1, \gamma')$ and $g \colon CF(\gamma_1, \gamma') \to CF(\gamma_2, \gamma')$, where γ_2 is a small translate of γ_0 intersecting it in two points. Finally, the associativity relation (obtained by counting immersed quadrilaterals) ensures that $f \circ g$ is homotopic to the natural map identifying $CF(\gamma_0, \gamma')$ with $CF(\gamma_2, \gamma')$.

This argument carries over largely unchanged to the admissible setting, but we must check that the counts of immersed triangles and quadrilaterals used to define the chain maps and homotopies are finite. We discuss the case of quadrilaterals; the argument for triangles is very similar. In order to check that there are only finitely many homotopy classes that contribute to the sum, we must show that for a given N, there are only finitely many periodic domains bounded by $\gamma_0, \gamma_1, \gamma_2$ and γ' whose multiplicities are all between $-N$ and N. To do this, it suffices to find a basis $\mathcal{P}_i$ for the space of periodic domains and points $z_i^{\pm}$ such that $n_{z_i^+}(\mathcal{P}_i) > 0$, $n_{z_i^-}(\mathcal{P}_i) < 0$, and $n_{z_i^{\pm}}(\mathcal{P}_j) = 0$ for $i \neq j$. In this case we say the points *separate* the periodic domains. Since we are in case (a) γ_0 and γ' are incommensurable, and all periodic domains are bounded by γ_0, γ_1, and γ_2. By choosing γ_2 appropriately, we can assume the configuration is as shown in Figure 40. The periodic domains $\mathcal{P}_1$ and $\mathcal{P}_2$ form a basis and are separated by the points shown in the figure. $\square$
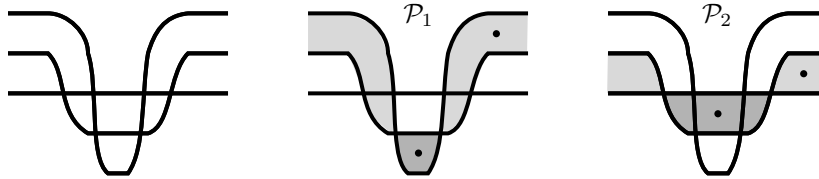
Figure 39. The curves $\gamma_0, \gamma_1, \gamma_2$ in the cylinder S_{γ_0}, together with a basis of periodic domains.

Recall that if γ, γ' are essential immersed curves in S, their *intersection number* is the minimal geometric intersection number between transverse curves $\widehat{\gamma}, \widehat{\gamma}'$ that are homotopic to γ, γ'. By a theorem of Freedman, Hass, and Scott [13], this minimal number is realized when $\widehat{\gamma}, \widehat{\gamma}'$ are shortest length geodesics in the free homotopy classes of γ, γ'. Similarly, if γ is primitive, the self-intersection number $i(\gamma)$ is the minimal number of self-intersections among immersed curves homotopic to γ. It is realized by the number of self-intersections of the shortest geodesic representing γ.

Corollary 4.10. *Suppose $\boldsymbol{\gamma} = (\gamma, k, A)$ and $\boldsymbol{\gamma}' = (\gamma', k', A')$ are primitive unobstructed immersed curves equipped with local systems. If γ and γ' are incommensurable, then*

$$\dim HF(\boldsymbol{\gamma}, \boldsymbol{\gamma}') = kk' \cdot i(\gamma, \gamma').$$

If $\gamma' = \gamma^{-1}$, then

$$\dim HF(\boldsymbol{\gamma}, \boldsymbol{\gamma}') = 2kk' \cdot i(\gamma) + 2\dim(\ker \Phi_{A,A'}),$$

where $\Phi_{A,A'}$ is the endomorphism of $\mathbb{F}^k \otimes \mathbb{F}^{k'}$ given by $\Phi_{A,A'} = I \otimes I + A \otimes A'$. As a consequence, $HF(\boldsymbol{\gamma}, \boldsymbol{\gamma})$ is always a nontrivial.

If γ, γ' are commensurable but not primitive, we can use Lemma 4.8 to reduce to the case where we have two parallel primitive curves. Hence the corollary enables us to compute $HF(\gamma, \gamma')$ for any two decorated curves γ, γ'.

Proof. Let $\widehat{\gamma}, \widehat{\gamma}'$ be shortest geodesics homotopic to γ, γ'. Since γ and γ' are primitive, we can assume they are either incommensurable or $\widehat{\gamma}' = \widehat{\gamma}^{\pm 1}$. In the first case, $CF(\boldsymbol{\gamma}, \boldsymbol{\gamma}') \cong CF(\widehat{\gamma}, \widehat{\gamma}')$, and $CF(\widehat{\gamma}, \widehat{\gamma}')$ has $kk' \cdot i(\gamma, \gamma')$ generators. To understand the differential, we consider the cover S_γ. By section 3 of [13], each component of $p_\gamma^{-1}(\widehat{\gamma}')$ intersects the lift of $\widehat{\gamma}$ in either 0 or 1 point. Hence $\pi_2(x, y) = 0$ for any two distinct generators of $CF(\widehat{\gamma}, \widehat{\gamma}')$. This proves the first statement.

If γ' is homotopic to γ^{-1}, $HF(\boldsymbol{\gamma}, \boldsymbol{\gamma}') \cong HF(\widehat{\gamma}, \gamma'')$, where γ'' is a small translate of $\widehat{\gamma}$ intersecting it in two points. In this case $p_\gamma^{-1}(\gamma'')$ consists of some arcs and a single loop that is a small translate of $\widehat{\widehat{\gamma}}$ and intersects it in two points. Since γ'' is a small translate of $\widehat{\gamma}$, as before, each of the arcs intersects $\widehat{\widehat{\gamma}}$ in a single point. There are two such intersections for each self-intersection of $\widehat{\gamma}$, so in total these contribute $2kk' \cdot i(\gamma)$ generators, and no differential. The final part of the chain complex is the summand $CF(\widehat{\widehat{\gamma}}, \widehat{\widehat{\gamma}})$ for which there are two intersection points, each contributing $\mathrm{Hom}(V', V) \cong \mathbb{F}^k \otimes \mathbb{F}^{k'}$ to the set of generators, and two embedded disks. The differential in this complex is given by $\Phi_{A,A'}$, so the dimension of its homology is $2\dim(\ker \Phi_{A,A'})$.

For the final claim, reverse the orientation on γ to get

$$HF((\gamma, k, A), (\gamma, k, A)) = HF((\gamma, k, A), (\gamma^{-1}, k, A'))$$

where $A' = (A^{-1})^T$. The operator $\Phi_{A,A'}$ can be viewed as a map from $\mathrm{End}(\mathbb{F}^k)$ to itself. Thinking of $\mathrm{End}(\mathbb{F}^k)$ as the space of $k \times k$ matrices with entries in $\mathbb{F}$, we have $\Phi_{A,A'}(X) = AXA^{-1} + X$. The $k \times k$ identity matrix is always in the kernel of Φ, so $HF((\gamma, k, A), (\gamma, k, A))$ is always nontrivial. $\square$

4.2. **From Type D structures to curves.** We now define a set $\mathcal{C}$, which is the target of the invariant $\widehat{HF}(M)$. Consider collections of finitely many oriented immersed unobstructed loops decorated with local systems in the marked torus; recall that by marked torus we mean the punctured torus T with a choice of parametrizing curves α and β. Each decorated curve is a triple (γ, k, A), where γ is an oriented unobstructed immersed curve, k is a positive integer, and A is a similarity class of $k \times k$ matrices over $\mathbb{F}$. As discussed earlier, the pair (k, A) determines a k-dimensional local system on γ.

We impose an equivalence relation on collections of such curves, generated by the following relations, which can be applied to any subset of a collection:

- If γ is homotopic to γ', (γ, k, A) is equivalent to (γ', k, A).
- (γ, k, A) is equivalent to $(-\gamma, k, A^{-1})$, where $-\gamma$ denotes γ with the orientation reversed.
- The pair $\{(\gamma, k_1, A_1), (\gamma, k_2, A_2)\}$ is equivalent to the single curve $(\gamma, k_1 + k_2, A_1 \oplus A_2)$.
- If γ is homotopic to $(\gamma')^m$ then (γ, k, A) is equivalent to (γ', mk, A'), where A' is a block matrix of the form shown in equation (2).

We define $\mathcal{C}$ to be the set of equivalence classes. Each equivalence class contains a representative which is minimal in the sense that the underlying curves are all primitive and no two of them are

homotopic. This representative is unique up to orientation reversal, conjugation of the monodromy, and homotopy of the underlying curves.

Given $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathcal{C}$, we define their pairing to be $\langle \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \rangle = HF(\boldsymbol{\gamma}_1, r(\boldsymbol{\gamma}_2))$, where r denotes the orientation reversing diffeomorphism of the marked torus that exchanges α and $-\beta$. Lemma 4.8 and Theorem 4.9 imply that the pairing is well-defined.

Let $\mathcal{D}$ denote the set of extendable type D structures over $\mathcal{A}$, up to homotopy equivalence. The set $\mathcal{D}$ is equipped with a pairing given by the box tensor product: if $N_1, N_2 \in \mathcal{D}$, their pairing is $\langle N_1, N_2 \rangle = N_1^A \boxtimes N_2$, where N_1^A is the $\mathcal{A}_\infty$ module corresponding to N^1 under the duality described in Section 2.3 (see Remark 2.3).

The main result of this section is:

Theorem 4.11. *There is a bijection $f \colon \mathcal{D} \to \mathcal{C}$, with inverse $g \colon \mathcal{C} \to \mathcal{D}$, satisfying $\langle N_1, N_2 \rangle \simeq \langle f(N_1), f(N_2) \rangle$ and, equivalently, $\langle \gamma_1, \gamma_2 \rangle \simeq \langle g(\gamma_1), g(\gamma_2) \rangle$.*

We first define the inverse $g \colon \mathcal{C} \to \mathcal{D}$. Given a decorated immersed curve $\boldsymbol{\gamma} = (\gamma, k, A)$ in the marked torus, we construct a train track by replacing γ with k parallel copies of γ, writing A as a product of elementary matrices $P_1 \cdots P_\ell$, and adding a crossover arrow between copies of γ for each elementary matrix, where an elementary matrix of type A_{ij} corresponds to an arrow connecting the ith copy of γ to the jth copy of γ and the arrows corresponding to $P_\ell, P_{\ell-1}, \ldots, P_1$ appear in order according to the orientation of γ. We assume that γ is homotoped so that it has minimal intersection with $\alpha \cup \beta$, and we take care that no crossover arrows intersect α or β. The result is a valid reduced train track; this corresponds to an extended type D structure $\widetilde{N}$, which has an underlying (extendable) type D structure N. We define $g(\boldsymbol{\gamma}) = N$. In doing so, we made some choices: namely a matrix representing the similarity class A and an elementary decomposition of that matrix. Changing either choice corresponds to sliding crossover arrows around γ by moves which preserve the homotopy equivalence class of the corresponding extended type D module; thus g is well-defined. Finally, if $\boldsymbol{\gamma} = \bigcup \gamma_i$ is a union of decorated curves, we define $g(\boldsymbol{\gamma}) = \bigoplus g(\gamma_i)$.

Next, we consider the effect of g on the pairing. Recall that if $\boldsymbol{\vartheta}$ is a curve-like train track (as in Definition 3.16), there is an associated local system (V, ϕ) on the underlying curve γ. If x is a point on γ, V_x is the $\mathbb{F}$ vector space spanned by vectors e_p for $p \in S_x := \boldsymbol{\vartheta} \cap \ell_x$, where ℓ_x is a small arc transverse to γ and passing through x. If α is an arc from x to y along γ, then $\phi_\alpha \colon V_x \to V_y$ is given by $\phi_\alpha(e_p) = \sum n_{pq} e_q$, where n_{pq} counts the number of paths on from p to q on θ_i which project to the path α on γ. Applying this procedure to the train track used to define $g(\boldsymbol{\gamma})$ returns the decorated curve $\boldsymbol{\gamma}$.

Now suppose that $\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2$ are two valid reduced train tracks, and consider the chain complex $\mathcal{C}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2)$, defined in Section 3.2. If $A(\boldsymbol{\vartheta}_1)$ and $D(\boldsymbol{\vartheta}_2)$ are in admissible position, $\mathcal{C}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2)$ is their intersection Floer chain complex. Note that $A(\boldsymbol{\vartheta}_1)$ is isotopic to $\boldsymbol{\vartheta}_1$, while $D(\boldsymbol{\vartheta}_2) = r(A(\boldsymbol{\vartheta}_2))$ is isotopic to $r(\boldsymbol{\vartheta}_2)$.

Lemma 4.12. *Suppose that $\boldsymbol{\vartheta}$ and $\boldsymbol{\vartheta}'$ are curve-like train tracks with associated decorated curves $A(\boldsymbol{\vartheta}) = \boldsymbol{\gamma} = (\gamma, V, \phi)$ and $D(\boldsymbol{\vartheta}_2) = \boldsymbol{\gamma}' = (\gamma', V', \phi')$, and that $\boldsymbol{\vartheta}$ and $\boldsymbol{\vartheta}'$ are in admissible position. Then $\mathcal{C}(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}') \cong CF(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$.*

Proof. Generators of $\mathcal{C}(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}')$ correspond to triples (x, p, q), where $x \in \gamma \cap \gamma'$, $p \in S_x$ and $q \in S'_x$, while generators of $CF(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$ correspond to pairs (x, A), where A is a generator of $V_x \otimes V'_x$. They are in bijection via the map that sends (x, p, q) to $(x, e_p \otimes e_q)$.

Similarly, immersed disks with boundaries on $\boldsymbol{\vartheta}$ and $\boldsymbol{\vartheta}'$ are in bijection with triples (ψ, α, α'), where ψ is an immersed disk with boundary on γ_1 and γ_2, α is a path on $\boldsymbol{\vartheta}$ which projects to the part of the boundary of ψ which lies on γ (and similarly for α'), and α and α' have the same endpoints.

Under the bijection above, the contribution to the differential on $\mathcal{C}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2)$ coming from ψ matches the contribution to the differential on $CF(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ coming from ψ. $\square$

Proposition 4.13. *If* $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathcal{C}$ *then* $\langle \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \rangle \simeq \langle g(\boldsymbol{\gamma}_1), g(\boldsymbol{\gamma}_2) \rangle$.

Proof. Let $N_i = g(\boldsymbol{\gamma}_i)$, let $\boldsymbol{\vartheta}_i$ be the curve-like train track associated with $\boldsymbol{\gamma}_i$, and consider the complex $\mathcal{C}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2)$. In order for $\mathcal{C}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2)$ to be defined, we need the pair to be admissible—there can be no immersed annuli carried by $A(\boldsymbol{\vartheta}_1)$ and $D(\boldsymbol{\vartheta}_2)$. To ensure this, we will modify $\boldsymbol{\vartheta}_2$ so that it has no component for which the underlying curve consists only of the following four types of segments:
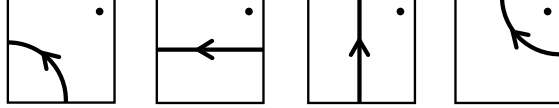
This ensures that the pair is admissible, since any immersed annulus, when cut by α and β, consists only of the connecting pieces shown in Figure 12. If any component of $\boldsymbol{\vartheta}_2$ does contain only the four segment types above, we choose one of the latter three types of segments and modify $\boldsymbol{\vartheta}_2$ by applying a finger-move isotopy to the given segment across either α or β, as in Figure 11. If $\boldsymbol{\vartheta}_2$ has parallel copies of the relevant segment we apply the isotopy to each copy, keeping them parallel, and ensure that the finger-move avoids any crossover arrows. Let $\boldsymbol{\vartheta}_2'$ denote the result of applying these finger-moves, if necessary, to $\boldsymbol{\vartheta}_2$. Note that $\boldsymbol{\vartheta}_2'$ is an almost reduced train track whose corresponding type D structure N_2' is homotopy equivalent to N_2, where the homotopy equivalence replaces certain ρ_{12}, ρ_{23}, or ρ_{123} arrows with zig-zags as in Figure 11; N_2' is both bounded and almost reduced.

By Theorem 2.2, $\mathcal{C}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2')$ is isomorphic to $N_1^A \boxtimes N_2'$, which is quasi-isomorphic to $N_1^A \boxtimes N_2 = \langle g(\boldsymbol{\gamma}_1), g(\boldsymbol{\gamma}_2) \rangle$. On the other hand, by Lemma 4.12, $\mathcal{C}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2') \cong CF(\boldsymbol{\gamma}_1, r(\boldsymbol{\gamma}_2'))$, where $\boldsymbol{\gamma}_2'$ is isotopic to $\boldsymbol{\gamma}_2$. Passing to homology gives the statement of the proposition. $\square$

To prove Theorem 4.11, we must show that g is injective and surjective. For surjectivity, we appeal to Theorem 1.5. Given an element of $\mathcal{D}$, we choose a reduced representative N and an extension $\widetilde{N}$ of N. By Theorem 1.5, N is homotopy equivalent to a reduced type D structure N' whose associated train track is a collection of curve-like train tracks. As we explained in the proof of Theorem 1.5, to such a train track we can naturally associate an immersed decorated multicurve $\boldsymbol{\gamma}$. To see that $\boldsymbol{\gamma}$ is an element of $\mathcal{C}$, we must check that each curve $\gamma \subset \boldsymbol{\gamma}$ is unobstructed, or equivalently, that $\widetilde{\gamma}$ is embedded in the universal cover. This follows from the fact that each individual segment of γ is embedded together with the fact that N' is reduced, so the path traced out by γ in the universal cover (thought of as a thickening of the regular 4-valent tree) never retraces itself. Then $g(\boldsymbol{\gamma}) = N'$ is homotopy equivalent to our original N, so g is surjective.

4.3. **Injectivity of g.** To complete the proof of Theorem 4.11, we show that g is injective. Suppose that $N_1 = g(\boldsymbol{\gamma}_1)$ and $N_2 = g(\boldsymbol{\gamma}_2)$ are homotopy equivalent to each other. Since N_1 and N_2 are homotopy equivalent, they have the same pairing behavior; that is,

$$H_*(N_1^A \boxtimes N) \cong H_*(N_2^A \boxtimes N)$$

for any reduced type D structure N. For $i = 1, 2$, let $\boldsymbol{\vartheta}_i$ be a curve-like train track in T that represents $\boldsymbol{\gamma}_i$. It follows from Theorem 2.2 that $H_*(\mathcal{C}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta})) \cong H_*(\mathcal{C}(\boldsymbol{\vartheta}_2, \boldsymbol{\vartheta}))$ for any reduced weakly valid train track $\boldsymbol{\vartheta}$. We will show that $\boldsymbol{\gamma}_i$ is determined by the groups $H_*(\mathcal{C}(\boldsymbol{\vartheta}_i, \boldsymbol{\vartheta}))$ as $\boldsymbol{\vartheta}$ varies; thus $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2$.

Since $\boldsymbol{\vartheta}_i$ is a curve-like train track, a component of $\boldsymbol{\vartheta}_i$ consists of the following data: an immersed curve γ in the punctured torus that intersects $\alpha \cup \beta$ minimally, an integer determining how many parallel copies of γ there are, and a collection of crossover arrows between parallel copies of γ. The

curve γ may be represented as a reduced cyclic word in $\{\alpha^{\pm 1}, \beta^{\pm 1}\}$ by traversing γ and recording the intersections with the parametrizing curves α and β, with negative crossings recorded as inverse letters. Equivalently, we realize γ as an element of the fundamental group of the marked torus, which is the free group on two generators. Here we take the generators α and β of the fundamental group to be the simple closed curves dual to the parametrizing curves α and β. Note that viewing γ as an element of the fundamental group requires choosing a basepoint, but choosing a different basepoint only changes the resulting word by cyclic permutation.

The underlying curves of $\boldsymbol{\vartheta}_i$ are determined by the results of pairing $\boldsymbol{\vartheta}_i$ with certain immersed line segments. To see this, we first define the immersed segments in question. Given a reduced word I in $\{\alpha^{\pm 1}, \beta^{\pm 1}\}$ we construct a train track $\boldsymbol{\vartheta}_I$ from a sequence of horizontal and vertical segments, with each segment having endpoints near the center of the marked torus and crossing either α or β exactly once. Each segment corresponds to a letter in the word I depending on which parametrizing curve it crosses, where negative crossings are interpreted as an inverse letter. In particular, each letter α in I gives rise to horizontal segment oriented rightward, and each letter β in I gives rise to a vertical segment oriented downward. We construct $\boldsymbol{\vartheta}_I$ by connecting these segments end to end, smoothing the corners, in the order determined by the word I; this results in an oriented immersed curve segment with the property that, when traversed, the intersections with α and β read off the word I. Some examples are shown in Figure 41 and Figure 42. Note that traversing the segment with the opposite orientation gives the inverse word I^{-1} in the free group, so $\boldsymbol{\vartheta}_I = -\boldsymbol{\vartheta}_{I^{-1}}$. We will generally not distinguish $\boldsymbol{\vartheta}_I$ from $\boldsymbol{\vartheta}_{I^{-1}}$, since the orientation is not relevant for pairing. Following Section 2, $\boldsymbol{\vartheta}_I$ represents a type D structure with one generator for each segment and arrows determined by Figure 8. Note that this type D structure satisfies $\partial^2 = 0$ but it is not extendable. Equivalently (by Proposition 3.6), the train track is only weakly valid.

For any given word I in the free group generated by α and β we can define a new word $\bar{I}$, which may be viewed as dual to I. Rather than define this new word algebraically, we define a procedure for constructing the immersed line segment $\boldsymbol{\vartheta}_{\bar{I}}$ from $\boldsymbol{\vartheta}_I$. More precisely, we will first construct $D(\boldsymbol{\vartheta}_{\bar{I}})$, and then determine $\boldsymbol{\vartheta}_{\bar{I}}$ from it. Assuming that the length of I is greater than 1, $D(\boldsymbol{\vartheta}_{\bar{I}})$ is chosen so that (1) the first segment of $D(\boldsymbol{\vartheta}_{\bar{I}})$ intersects the first segment of $A(\boldsymbol{\vartheta}_I)$, (2) the last segment of $D(\boldsymbol{\vartheta}_{\bar{I}})$ intersects the last segment of $A(\boldsymbol{\vartheta}_I)$, and (3) the paths in $D(\boldsymbol{\vartheta}_{\bar{I}})$ and $A(\boldsymbol{\vartheta}_I)$ connecting these first and last intersections are homotopic rel endpoints; see Figure 41. These rules uniquely determine the sequence of horizontal and vertical segments appearing in $D(\boldsymbol{\vartheta}_{\bar{I}})$: for each step after the first, we choose the segment so that it exits the square (the fundamental domain for T) by the same side as $A(\boldsymbol{\vartheta}_I)$. In the special case where the length of I is 1, we introduce the convention that $\bar{I} = I$, so $D(\boldsymbol{\vartheta}_{\bar{I}})$ intersects $A(\boldsymbol{\vartheta}_I)$ in a single point. Note that reflecting across the antidiagonal in the torus takes $A(\boldsymbol{\vartheta}_I)$ to $D(\boldsymbol{\vartheta}_I)$ and $D(\boldsymbol{\vartheta}_{\bar{I}})$ to $A(\boldsymbol{\vartheta}_{\bar{I}})$ so that applying the dual operation to $\bar{I}$ recovers I.

These immersed curve segments will be used as test train tracks against which we will pair the curve-like train tracks $\boldsymbol{\vartheta}_i$ we hope to identify. It turns out that for any word I, the result of pairing with $\boldsymbol{\vartheta}_{\bar{I}}$ gives us information about how many times the word I appears in the cyclic word defining a curve-like train track. More precisely, for a curve-like train track $\boldsymbol{\vartheta}_i$ and a word I, let $n_I(\boldsymbol{\vartheta}_i)$ be the number of times either I or I^{-1} appears in the cyclic words representing the underlying curves of $\boldsymbol{\vartheta}_i$, where each appearance is counted with multiplicity associated with the relevant cyclic word (i.e. the dimension of the local system associated with the corresponding curve component of $\boldsymbol{\vartheta}_i$).

Lemma 4.14. *For a curve-like train track $\boldsymbol{\vartheta}$, $n_I(\boldsymbol{\vartheta})$ is determined by the dimension of the homology of $\mathcal{C}(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_{\bar{I}})$ together with the values of $n_J(\boldsymbol{\vartheta})$ for all words J for which $\bar{J}$ is a subword of $\bar{I}$.*

Proof. The homology of the complex $\mathcal{C}(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_{\bar{I}})$ does not depend on the local system on $\boldsymbol{\vartheta}$ (that is, on the crossover arrows) since $\boldsymbol{\vartheta}_{\bar{I}}$ is not a closed curve. Without loss of generality then, assume that all local systems are trivial and $\boldsymbol{\vartheta}$ has no crossover arrows. Each component of $\boldsymbol{\vartheta}$ is represented by a cyclic word in α and β with some multiplicity. Note that in $A(\boldsymbol{\vartheta})$ there is a horizontal segment for each instance of α or α^{-1} in the words representing $\boldsymbol{\vartheta}$ (counted with multiplicity), and there is
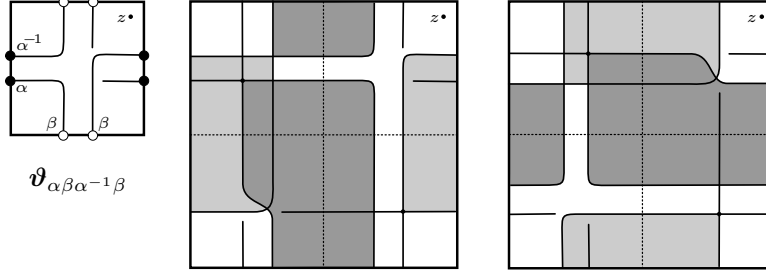
Figure 40. Starting from the word $I = \alpha\beta\alpha^{-1}\beta$ we construct $\boldsymbol{\vartheta}_I$ (left). Next (center) we construct $A(\boldsymbol{\vartheta}_I)$ by applying the operations of Section 2.5, and then find $D(\boldsymbol{\vartheta}_{\bar{I}})$ via the construction in the text. The shaded bigons illustrate the homotopy rel endpoints between the two curves. Note that the start and end points of $D(\boldsymbol{\vartheta}_{\bar{I}})$ (as well as any corners) must lie in the lower left quadrant of the square. Finally, we reflect across the antidiagonal (right) to obtain $A(\boldsymbol{\vartheta}_{\bar{I}})$, and hence $\boldsymbol{\vartheta}_{\bar{I}}$. This determines $\bar{I} = \alpha\alpha\beta^{-1}$. Since $A(\boldsymbol{\vartheta}_{\bar{I}})$ and $D(\boldsymbol{\vartheta}_I)$ satisfy the same relationship as $A(\boldsymbol{\vartheta}_I)$ and $D(\boldsymbol{\vartheta}_{\bar{I}})$, $\bar{\bar{I}} = I$.

a vertical segment for each instance of β or β^{-1}. Moreover, the immersed arc obtained by following $A(\boldsymbol{\vartheta})$ from one horizontal or vertical segment to another can be interpreted as $A(\boldsymbol{\vartheta}_I)$ for a subword I of the cyclic word representing the relevant component of $\boldsymbol{\vartheta}$, and the arc $A(\boldsymbol{\vartheta}_I)$ (ignoring orientation) appears for each instance of I or I^{-1} in the cyclic words representing $\boldsymbol{\vartheta}$.

We first observe that the dimension of $H_*(\mathcal{C}(\boldsymbol{\vartheta},\boldsymbol{\vartheta}_\alpha))$ is precisely $n_\alpha(\boldsymbol{\vartheta})$. This follows from the fact that $D(\boldsymbol{\vartheta}_\alpha)$ consists of a single vertical segment and no bigons are formed in this case, so intersecting $D(\boldsymbol{\vartheta}_\alpha)$ with $A(\boldsymbol{\vartheta})$ simply counts the number of horizontal segments in $A(\boldsymbol{\vartheta})$. Similarly, $n_\beta(\boldsymbol{\vartheta})$ is determined by $H_*(\mathcal{C}(\boldsymbol{\vartheta},\boldsymbol{\vartheta}_\alpha))$. We will demonstrate the proof for longer words I in an example before commenting on the general case. Let $I = \alpha\beta^{-1}\alpha$, so $A(\boldsymbol{\vartheta}_I)$ and $D(\boldsymbol{\vartheta}_{\bar{I}})$ are as shown in Figure 42; note that $\bar{I} = \alpha^{-1}\beta\alpha^{-1}$. To compute the homology of $\mathcal{C}(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_{\bar{I}})$ we first find the generators. These count α-α or β-β intersections, so there are $n_\alpha(\boldsymbol{\vartheta})$ generators for each α or α^{-1} in I and $n_\beta(\boldsymbol{\vartheta})$ for each β or β^{-1} in I. In this example, the number of generators is given by $2n_\alpha(\boldsymbol{\vartheta}) + n_\beta(\boldsymbol{\vartheta})$. We next consider the differential. The length two subwords of $\bar{I}$ are $\bar{J}_1 = \alpha^{-1}\beta$ and $\bar{J}_2 = \beta\alpha^{-1}$, with dual words $J_1 = \alpha\beta^{-1}$ and $J_2 = \beta^{-1}\alpha$ (note that in this example J_1 and J_2 are subwords of I, but this need not be the case). Letting J be either J_1 or J_2, for each instance of J or J^{-1} in the cyclic words representing $\boldsymbol{\vartheta}$, there is a bigon between the portion of $A(\boldsymbol{\vartheta})$ corresponding to J and the portion of $D(\boldsymbol{\vartheta}_{\bar{I}})$ corresponding to $\bar{J}$. These bigons take the form of the lightly shaded bigon in Figure 42 for $J = J_1$ or the darkly shaded bigon in the figure for $J = J_2$. It is not difficult to see that all bigons contributing to the differential occur in this way. If all such bigons were disjoint then we would simply cancel two generators for each bigon. However, if two bigons share an endpoint then this computation is incorrect, as it would over-cancel. We next observe that if $A(\boldsymbol{\vartheta})$ contains a segment that looks like $A(\boldsymbol{\vartheta}_{\alpha\beta^{-1}\alpha})$ then intersecting with $D(\boldsymbol{\vartheta}_{\bar{I}})$ produces a pair of bigons sharing an endpoint as in Figure 42, and any pair of bigons formed with $D(\boldsymbol{\vartheta}_{\bar{I}})$ and sharing an endpoint must have this form. Thus we can correct for the over-cancellation by adding $2n_{\alpha\beta^{-1}\alpha}(\boldsymbol{\vartheta})$, and we have

$$\dim H_*(\mathcal{C}(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_{\bar{I}})) = 2n_\alpha(\boldsymbol{\vartheta}) + n_\beta(\boldsymbol{\vartheta}) - 2n_{\alpha\beta^{-1}}(\boldsymbol{\vartheta}) - 2n_{\beta^{-1}\alpha}(\boldsymbol{\vartheta}) + 2n_{\alpha\beta^{-1}\alpha}(\boldsymbol{\vartheta})$$

The last term on the right is $n_I(\boldsymbol{\vartheta})$ and all other terms on the right are $n_J(\boldsymbol{\vartheta})$ where $\bar{J}$ is a subword of $\bar{I}$, as desired.

We now consider the case of an arbitrary word I. A key fact implicit in the example above is that any intersection point x between $A(\boldsymbol{\vartheta})$ and $D(\boldsymbol{\vartheta}_{\bar{I}})$ is an end of at most two bigons, and if there are two then they both start at x or they both end at x. This fact holds in the general case, as we will show below. Assuming this, note that the Floer complex $C(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_{\bar{I}})$ is a direct sum of *zig-zag chains*, i.e. complexes that can be represented by a chain of dots along a line connected by arrows on that line pointing in alternating directions. The dimension of $H_*(C(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_{\bar{I}}))$ is a function of the number of generators and the number of zig-zag chains of length k for each $k > 0$; note that when
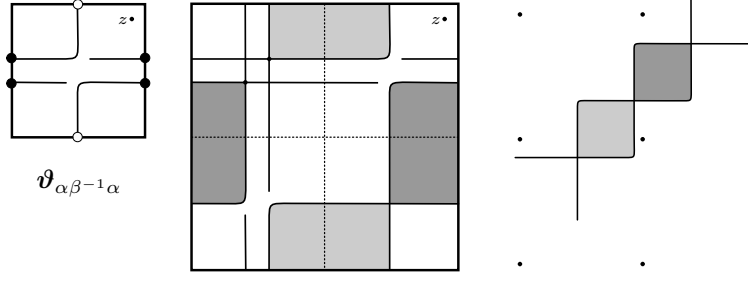
Figure 41. As simple example illustrating the key step in the proof of Lemma 4.14, the pairing of $A(\boldsymbol{\vartheta}_I)$ with $D(\boldsymbol{\vartheta}_{\bar{I}})$ for $I = \alpha\beta^{-1}\alpha$, shown both in the punctured torus and a covering space.

we count zig-zag chains of length k we will include subchains of longer chains, so for instance a chain of length three also contributes two to the count of length two chains and three to the count of length one chains. Now consider a bigon or chain of bigons starting at an intersection point x and ending at an intersection point y, contributing a zig-zag chain to $C(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_{\bar{I}})$ (we do not assume the chain of bigons is maximal, so the zig-zag chain may be a subchain of a longer zig-zag chain in the complex). The portion of $D(\boldsymbol{\vartheta}_{\bar{I}})$ from the segment containing x to the segment containing y is $D(\boldsymbol{\vartheta}_{\bar{J}})$ for some subword $\bar{J}$ of $\bar{I}$. The portion of $A(\boldsymbol{\vartheta})$ from the segment containing x to the segment containing y runs parallel to $D(\boldsymbol{\vartheta}_{\bar{J}})$ and so by the definition of the bar operation on segments this portion of $A(\boldsymbol{\vartheta})$ is $A(\boldsymbol{\vartheta}_J)$. This indicates that J or J^{-1} appears in the cyclic word representing $\boldsymbol{\vartheta}$. Conversely, each time J or J^{-1} appears in the word representing $\boldsymbol{\vartheta}$ a corresponding subset of $A(\boldsymbol{\vartheta})$ agrees with $A(\boldsymbol{\vartheta}_J)$ and forms such a chain of bigons with the subset $D(\boldsymbol{\vartheta}_{\bar{J}})$ of $D(\boldsymbol{\vartheta}_{\bar{I}})$. It follows that chains of bigons of that particular form are precisely counted by $n_J(\boldsymbol{\vartheta})$, and the counts of all chains is determined by $n_J(\boldsymbol{\vartheta})$ for all J for which $\bar{J}$ is a subword of $\bar{I}$. If $n_J(\boldsymbol{\vartheta})$ is known for all J for which $\bar{J}$ is a strict subword of $\bar{I}$, then $n_I(\boldsymbol{\vartheta})$ can be determined from $\dim(H_*(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_{\bar{I}}))$.

It remains to show that the complex $C(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_{\bar{I}})$ breaks into zig-zag chains. This ultimately relies on the fact that $\boldsymbol{\vartheta}$ and $\boldsymbol{\vartheta}_{\bar{I}}$ are reduced, but it is especially apparent given the particular pairing position we assume for $C(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_{\bar{I}})$. In particular, recall that any bigon contributing to the differential in $C(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_{\bar{I}})$ must decompose into pieces of the forms shown in Figure 12. We can exclude the pieces of small bigons in the top row of the figure because $\boldsymbol{\vartheta}_{\bar{I}}$ is assumed to be reduced rather than only almost reduced. By inspecting these pieces it is straightforward to check that x cannot be both a starting and an ending point of a bigon. For example, if x is in the top left quadrant of the square, a bigon starting at x is only possible if the path in $A(\boldsymbol{\vartheta})$ starting rightward from x first leaves the square on the top edge or if the path in $D(\boldsymbol{\vartheta}_{\bar{I}})$ starting downward from x first leaves the square on the left edge, but a bigon ending at x is only possible if both these paths first leave the square on the right or bottom edge.

Inspecting the pieces in Figure 12 reveals that an intersection point x can be the start or end of at most one bigon covering a given quadrant locally near x. Indeed, the paths in $A(\boldsymbol{\vartheta})$ and $D(\boldsymbol{\vartheta}_{\bar{I}})$ starting from x in the directions required to surround a given quadrant near x uniquely determine a chain of pieces like those in Figure 12, which ends the first time the paths leave the square on different sides. A bigon is formed only if the paths eventually cross, so that the chain of pieces gets to an ending piece, and once this happens there is no way of adding more pieces to get a second bigon with a corner at x. More precisely, again by inspecting Figure 12, observe that both the left and right boundary of any bigon, viewed as paths from the starting intersection point to the ending intersection point, must move monotonically upward and leftward in the following sense: considering only the midpoints of the horizontal and vertical segments making up the path (recall that these are where intersections occur), the projection to the vector $(-1, 1)$ in the plane is non-decreasing, and it is strictly increasing except possibly at the beginning or end of the path (this happens in the case of the first or third starting piece or the first or third ending piece in Figure 12). Now suppose there are two bigons starting at x covering the same quadrant locally near x (the argument for bigons ending at x is similar). Call these bigons B_1 and B_2, call their endpoints y_1 and y_2, call

their left boundaries by $\partial_L B_1$ and $\partial_L B_2$ and call their right boundaries by $\partial_R B_1$ and $\partial_R B_2$. Let B_1 be the bigon with the shorter left boundary, so that $\partial_L B_1 \subset \partial_L B_2$; it can be shown that $\partial_R B_2$ must a subset of $\partial_R B_1$, since otherwise B_2 would have a region of negative multiplicity near y_1 (at least in an appropriate cover). In fact, $\partial_L B_2$ must intersect $\partial_R B_1$ at some other point y_3 between y_2 and y_1. But in this case y_3 occurs before y_1 on $\partial_R B_1$ and after y_1 on $\partial_L B_2$, violating the fact that both paths move monotonically in the upward/leftward direction. (Note that monotonicity of $\partial_L B_2$ implies that y_3 is strictly above/left from y_1 since neither of them is an end of $\partial_L B_2$). $\square$

The preceding lemma (and induction on the length of $\boldsymbol{\vartheta}_{\bar{I}}$) shows that if the dimension of the homology of $\mathcal{C}(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_{\bar{J}})$ is known for every word J, then the number of instances of I or I^{-1} in $\boldsymbol{\vartheta}$ can be determined for any word I. Since $\boldsymbol{\vartheta}$ has a finite number of curves that have a finite length, sufficiently long words will only occur a nonzero number of times if they repeat the entire cyclic word for one of the immersed curves in $\boldsymbol{\vartheta}$. In this way, the underlying immersed curves in $\boldsymbol{\vartheta}$ can be determined, with multiplicity.

It remains to show that if $\boldsymbol{\vartheta}_1$ and $\boldsymbol{\vartheta}_2$ are curve-like train tracks with the same underlying curves and multiplicities but with different local systems, then there is some $\boldsymbol{\vartheta}'$ with $\mathcal{C}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}') \not\simeq \mathcal{C}(\boldsymbol{\vartheta}_2, \boldsymbol{\vartheta}')$. Let γ be an underlying curve of $\boldsymbol{\vartheta}_1$ and $\boldsymbol{\vartheta}_2$ for which the local systems are non-isomorphic. We will take $\boldsymbol{\vartheta}'$ to be curve-like with underlying curve γ. Since the local systems on the other components of $\boldsymbol{\vartheta}_1$ and $\boldsymbol{\vartheta}_2$ have the same multiplicities, they will have the same pairing with $\boldsymbol{\vartheta}'$. Thus we may assume that $\boldsymbol{\vartheta}_i$ has a single underlying curve γ and that the monodromy of the local system is some $k \times k$ matrix A_i. Referring to the calculation in Corollary 4.10, we see that it suffices to prove the following result.

Proposition 4.15. *If A_1 and A_2 are $k \times k$ matrices over $\mathbb{F}$ which are not similar, then there exists a matrix B such that* $\mathrm{rk}(A_1 \otimes B + I \otimes I) \neq \mathrm{rk}(A_2 \otimes B + I \otimes I)$.

Proof. We outline the key steps but leave the details of the proof as an exercise to the reader. We assume that A_1, A_2, and B are written in rational canonical form, and we may choose B to have a single block, so that

$$B = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & b_1 \\ 0 & 1 & \cdots & 0 & b_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & b_{n-1} \end{bmatrix}$$

It is sufficient to consider the case that A_1 and A_2 have this form as well. It follows that

$$A \otimes B + I = \begin{bmatrix} I & 0 & \cdots & 0 & A \\ A & I & \cdots & 0 & b_1 A \\ 0 & A & \cdots & 0 & b_2 A \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & I + b_{n-1} A \end{bmatrix}$$

which row reduces to

$$\begin{bmatrix} I & 0 & \cdots & 0 & M_0 \\ 0 & I & \cdots & 0 & M_1 \\ 0 & 0 & \cdots & 0 & M_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & I + M_{n-1} \end{bmatrix}$$

where $M_0 = A$ and $M_i = A(b_i + M_{i-1})$ for $i > 0$. The rank of this is determined by the rank of the lower right block. We choose the coefficients in B so that $p(x) = x^n + b_1 x^{n-1} + \ldots + b_{n-1} x + 1$ is the minimal polynomial of A_1, which we assume without loss of generality does not divide the minimal polynomial of A_2. It follows that $A_1 \otimes B + I \otimes I$ has rank $k(n-1)$ and $A_2 \otimes B + I \otimes I$ has rank strictly greater than $k(n-1)$. $\square$

In summary, we have shown that the map g is injective: if $g(\boldsymbol{\gamma}_1) = g(\boldsymbol{\gamma}_2)$, then Lemma 4.14 shows that $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ have the same underlying curves with multiplicities, and Proposition 4.15 shows that the local systems on each curve are isomorphic.

We have now shown that g is both injective and surjective. Together with Proposition 4.13, this completes the proof of Theorem 4.11. $\hfill\square$

We conclude this section with the observation that, in particular, any two extensions of an extendable type D module correspond to the same decorated immersed curve and are thus homotopy equivalent.

Proposition 4.16. *An extendable type D module over $\mathcal{A}$ has a unique extension up to homotopy equivalence.*

5. Bordered Floer invariants as decorated immersed curves

We now return our attention to Floer theoretic invariants of manifolds with torus boundary. Our goal in this section is to prove Theorems 1.1 and 1.2. Suppose that (M, α, β) is a parametrized three-manifold with torus boundary. By Theorem 1.4 (proved in the Appendix), the type D structure $\widehat{CFD}(M, \alpha, \beta)$ is extendable. The results of the preceding two sections imply that there is a well-defined collection of immersed curves with local systems associated with $\widehat{CFD}(M, \alpha, \beta)$. This collection of curves lives in the abstract torus T, which we identify with T_M via a map that identifies the vertical edge of T with $\alpha \subset \partial M$ and the horizontal edge of T with $\beta \subset \partial M$. We denote the result by $f_{\alpha,\beta}(\widehat{CFD}(M, \alpha, \beta))$.

It remains to check that this collection of curves does not depend on the choice of parametrization (α, β).

Proposition 5.1. *There are equivalences*

(3) $$f_{\alpha,\beta}(\widehat{CFD}(M, \alpha, \beta)) \cong f_{\alpha,\beta\pm\alpha}(\widehat{CFD}(M, \alpha, \beta \pm \alpha))$$

(4) $$f_{\alpha,\beta}(\widehat{CFD}(M, \alpha, \beta)) \cong f_{\alpha\pm\beta,\beta}(\widehat{CFD}(M, \alpha \pm \beta, \beta))$$

where $\cong$ denotes regular homotopy of curves with local systems.

Proof. We will establish the equivalence (3); the equivalence (4) is nearly identical, and is left to the reader. To this end, we recall the bimodules associated with a Dehn twist established in [38]. These give rise to

$$\widehat{T}_\alpha^{\pm 1} \boxtimes \widehat{CFD}(M, \alpha, \beta) \cong \widehat{CFD}(M, \alpha, \beta \pm \alpha)$$

where $\widehat{T}_\alpha$ is the type DA bimodule denoted $\widehat{CFDA}(\tau_\mu)$ in the notation of [38, Section 10]. This bimodule, and its inverse, are shown in Figure 43. Thus we need to show that $f_{\alpha,\beta}(N) \cong f_{\alpha,\beta\pm\alpha}(\widehat{T}_\alpha^{\pm 1} \boxtimes N)$, where N is the extendable type D structure $\widehat{CFD}(M, \alpha, \beta)$.

We may work with one component of $\widehat{CFD}(M, \alpha, \beta)$ at a time, so we will assume that $f_{\alpha,\beta}(N)$ consists of a single immersed curve γ decorated by a local system. Let $\boldsymbol{\vartheta}$ be the corresponding curve-like train track, consisting of parallel copies of γ connected by crossover arrows. Recall that the valid reduced train track $\boldsymbol{\vartheta}$ determines an extendable type D structure which is homotopy equivalent to N; since we may modify N by homotopy equivalence, we will assume $\boldsymbol{\vartheta}$ corresponds directly to N. We can view the crossover arrows as all being concentrated in a single box, which slides freely along γ. By sliding this box, we can ensure that no crossover arrow has its head or tail on a segment of curve corresponding to a ρ_3 arrow.

We claim that the effect on $\boldsymbol{\vartheta}$ of tensoring N with $\widehat{T}_\alpha$ is easy to describe: it is applying a negative Dehn twist along α (and homotoping the result if necessary to achieve minimal intersection with β). Let $\boldsymbol{\vartheta}'$ be the train track obtained from $\boldsymbol{\vartheta}$ by applying a negative Dehn twist in a thin strip
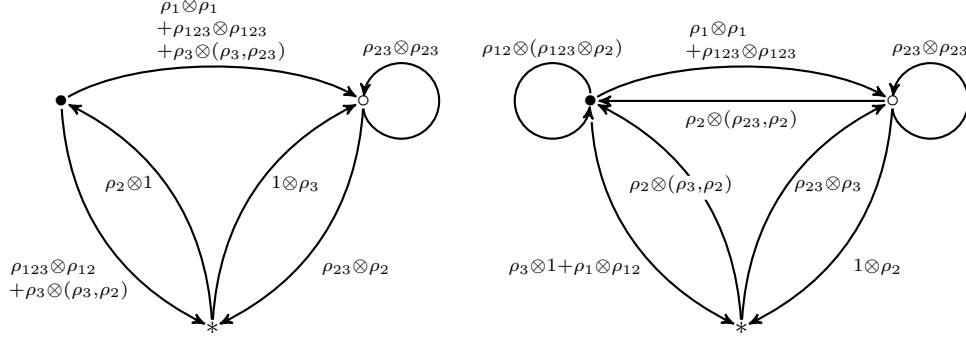
Figure 42. Graphical representations of the Dehn twist bimodules $\widehat{T}_\alpha$ (left) and $\widehat{T}_\alpha^{-1}$ (right), following [38, Section 10].

along the left edge of the square $T \setminus (\alpha \cup \beta)$ (see Figure 44). The effect of this transformation can be described as follows:

- For each intersection x with α we add a new intersection point x' with β, such that x and x' are connected by a segment connecting the bottom and left edges of the square;
- Each path from the bottom of the square to the left of the square ending at x becomes a path from the bottom to the top ending at x';
- Each path from the right of the square to the left ending at x becomes a path from the right to the top ending at x';
- Each path from the top of the square to the left ending at x becomes a path from the top edge of the square to itself ending at x'; and
- intersection points with β and curve segments that do not meet the left edge of the square are uneffected.

We can attempt to read off a type D structure from this new train track in the usual way, although we note that the result may not be a valid type D structure (it may fail to satisfy $\partial^2 = 0$). In any case, the result is precisely what is obtained from N by tensoring with $\widehat{T}_\alpha$, if we ignore the two operations in $\widehat{T}_\alpha$ with multiple $\mathcal{A}_\infty$ inputs. Indeed, referring to Figure 43, this has the following effect on N:

- For each ι_0 generator x we add a new ι_1 generator x' such that x and x' are connected by a ρ_2 arrow;
- Each ρ_2 arrow ending at x becomes a ρ_{23} arrow ending at x';
- Each ρ_{12} arrow ending at x becomes a ρ_{123} arrow ending at x';
- Each path ρ_3 arrow starting at x becomes an arrow labelled by $1 = \rho_\varnothing$ starting at x'; and
- ι_0 generators, ρ_1, ρ_{123} and ρ_{23} arrows are unchanged.

To be precise, the new generator x' corresponds to $* \boxtimes x$, while the new x (in $\widehat{T}_\alpha \boxtimes N$) corresponds to $\bullet \boxtimes x$. Note that each ρ_3 in N gives rise to a $\rho_\varnothing$ labelled arrow in the tensor product, and a corresponding bigon between $\boldsymbol{\vartheta}'$ and β. We will now use the assumption that $\boldsymbol{\vartheta}$ is curve-like and that there are no crossover arrows between segments corresponding to ρ_3 arrows in N; it follows that bigons between $\boldsymbol{\vartheta}'$ and β appear in one of the three configurations shown in Figure 45. These three configurations arise from ρ_3 arrows in $\boldsymbol{\vartheta}$ that are followed by, respectively, backwards ρ_1 arrows, ρ_{23} arrows, and ρ_2 arrows. Note that in the latter two configurations, adding the gray segments shown in the figure corresponds to incorporating the two operations of $\widehat{T}_\alpha$ that we ignored until now into the box tensor product with N—each (ρ_3, ρ_{23}) or (ρ_3, ρ_2) sequence in N gives becomes a ρ_3 in the tensor product. Let $\boldsymbol{\vartheta}''$ be the result of adding these gray lines to $\boldsymbol{\vartheta}'$. Note that $\boldsymbol{\vartheta}''$ is a valid (not necessarily reduced) train track and the corresponding type D structure is precisely $\widehat{T}_\alpha \boxtimes N$.
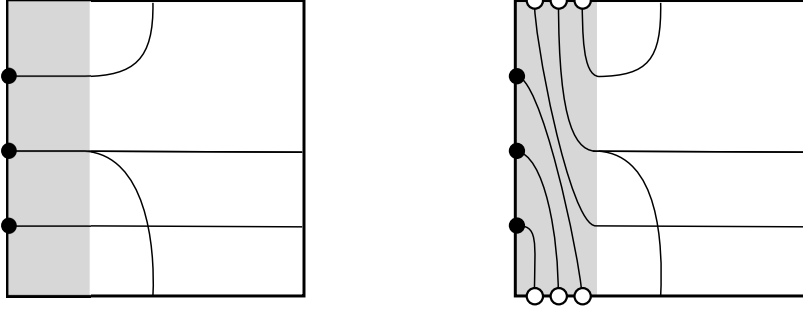
Figure 43. Applying a Dehn twist to a train track

Now consider the effect of removing the bigons between $\boldsymbol{\vartheta}''$ and β. In the first configuration, removing the bigons by homotoping the parallel immersed curves has precisely the effect of canceling $\rho_\varnothing$ arrows in the underlying type D structures. Each ρ_1 arrow into a generator z for which there is a $\rho_\varnothing$ arrow from x' to z and a ρ_2 arrow from x' to x becomes a ρ_{12} into x. In the second two configurations, the segments of $\boldsymbol{\vartheta}''$ forming the bigon can simply be deleted. This is because in the corresponding type D structure, the terminal endpoints of the $\rho_\varnothing$ arrows have no other incoming arrows, so the endpoints of the $\rho_\varnothing$ arrows can be canceled without adding any new arrows. In all three configurations, it is clear that the result of removing the bigons from $\boldsymbol{\vartheta}''$ in the appropriate way is simply the result of homotoping $\boldsymbol{\vartheta}'$ to be in minimal position with β.

We have shown that a curve-like train track associated with $\widehat{T}_\alpha \boxtimes N$ is obtained from $\boldsymbol{\vartheta}$ by applying a negative Dehn twist about α. Thus, interpreting curve-like train tracks as decorated immersed curves, we have that $f_{\alpha,\beta}(\widehat{T}_\alpha \boxtimes N)$ is obtained from $f(N)$ by applying the same Dehn twist to the underlying curve γ of $f(N)$. It follows that $f_{\alpha,\beta}(N) \cong f_{\alpha,\beta+\alpha}(\widehat{T}_\alpha \boxtimes N)$, since the change of marking cancels the action of the Dehn twist.

A similar argument applies to the bimodule $\widehat{T}_\alpha^{-1}$, where tensoring by this bimodule has the effect of applying a positive Dehn twist about α to the corresponding train track and to $f(N)$. However, it is easier to observe that $f_{\alpha,\beta}(N) = f_{\alpha,\beta}(\widehat{T}_\alpha \boxtimes \widehat{T}_\alpha^{-1} \boxtimes N)$ is obtained from $f_{\alpha,\beta}(\widehat{T}_\alpha^{-1} \boxtimes N)$ by applying a negative Dehn twist, so $f_{\alpha,\beta}(\widehat{T}_\alpha^{-1} \boxtimes N)$ is obtained from $f_{\alpha,\beta}(N)$ by applying a positive Dehn twist. Again, we see that $f_{\alpha,\beta}(N) \cong f_{\alpha,\beta-\alpha}(\widehat{T}_\alpha^{-1} \boxtimes N)$, since the change of marking cancels the action of the Dehn twist. $\qquad\square$

Since we can move between any two markings/bordered structures by a sequence of Dehn twists, Proposition 5.1 ensures that

$$f_{\alpha,\beta}(\widehat{CFD}(M,\alpha,\beta)) \cong f_{\alpha',\beta'}(\widehat{CFD}(M,\alpha',\beta'))$$

for any two pairs of parametrizing curves (α,β) and (α',β').

We will write

$$\widehat{HF}(M) = f_{\alpha,\beta}\left(\widehat{CFD}(M,\alpha,\beta)\right).$$

As we have shown, this definition is independent of the choice of parametrization (α,β). This completes the proof of Theorem 1.1.

We now turn to the proof of Theorem 1.2. Recalling the set-up of the theorem, suppose M_0 and M_1 are manifolds with torus boundary and $h\colon \partial M_1 \to \partial M_2$ is a gluing map. Chose a parametrization (α_1,β_1) for ∂M_1 and fix the parametrization (α_0,β_0) of ∂M_0 by setting $\alpha_0 = h(\beta_1)$ and $\beta_0 = h(\alpha_1)$. Following the conventions for bordered Heegaard Floer invariants (see Section 2.2), these parametrizations are consistent with the gluing map h; that is, by the pairing theorem in [39], $\widehat{HF}(M_0 \cup_h M_1)$ is given by the homology of $\widehat{CFA}(M_0,\alpha_0,\beta_0) \boxtimes \widehat{CFD}(M_1,\alpha_1,\beta_1)$. By the equivalence
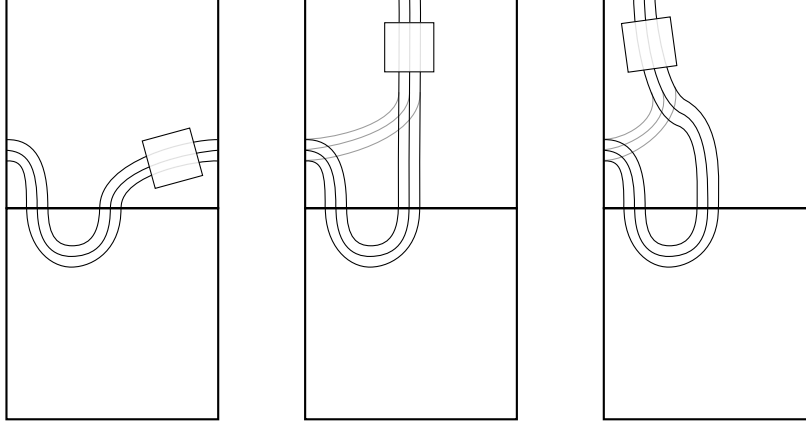
Figure 44. Possible configurations of bigons in $\boldsymbol{\vartheta}'$ corresponding to ρ_3 arrows in N that are followed by backward ρ_1 arrows, ρ_{23} arrows, and ρ_2 arrows, respectively. The assumption that $\boldsymbol{\vartheta}$ has no crossover arrows between segments corresponding to ρ_3 arrows ensures that the bigons do not involve any crossover arrows; there could be crossover arrows in the boxes shown. The black segments are part of $\boldsymbol{\vartheta}'$ and $\boldsymbol{\vartheta}''$ is obtained from $\boldsymbol{\vartheta}'$ by adding the gray segments in the second and third configuration.

of categories proved Section 4, the homology of this box tensor product agrees with the pairing of the decorated immersed curves $f(\widehat{CFD}(M_0, \alpha_0, \beta_0))$ and $f(\widehat{CFA}(M_1, \alpha_1, \beta_1))$. To pair these curves we include them in the same marked torus, reflecting one across the anti-diagonal, and take intersection Floer homology. Note that as 3-manifold invariants, we think of $\widehat{HF}(M_0) = f(\widehat{CFD}(M_0, \alpha_0, \beta_0))$ and $\widehat{HF}(M_1) = f(\widehat{CFA}(M_1, \alpha_1, \beta_1))$ as living in two different marked tori, the parametrized boundaries of M_0 and M_1. Identifying these marked tori with a reflection across the anti-diagonal corresponds to a gluing map which takes α_1 to $-\beta_0$ and β_1 to α_0; this is precisely the map $\bar{h}$, the elliptic involution composed with h. Letting $\boldsymbol{\gamma_0}$ denote $\widehat{HF}(M_0)$ and $\boldsymbol{\gamma_1}$ denote the image of $\widehat{HF}(M_1)$ under $\bar{h}$, the pairing is given by $HF(\boldsymbol{\gamma_0}, \boldsymbol{\gamma_1})$, as claimed.

Remark 5.2. Theorem 1.2, which follows from our graphical representation of type D structures in terms of train tracks (Section 2) and the structure theorem for these tracks in terms of immersed curves with local systems (Section 3), inherits some features from the box tensor product in bordered Floer homology. For instance, the appearance of $\bar{h}$ instead of h was initially a surprise to us. Of course, if $\widehat{HF}(M_1)$ is invariant under the elliptic involution, there is no difference between using h and $\bar{h}$. This symmetry is known to hold when M_1 is a knot complement in S^3 by a result of Xiu [51], and when M_1 is a graph manifold as discussed in [23]. In the companion to this paper [22], we show that it holds for all manifolds with torus boundary.

6. Gradings

The bordered invariants of M come equipped with a mod 2 and a Spin^c grading. In this section, we explain how the invariant $\widehat{HF}(M)$ can be enhanced to capture this information, leading to a proof of Theorem 1.7. While this does not capture all of the graded information in the bordered Heegaard Floer package, it is sufficient for the applications in this paper; a more thorough discussion of gradings, including the Maslov grading, appears in the companion to this paper [22].

6.1. **The $\mathbb{Z}/2\mathbb{Z}$ grading.** We first recall the $\mathbb{Z}/2\mathbb{Z}$ grading in bordered Floer homology. For each spin^c structure $\mathfrak{s}$, $\widehat{CFD}(M, \alpha, \beta; \mathfrak{s})$ admits a relative $\mathbb{Z}/2\mathbb{Z}$ grading gr^D as defined by Petkova in [45]. (This grading may be identified with a specialization of the full grading package on the bordered invariants; see [29, Appendix A].) The relative $\mathbb{Z}/2\mathbb{Z}$ grading satisfies $\text{gr}^D(\partial x) = \text{gr}^D(x) - 1$ and $\text{gr}^D(a \otimes x) = \text{gr}^D(a) + \text{gr}^D(x)$ for $a \in \mathcal{A}$, where $\text{gr}^D(\rho_1) = \text{gr}^D(\rho_3) = 0$ and $\text{gr}^D(\rho_2) = \text{gr}^D(\rho_{123}) =$

Edge type	Vertex sign	Orientation	Edge type	Vertex sign	Orientation
$\mathbf{I_\bullet}$	$+$	$\bullet \longleftarrow$	$\mathbf{I_\circ}$	$+$	$\circ \longrightarrow$
$\mathbf{I_\bullet}$	$-$	$\bullet \longrightarrow$	$\mathbf{I_\circ}$	$-$	$\circ \longleftarrow$
$\mathbf{II_\bullet}$	$+$	$\bullet \longrightarrow$	$\mathbf{II_\circ}$	$+$	$\circ \longleftarrow$
$\mathbf{II_\bullet}$	$-$	$\bullet \longleftarrow$	$\mathbf{II_\circ}$	$-$	$\circ \longrightarrow$

Table 1. Producing the oriented graph given a signed $\mathcal{A}$-decorated graph Γ and the vertex types from Figure 46.

$\mathrm{gr}^D(\rho_{12}) = \mathrm{gr}^D(\rho_{23}) = 1$. Note that if $\widehat{CFD}(M, \alpha, \beta; \mathfrak{s})$ is connected (see below), the relative $\mathbb{Z}/2\mathbb{Z}$ grading is completely determined by this condition. The generators of $\widehat{CFA}(M, \alpha, \beta; \mathfrak{s})$ inherit a grading gr^A from the corresponding generators in $\widehat{CFD}(M, \alpha, \beta; \mathfrak{s})$, where the grading of generators with idempotent ι_0 is reversed. A generator $x_0 \otimes x_1$ in a box tensor product $\widehat{CFA}(M_0, \alpha_0, \beta_0) \boxtimes \widehat{CFD}(M_1, \alpha_1, \beta_1)$ inherits the grading $\mathrm{gr}^A(x_0) + \mathrm{gr}^D(x_1)$, which recovers the relative $\mathbb{Z}/2\mathbb{Z}$ grading on $\widehat{CF}(M_0 \cup M_1)$.

The graph representing a $\mathbb{Z}/2\mathbb{Z}$ graded type D structure can be enhanced, replacing the vertex labeling $\{\bullet, \circ\}$ with $\{\bullet^+, \bullet^-, \circ^+, \circ^-\}$, where $+$ designates gr^A grading 0 and $-$ designates gr^A grading 1. Referring to the conventions above, we see that an edge labeled 2 or 123 joins two vertices with opposite sign, while an edge with any other label joins two vertices with the same sign.

We first observe that this data is equivalent to a choice of orientation on each edge of the underlying (undecorated) graph. Given a vertex v, we partition the edges incident to v into two types, as shown in Figure 46. That is, if v has label $\bullet$, incident edges are either of type $\mathbf{I_\bullet}$ or type $\mathbf{II_\bullet}$, and similarly if v has label $\circ$. This is a natural partition to consider in the context of extendable type D structures:
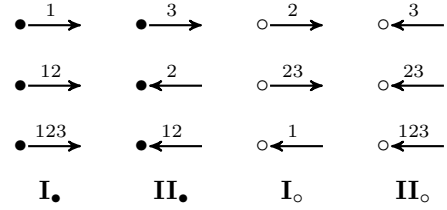
Figure 45. Vertex types for $\mathcal{A}$-decorated graphs.

Proposition 6.1. *Suppose that Γ is an $\mathcal{A}$-decorated graph associated with a reduced extendable type D structure N. Then every vertex $\bullet$ has at least one incident edge of type $\mathbf{I_\bullet}$ and at least one incident edge of type $\mathbf{II_\bullet}$, and every vertex $\circ$ has at least one incident edge of type $\mathbf{I_\circ}$ and at least one incident edge of type $\mathbf{II_\circ}$.*

Proof. This is essentially [23, Proposition 3.3]. $\square$

In particular, when Γ has valence 2 (and hence the associated train track is an immersed curve), there is exactly one of each edge type incident at every vertex.

This partition allows us to convert the system of signed vertices into an orientation on the underlying graph. The orientations on edges incident to any given vertex are assigned following the rules in Table 1. See Figure 47 for this process carried out in an example. Note that reversing the overall orientation on this graph corresponds to switching the grading of every vertex, and vice versa. For consistency with the gradings on the relevant bordered invariants, if an orientation in a given component is to be changed one must reverse orientations on every connected component corresponding to the relevant summand $\widehat{CFD}(M, \alpha, \beta; \mathfrak{s})$.

Lemma 6.2. *The conventions of Table 1 define a consistent orientation on each edge of the graph.*

Proof. It is enough to check that edges labelled 2 and 123 change the grading and that all other labelings preserve it. It is easy to see from the table that the two ends of an edge have the same
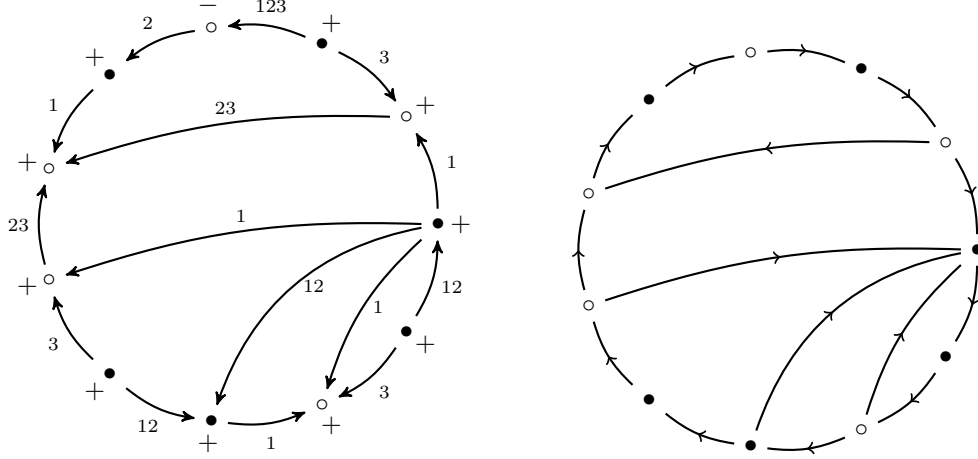
Figure 46. A $\mathbb{Z}/2\mathbb{Z}$ augmentation of an $\mathcal{A}$-decorated graph (left), and the associated orientation graph (right).

mod 2 grading if they are both of the same type (*i.e.* both **I** or both **II**). Referring to Figure 46, we see that edges labeled 2 and 123 have one end of type **I** and one of type **II**; other edges have ends of the same type. □

Remark 6.3. The condition from Proposition 6.1 ensures that every vertex of the graph has both an inward pointing and an outward pointing edge. If the graph is a loop corresponding to an immersed curve γ in T_M, it follows that the $\mathbb{Z}/2\mathbb{Z}$ grading is completely determined by the choice of an orientation on γ. More generally, if the graph corresponds to an immersed curve with a connected local system, all the parallel curves in the local system must carry the same orientation. Thus to specify the $\mathbb{Z}/2\mathbb{Z}$ grading it is again enough to specify an orientation on the underlying immersed curve.

6.2. **The Spin^c grading.** Now we review the Spin^c grading on bordered Floer homology. Recall that each generator x of $\widehat{CFD}(M, \alpha, \beta)$ has an associated spin^c structure $\mathfrak{s}(x) \in \mathrm{Spin}^c(M)$. The elements of $\mathrm{Spin}^c(M)$ are homology classes of nonvanishing vector fields on M, and $\mathrm{Spin}^c(M)$ has the structure of an affine set modeled on $H^2(M) \cong H_1(M, \partial M)$ (*i.e.* an $H^2(M)$–torsor.) The same decomposition holds for $\widehat{CFA}(M, \alpha, \beta)$, so that

$$\widehat{CFD}(M, \alpha, \beta) = \bigoplus_{\mathfrak{s} \in \mathrm{Spin}^c(M)} \widehat{CFD}(M, \alpha, \beta; \mathfrak{s}) \quad \text{and} \quad \widehat{CFA}(M, \alpha, \beta) = \bigoplus_{\mathfrak{s} \in \mathrm{Spin}^c(M)} \widehat{CFA}(M, \alpha, \beta; \mathfrak{s}).$$

Restricting attention to the generators in a particular idempotent ι, we can also define a refined spin^c grading $\bar{\mathfrak{s}}_\iota(x)$, which lives in an affine set $\mathrm{Spin}^c(M, \iota)$ modeled on $H^2(M, \partial M) \cong H_1(M)$. Elements of $\mathrm{Spin}^c(M, \iota)$ are homology classes of nonvanishing vector fields with prescribed behavior on ∂M, and $\mathfrak{s}(x)$ is the image of $\bar{\mathfrak{s}}_\iota(x)$ in $\mathrm{Spin}^c(M)$.

To compare the refined gradings of two generators, we adopt the following. Given $\mathfrak{s} \in \mathrm{Spin}^c(M)$, let

$$\mathrm{Spin}^c(M, \iota, \mathfrak{s}) = \{\bar{\mathfrak{s}} \in \mathrm{Spin}^c(M, \iota) \,|\, \bar{\mathfrak{s}} = \mathfrak{s} \text{ in } \mathrm{Spin}^c(M)\}.$$

If $j_* \colon H_1(\partial M) \to H_1(M)$ is the map induced by inclusion, $\mathrm{Spin}^c(M, \iota, \mathfrak{s})$ is an affine set modeled on $H_M = \mathrm{im}\, j_* \cong H_1(\partial M)/\ker j_*$. We define $\mathrm{Spin}^c(M, \mathfrak{s}) = \mathrm{Spin}^c(M, \iota_0, \mathfrak{s}) \cup \mathrm{Spin}^c(M, \iota_1, \mathfrak{s})$. If x is a generator of $\widehat{CFD}(M, \alpha, \beta, \mathfrak{s})$ in idempotent ι, we write $\bar{\mathfrak{s}}(x) = \bar{\mathfrak{s}}_\iota(x) \in \mathrm{Spin}^c(M, \mathfrak{s})$.

Let $2S \subset H_1(\partial M)$ be the kernel of the linear map $H_1(\partial M) \to \mathbb{Z}/2\mathbb{Z}$ given by $\alpha \mapsto 1$, $\beta \mapsto 1$, and consider the lattices $\frac{1}{2}H_1(M)$ and $S = \frac{1}{2}(2S)$ inside $H_1(\partial M; \mathbb{R})$. We have inclusions $H_1(M) \subset S \subset \frac{1}{2}H_1(M)$, and each lattice is index two in the next larger one. Let S_M be the image of S in $H_1(\partial M; \mathbb{R})/\ker j_*$ so we have inclusions $H_M \subset S_M \subset \frac{1}{2}H_M$, where each subgroup is index two in

Labeled edge	$\bar{\mathfrak{s}}(y) - \bar{\mathfrak{s}}(x)$	Labeled edge	$\bar{\mathfrak{s}}(y) - \bar{\mathfrak{s}}(x)$
$x \xrightarrow{\ 1\ } y$	$-(\alpha + \beta)/2$	$x \xrightarrow{\ 12\ } y$	$-\beta$
$x \xrightarrow{\ 2\ } y$	$(\alpha - \beta)/2$	$x \xrightarrow{\ 23\ } y$	α
$x \xrightarrow{\ 3\ } y$	$(\alpha + \beta)/2$	$x \xrightarrow{\ 123\ } y$	$(\alpha - \beta)/2$

Table 2. Grading shifts in $\widehat{CFD}(M, \alpha, \beta)$ associated with labelled edges.

the next larger one. (Note that S and S_M depend on α and β, although this is not reflected in the notation.)

Lemma 6.4. $\mathrm{Spin}^c(M, \mathfrak{s})$ *can be given the structure of an* S_M*–torsor such that (a) the action of* $H_M \subset S_M$ *agrees with the pre-existing action of* H_M *on* $\mathrm{Spin}^c(M, \mathfrak{s})$ *and (b) if* x, y *are generators of* $\widehat{CFD}(M, \alpha, \beta, \mathfrak{s})$ *that are joined by an arrow, then the difference* $\bar{\mathfrak{s}}(y) - \bar{\mathfrak{s}}(x)$ *is given by Table 2.*

Proof. This is a rephrasing of [46, Lemma 3.9]; compare [39, Lemma 11.42]. Specifically, the action of H_M on S_M has two cosets H and H'. Each coset is an H_M–torsor, and $H' = H - (\alpha + \beta)/2$. Choose some H_M-equivariant identification $\varphi \colon \mathrm{Spin}^c(M, \iota_0, \mathfrak{s}) \to H$. For $\bar{\mathfrak{s}} \in \mathrm{Spin}^c(M, \iota_1, \mathfrak{s})$ define $\varphi(\bar{\mathfrak{s}}) = \varphi(i^{-1}(\bar{\mathfrak{s}})) - (\alpha + \beta)/2$, where $i \colon \mathrm{Spin}^c(M, \iota_0, \mathfrak{s}) \to \mathrm{Spin}^c(M, \iota_1, \mathfrak{s})$ is the H_M-equivariant map defined in [46, Lemma 3.9]. Then $\varphi(\mathrm{Spin}^c(M, \iota_1, \mathfrak{s})) = H'$, so φ defines an H_M equivariant bijection between $\mathrm{Spin}^c(M, \mathfrak{s})$ and S_M. This gives $\mathrm{Spin}^c(M, \mathfrak{s})$ the structure of an S_M torsor satisfying property (a). Property (b) follows from the corresponding list of grading shifts in [46, Lemma 3.9]. $\qquad\square$

As in the introduction, we define $p \colon \overline{T}_M \to T_M$ to be the covering map defined by the condition that $\pi_1(\overline{T}_M)$ is the kernel of the composite map $\pi_1(T_M) \to H_1(T_M) \to H_1(\partial M) \to H_1(M)$. Equivalently, if ℓ generates the kernel of $j_* \colon H_1(\partial M; \mathbb{Z}) \to H_1(M; \mathbb{Z})$, then $\overline{T}_M$ is the quotient of $H_1(\partial M; \mathbb{R}) \setminus H_1(\partial M; \mathbb{Z})$ by the action of ℓ. The group of deck transformations on $\overline{T}_M$ is isomorphic to H_M. The class $\ell \in H_1(\partial M; \mathbb{Z})$ is the *homological longitude*. It is an integer multiple of a primitive class $\lambda \in H_1(\partial M; \mathbb{Z})$, which is called the *rational longitude*.

Let $q_\alpha, q_\beta \in T_M$ be the midpoints of the arcs α and β, and define $S_\alpha = p^{-1}(q_\alpha) \subset \overline{T}_M$, $S_\beta = p^{-1}(q_\beta) \subset \overline{T}_M$. Then $S_\alpha \cup S_\beta$ is naturally an S_M–torsor and the subsets S_α and S_β are identified with the H_M–cosets H and H'. By Lemma 6.4, the S_M–torsor $\mathrm{Spin}^c(M, \mathfrak{s})$ can be identified with $S_\alpha \cup S_\beta$ (as S_M–torsors) in such a way that $\mathrm{Spin}^c(M, \iota_0, \mathfrak{s})$ maps to S_α and $\mathrm{Spin}^c(M, \iota_1, \mathfrak{s})$ maps to S_β. Equivalently, by identifying an arc with its midpoint, we can identify $\mathrm{Spin}^c(M, \mathfrak{s})$ with the set of lifts of α and β to $\overline{T}_M$. We fix one such identification and write $\alpha_{\bar{\mathfrak{s}}}$ or $\beta_{\bar{\mathfrak{s}}}$ for the lift of α or β corresponding to $\bar{\mathfrak{s}}$. Note that any other such identification differs from our fixed one by the action of an element of H_M.

In Section 2 we explained how to assign a train track $\boldsymbol{\vartheta}(M, \alpha, \beta)$ in T_M to the type D structure $\widehat{CFD}(M, \alpha, \beta)$. Since the differential on $\widehat{CFD}(M, \alpha, \beta)$ respects the decomposition into Spin^c structures, $\boldsymbol{\vartheta}(M, \alpha, \beta)$ decomposes as a union of train tracks $\boldsymbol{\vartheta}(M, \alpha, \beta; \mathfrak{s})$, where $\mathfrak{s}$ runs over $\mathrm{Spin}^c(M)$.

Proposition 6.5. $\boldsymbol{\vartheta}(M, \alpha, \beta; \mathfrak{s})$ *lifts to a unique train track* $\overline{\boldsymbol{\vartheta}}(M, \alpha, \beta; \mathfrak{s})$ *in* $\overline{T}_M$ *with the property that if* x *is a switch of* $\boldsymbol{\vartheta}(M, \alpha, \beta; \mathfrak{s})$ *lying on* α *(resp.* β*) its lift lies on* $\alpha_{\bar{\mathfrak{s}}(x)}$ *(resp.* $\beta_{\bar{\mathfrak{s}}(x)}$*).*

Proof. We lift each switch x on α (resp. β) to its preimage on $\alpha_{\bar{\mathfrak{s}}(x)}$ (resp. $\beta_{\bar{\mathfrak{s}}(x)}$). It remains to check that we can lift each section of track, which corresponds to some arrow in $\widehat{CFD}(M, \alpha, \beta)$, in a way that is compatible with the lifts of its endpoints. Comparing the grading shifts in Table 2 with

Figure 8, we see that this is indeed the case. Uniqueness of $\overline{\boldsymbol{\vartheta}}(M, \alpha, \beta; \mathfrak{s})$ is clear, since the given condition determines the positions of all the switches in the lift. $\qquad\square$

Finally, note that if we had chosen a different identification of $\mathrm{Spin}^c(M, \mathfrak{s})$ with $S_\alpha \cup S_\beta$, the resulting lift of $\boldsymbol{\vartheta}(M, \alpha, \beta, \mathfrak{s})$ would differ from the one constructed above by the action of an element of H_M.

Proof of Theorem 1.7. Orient the track $\boldsymbol{\vartheta}(M, \alpha, \beta; \mathfrak{s})$ and lift it to a collection of oriented tracks $\overline{\boldsymbol{\vartheta}}(M, \alpha, \beta; \mathfrak{s})$ in $\overline{T}_M$. The result is a compact train track in the cover $\overline{T}_M$, which we now simplify as in Section 3. In fact, we can simplify $\overline{\boldsymbol{\vartheta}}(M, \alpha, \beta; \mathfrak{s})$ using exactly the same moves as we used to simplify $\boldsymbol{\vartheta}(M, \alpha, \beta)$ down in T_M. Each move we perform involves some arcs and/or crossover arrows that are supported in the 0-handle of the downstairs track. It is easy to see from the definition that all of the arcs involved lift up to a single 0-handle in the cover, so the move can be performed on the cover as well.

It is also easy to see that these moves respect the relative $\mathbb{Z}/2\mathbb{Z}$ grading. For example, consider move (M1), which eliminates a clockwise running crossover arrow which runs from x to y, as in Proposition 3.10. Let z be the generator connected to y by the two-way arc, so that ∂y contains a term of the form $\rho_J \otimes z$. The presence of the crossover arrow implies that ∂x contains a term of the form $\rho_{IJ} \otimes z$. It follows that x and $\rho_I \otimes y$ have the same $\mathbb{Z}/2\mathbb{Z}$ grading, so the change of basis in which we replace x by $x + \rho_I \otimes y$ respects the $\mathbb{Z}/2\mathbb{Z}$ grading. The argument for move (M2) is very similar.

We define $\overline{\boldsymbol{\vartheta}}(M, \mathfrak{s})$ to be this simplified train track; its image in T_M is the track obtained by simplifying $\boldsymbol{\vartheta}(M, \mathfrak{s})$. Finally, the same argument as in Section 5 shows that $\overline{\boldsymbol{\vartheta}}(M, \mathfrak{s})$ does not depend on the choice of parametrization. We denote the resulting invariant by $\widehat{HF}(M, \mathfrak{s})$. $\qquad\square$

Let $\widehat{T}_M = H_1(\partial M; \mathbb{R})/H_1(\partial M; \mathbb{Z})$ be T_M with the puncture filled in.

Corollary 6.6. *Each component of $\widehat{HF}(M)$ is homotopic to some multiple of ℓ in $\widehat{T}_M$. If $b_1(M) = 1$, $\widehat{HF}(M, \mathfrak{s})$ represents the homology class $\ell \in H_1(\widehat{T}_M)$ for each $\mathfrak{s} \in \mathrm{Spin}^c(M)$. If $b_1(M) > 1$, $\widehat{HF}(M, \mathfrak{s})$ is null homologous in $\widehat{T}_M$.*

Proof. Filling in the punctures of $\overline{T}_M$ gives a covering space $\widehat{\overline{T}}_M$ of $\widehat{T}_M$ that is homeomorphic to the cylinder $H_1(\partial M; \mathbb{R})/\langle \ell \rangle$. The image of a component of $\widehat{HF}(M, \mathfrak{s})$ in this covering space is homotopic to some multiple of ℓ. Pushing everything back down to $\widehat{T}_M$ gives the first statement.

For the second, recall that the $\mathbb{Z}/2$ grading induces a well-defined orientation on the components of $\widehat{HF}(M)$, so it makes senses to talk about the homology class it represents. To compute this class, we Dehn fill M along some slope α which is not the rational homological longitude, so $b_1(M(\alpha)) = b_1(M) - 1$. If $\mathfrak{s} \in \mathrm{Spin}^c(M(\alpha))$, $\widehat{HF}(M(\alpha), \mathfrak{s})$ is obtained by pairing $\widehat{HF}(M, \mathfrak{s}|_M)$ with the image of some line $L_{\alpha, \mathfrak{s}}$ with slope α in $\overline{T}_M$. Since α is not a multiple of ℓ, the image of $L_{\alpha, \mathfrak{s}}$ in $\overline{T}_M$ is noncompact and generates the Borel-Moore homology $H_1^{BM}(\widehat{\overline{T}}_M) \simeq \mathbb{Z}$. In particular, the intersection number $L_{\alpha, \mathfrak{s}} \cdot \ell = 1$, so if $[\widehat{HF}(M, \mathfrak{s})] = k\ell \in H_1(\widehat{\overline{T}}_M)$, then $L_{\alpha, \mathfrak{s}} \cdot \widehat{HF}(M, \mathfrak{s}) = k$. On the other hand, $L_{\alpha, \mathfrak{s}} \cdot \widehat{HF}(M, \mathfrak{s}) = \chi(\widehat{HF}(M(\alpha), \mathfrak{s}))$. The result now follows from the well-known fact that $\chi(\widehat{HF}(M(\alpha), \mathfrak{s})) = 1$ if $b_1(M(\alpha)) = 0$, and is 0 otherwise. $\qquad\square$

To sum up, for each $\mathfrak{s} \in \mathrm{Spin}^c(M)$, $\widehat{HF}(M, \mathfrak{s})$ is a well-defined collection of curves equipped with local systems in the cover $\overline{T}_M$. When we want to emphasize the fact that we are thinking of the copy of $\overline{T}_M$ containing $\widehat{HF}(M, \mathfrak{s})$, we will denote it by $\overline{T}_{M, \mathfrak{s}}$.

It is tempting to try to combine all the $\widehat{HF}(M, \mathfrak{s})$ for different $\mathfrak{s}$ into a single well-defined collection of curves in $\overline{T}_M$—for example, by identifying relative Spin^c structures with the same value of $\langle c_1(\bar{\mathfrak{s}}), x \rangle$, where x is a generator of $H_2(M, \partial M)$. Although this can sometimes be made to work, it is not possible in general. A good example to consider is when M is the twisted I-bundle over the Klein bottle (Example 7.9b) where $\mathrm{Spin}^c(M)$ consists of two Spin^c structures $\mathfrak{s}_0, \mathfrak{s}_1$ with the property that if $\bar{\mathfrak{s}} \in \mathrm{Spin}^c(M, \mathfrak{s}_i)$ then $\frac{1}{2}\langle c_1(\bar{\mathfrak{s}}), x \rangle \equiv i \mod 2$.

6.3. **The refined pairing theorem.** Suppose that $Y = M_0 \cup_h M_1$. The relative Spin^c grading on $\widehat{CF}(Y)$ can be recovered from the box tensor product $\widehat{CFA}(M_0, \alpha_0, \beta_0) \boxtimes \widehat{CFD}(M_1, \alpha_1, \beta_1)$ as follows. Consider the box tensor products $x_0 \boxtimes x_1$ and $y_0 \boxtimes y_1$, where x_0, y_0 in $\widehat{CFA}(M_0, \alpha_0, \beta_0)$ and x_1, y_1 in $\widehat{CFD}(M_1, \alpha_1, \beta_1)$. We have

$$\bar{\mathfrak{s}}(x_i) - \bar{\mathfrak{s}}(y_i) \in S_{M_i}.$$

Note that x_0 and y_0 are in the same idempotent if and only if x_1 and y_1 are, so either both differences are in H_{M_i} or neither one is. Recalling that $H_{M_i} = H_1(\partial M_i)/\ker j_{i,*}$, where $j_{i,*}$ is the map induced by the inclusion $j_i\colon H_1(\partial M_i) \to H_1(M_i)$, we see that the sum $\bar{\mathfrak{s}}(x_0) - \bar{\mathfrak{s}}(y_0) + \bar{\mathfrak{s}}(x_1) - \bar{\mathfrak{s}}(y_1)$ is a well defined element of

$$H_1(\partial M_0)/\left(\ker(j_{0,*}) \oplus \ker(j_{1,*})\right) \cong H_1(Y).$$

It is equal to the difference $\mathfrak{s}(x_0 \boxtimes x_1) - \mathfrak{s}(y_0 \boxtimes y_1)$.

Restriction gives a surjective map

$$\pi\colon \mathrm{Spin}^c(Y) \to \mathrm{Spin}^c(M_1) \times \mathrm{Spin}^c(M_2).$$

If $\mathfrak{s}_i \in \mathrm{Spin}^c(M_i)$, it is not hard to see that $\pi^{-1}(\mathfrak{s}_0 \times \mathfrak{s}_1)$ is a torsor over $H_Y = H_1(\partial M_0)/\langle \ell_0, h_*(\ell_1) \rangle)$, where ℓ_i is the homological longitude of M_i. Let $\overline{T}_Y$ be the covering space of T_{M_0} whose fundamental group is the kernel of the natural map $\pi_1(T_{M_0}) \to H_1(T_{M_0}) \to H_Y$; $\overline{T}_Y$ is the largest covering space of T_M that is covered by both $\overline{T}_{M_0}$ and $\overline{T}_{M_1}$.

Let $p_i\colon \overline{T}_{M_i} \to \overline{T}_Y$ be the projection, and let $\bar{p}_1$ be the composition with p_1 with the elliptic involution on $\overline{T}_Y$. (The elliptic involution is the map which descends from the map $w \mapsto -w$ on the universal cover. It is well defined up to the action of the deck group.) The images $p_0(\widehat{HF}(M_0, \mathfrak{s}_0))$ and $\bar{p}_1(\widehat{HF}(M_1, \mathfrak{s}_1))$ are well defined up to the action of the deck group H_Y.

Proposition 6.7. *Suppose that $\mathfrak{s} \in \mathrm{Spin}^c(Y)$ has $\pi(\mathfrak{s}) = \mathfrak{s}_0 \times \mathfrak{s}_1$. There is some $\alpha_{\mathfrak{s}} \in H_Y$ such that $\widehat{HF}(Y, \mathfrak{s})$ is given by the pairing of $p_0(\widehat{HF}(M_0, \mathfrak{s}_0))$ with $\alpha_{\mathfrak{s}} \cdot \bar{p}_1(\widehat{HF}(M_1, \mathfrak{s}_1))$ in $\overline{T}_Y$. Moreover, the relative $\mathbb{Z}/2\mathbb{Z}$ grading of a generator of $\widehat{HF}(Y, \mathfrak{s})$ is given by the sign of the corresponding intersection point.*

Proof. It suffices to check that the statement about Spin^c structures holds for the train tracks $\boldsymbol{\vartheta}(M_0; \mathfrak{s}_0), \boldsymbol{\vartheta}(M_1, \mathfrak{s}_1)$ (before simplification), since the pairings before and after simplification are the same. Suppose x_0, x_1 are generators of $\widehat{CFA}(M_0), \widehat{CFD}(M_1)$ that pair to give a generator of $\widehat{CF}(Y, \mathfrak{s})$. Then we must have $x_0 \in \widehat{CFA}(M_0, \mathfrak{s}_0)$ and $x_1 \in \widehat{CFD}(M_1, \mathfrak{s}_1)$, and we may choose $\alpha_{\mathfrak{s}}$ so that the intersection between the train tracks $p_0(\boldsymbol{\vartheta}(M_0); \mathfrak{s})$ and $\alpha_{\mathfrak{s}} \cdot \bar{p}_1(\boldsymbol{\vartheta}(M_1); \mathfrak{s})$ contains the intersection between the segments corresponding to x_0 and x_1. If y_0, y_1 are other generators of $\widehat{CFA}(M_0, \mathfrak{s}_0)$ and $\widehat{CFD}(M_1, \mathfrak{s}_1)$, $y_0 \boxtimes y_1$ will be a generator of $\widehat{CF}(Y, \mathfrak{s})$ if and only if the image of $\bar{\mathfrak{s}}(x_0) - \bar{\mathfrak{s}}(y_0) - \bar{\mathfrak{s}}(x_1) + \bar{\mathfrak{s}}(y_1)$ is 0 in $H_1(Y)$. This is equivalent to saying that $\bar{\mathfrak{s}}(x_0) - \bar{\mathfrak{s}}(y_0)$ and $\bar{\mathfrak{s}}(x_1) - \bar{\mathfrak{s}}(y_1)$ have the same image in $H_1(Y)$, and this occurs if and only if the segments corresponding to y_0 and y_1 intersect in $p_0(\boldsymbol{\vartheta}(M_0; \mathfrak{s}_0)) \cap \alpha_{\mathfrak{s}} \cdot \bar{p}_1(\boldsymbol{\vartheta}(M_1; \mathfrak{s}_1))$.

Next, consider the relative $\mathbb{Z}/2\mathbb{Z}$ grading. Recall that fixing the $\mathbb{Z}/2\mathbb{Z}$ grading on $\widehat{CFA}(M_0, \alpha_0, \beta_0)$ corresponds to choosing an orientation of each curve in $\boldsymbol{\gamma}_0 = \widehat{HF}(M_0)$. In particular, upward (respectively downward) oriented α_0 segments in $\boldsymbol{\gamma}_0$ correspond to $\circ^+$ (respectively $\circ^-$) generators of $\widehat{CFA}(M_0, \alpha_0, \beta_0)$, and rightward (respectively leftward) oriented β_0 segments in $\boldsymbol{\gamma}_0$ correspond to

$\bullet^+$ (respectively $\bullet^-$) generators of $\widehat{CFA}(M_0, \alpha_0, \beta_0)$. Fixing the $\mathbb{Z}/2\mathbb{Z}$ grading on $\widehat{CFD}(M_1, \alpha_1, \beta_1)$ corresponds to choosing orientations on $\boldsymbol{\gamma_1} = \widehat{HF}(M_1)$ with similar identification except that the grading is reversed on $\bullet$ generators. Reflecting across the anti-diagonal, it follows that upward (respectively downward) oriented segments in $\boldsymbol{\gamma}_1$ correspond to $\bullet^+$ (respectively $\bullet^-$) generators of $\widehat{CFD}(M_1, \alpha_1, \beta_1)$ and rightward (respectively leftward) oriented segments in $\boldsymbol{\gamma}_1$ correspond to $\circ^-$ (respectively $\circ^+$) generators of $\widehat{CFD}(M_1, \alpha_1, \beta_1)$. The (relative) $\mathbb{Z}/2\mathbb{Z}$ grading on $\widehat{HF}(M_0 \cup_h M_1)$ is given by $\mathrm{gr}(x \otimes y) = \mathrm{gr}^A(x) + \mathrm{gr}^D(y)$, while the grading of $HF(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1)$ is given by intersection signs. It is straightforward to check that generators of $\widehat{HF}(M_0 \cup_h M_1)$ of the form $\circ^+ \otimes \circ^+$, $\circ^- \otimes \circ^-$, $\bullet^+ \otimes \bullet^+$, and $\bullet^- \otimes \bullet^-$, which all have the same $\mathbb{Z}/2\mathbb{Z}$ grading, correspond exactly to positive intersection points and the remaining generators correspond to negative intersection points. □

6.4. **Examples.** We conclude with some examples of how the Spin^c grading is computed.

Example 6.8. Suppose that M is a homology $S^1 \times D^2$. In this case the homological longitude ℓ coincides with the rational longitude λ. Take $x \in H_2(M, \partial M)$ with $\partial x = \lambda$, and $\mu \in H_1(M)$ with $\mu \cdot \lambda = 1$, so that $\mu \cdot x = 1$. If we use the parametrization $\alpha = \mu$ and $\beta = \lambda$, we have

$$\iota_0 \widehat{CFD}(M, \mu, \lambda) \simeq SFH(M, \gamma_\mu) \simeq \widehat{HFK}(K_\mu)$$

where K_μ is the core of the Dehn filling $M(\mu)$.

There is a unique Spin^c structure $\mathfrak{s}$ on M, and the set $\mathrm{Spin}^c(M, \iota_0, \mathfrak{s})$ is a $\mathbb{Z}$–torsor. Let $\bar{\mathfrak{s}}_i \in \mathrm{Spin}^c(M, \mathfrak{s})$ be the relative Spin^c structure with $\langle c_1(\bar{\mathfrak{s}}), x \rangle = 2i$, and let μ_i be the associated lift of μ. There is a natural height function $h \colon \overline{T}_M \to \mathbb{R}$ such that the height of the midpoint of μ_i associated to $\bar{\mathfrak{s}}$ is i, and $HF(\widehat{HF}(M, \mathfrak{s}_0), \mu_i) = \widehat{HFK}(K_\mu, i)$ is the knot Floer homology of K_μ in Alexander grading i. Figure 48 illustrates the computation for $\widehat{HFK}(T(2, 3))$. The vertical segments in the middle of the diagram are μ_{-1}, μ_0, and μ_1, so there is one generator in each of $\bar{\mathfrak{s}}_{-1}, \bar{\mathfrak{s}}_0$, and $\bar{\mathfrak{s}}_1$.
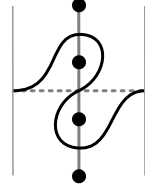
Figure 47. Lifts of μ and their intersections with the curve representing the right-hand trefoil.
.

More generally, the same method can be used to compute $\widehat{HFK}(K_\alpha)$ from $\widehat{HF}(M)$, where α is any primitive curve on ∂M. This and other relations between $\widehat{HF}(M)$ and knot Floer homology are discussed in [22].

Example 6.9. Now suppose that M_0 and M_1 are homology $S^1 \times D^2$'s, and that h identifies their homological longitudes, so that $H_1(Y) \simeq \mathbb{Z}$. There is a unique Spin^c structure $\mathfrak{s}_i$ on M_i, but $\mathrm{Spin}^c(Y)$ is a $\mathbb{Z}$-torsor, and $\overline{T}_Y = \overline{T}_{M_0} = \overline{T}_{M_1}$.

Let $\mathfrak{t}_i \in \mathrm{Spin}^c(Y)$ satisfy $\langle c_1(\mathfrak{t}_i), y \rangle = 2i$, where y generates $H_2(Y)$. Then $\widehat{HF}(Y, \mathfrak{t}_i) = HF(\gamma_0, \alpha_i \cdot \gamma_1)$ where $\gamma_0 = \widehat{HF}(M_0, \mathfrak{s}_0)$ and γ_1 is obtained by starting with $\widehat{HF}(M_1, \mathfrak{s}_1)$ and applying the elliptic involution, which sends μ_k to μ_{-k}. The action of α_i shifts the height by i units, so the net effect is to superimpose γ_0 and γ_1 in such a way that μ_k in the first diagram is identified with μ_{i-k} in the second.

In general, suppose we are given M whose rational longitude λ has order k in $H_1(M)$. Choose a class $x \in H_2(M, \partial M)$ with $\partial x = k\lambda$, and a class μ in $H_1(M)$ with $\mu \cdot \lambda = 1$, so that $\mu \cdot x = k$. Then $H_M \cong \mathbb{Z} \oplus \mathbb{Z}/k$, and there is a height function $h_x \colon \overline{T}_{M, \mathfrak{s}} \to \mathbb{R}$ that satisfies $h_x(q_{\bar{\mathfrak{s}}}) = \frac{1}{2}\langle c_1(\bar{\mathfrak{s}}), x \rangle$. If $\bar{\mathfrak{s}}, \bar{\mathfrak{s}}' \in \mathrm{Spin}^c(M, \iota_0, \mathfrak{s})$, then

$$h_x(q_{\bar{\mathfrak{s}}}) - h_x(q_{\bar{\mathfrak{s}}'}) = \frac{1}{2} PD(c_1(\bar{\mathfrak{s}}) - c_1(\bar{\mathfrak{s}}')) \cdot x$$

is a multiple of k. Hence all the heights for a given $\mathfrak{s}$ are congruent to a single value modulo k, and there are k meridians at each such height.

Example 6.10. Now suppose that we have two manifolds M_0 and M_1, with rational longitudes λ_i of order k_i, that are identified by h. Let $k = \gcd(k_1, k_2)$, and define $c_i = k_i/k$. Given $x_i \in H_2(M_i, \partial M_i)$ as above, we can form a primitive class $y = c_2 x_1 - c_1 x_2 \in H_2(Y)$. Given $\mathfrak{s}_i \in \mathrm{Spin}^c(M_i)$, let $\gamma_i = \widehat{HF}(M_i, \mathfrak{s}_i)$. Then for $\mathfrak{t} \in \pi^{-1}(\mathfrak{s}_0, \mathfrak{s}_1)$,

$$\widehat{HF}(Y, \mathfrak{t}) = HF(p_0(\gamma_0), \alpha_{\mathfrak{t}} \cdot \overline{p}_1(\gamma_1)),$$

where $\alpha_{\mathfrak{t}} \in H_Y = H_1(\partial M_0)/\langle k_0 \lambda, k_1 \lambda \rangle \simeq \mathbb{Z} \oplus \mathbb{Z}/k\mathbb{Z}$.

The height functions h_{x_i} on $\overline{T}_{M_i}$ induce height functions $h_{y,i} : \overline{T}_Y \to \mathbb{R}$ given by $h_{y,1}(p_1(w)) = c_2 h_{x_1}(w)$ and $h_{y,2}(p_1(w)) = -c_1 h_{x_1}(w)$. Then $\langle c_1(\mathfrak{t}), y \rangle$ measures the shift in heights induced by $\alpha_{\mathfrak{t}}$ so that

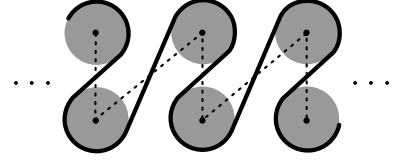$$\langle c_1(\mathfrak{t}), y \rangle = 2(h_{y,1}(\alpha_t \cdot w) - h_{y,2}(w))$$

for all $w \in \overline{T}_Y$.

7. Applications

With invariance (Theorem 1.1) and pairing (Theorem 1.2) for $\widehat{HF}(M)$ established, we can now provide proofs for the applications stated in the introduction.

7.1. Peg-board diagrams.
To facilitate the proofs, we will first discuss a convenient way of arranging the curves $\widehat{HF}(M)$.

Suppose γ is a component of $\widehat{HF}(M)$. To understand the pairing between γ and some other curve it is helpful to "pull the curves tight." Intuitively, we can imagine T_M as being a flat Euclidean torus (like a board with the sides identified) and the puncture as represented by a peg pounded into the board. We place a rubber band in the homotopy class determined by γ, and see what position it settles into. To draw this curve, it is convenient to first lift it to the cover $\overline{T}_M$ and

Figure 48. Pulling the invariant for the trefoil tight.

then take the preimage of this lift in the larger cover $\widetilde{T}_M$. The latter space can be identified with flat Euclidean $\mathbb{R}^2$ with a lattice of pegs—in other words, a peg-board. Hence we refer to these curves as *peg-board diagrams*. Figure 49 illustrates this process for the trefoil complement. We offer three different mathematical formulations of this idea, and explain how they are related.

We first describe the notion of an ϵ-*geodesic*. Fix a flat metric g on $\widehat{T}_M = H_1(\partial M; \mathbb{R})/H_1(\partial M, \mathbb{Z})$ by choosing a parametrization of ∂M and using it to identify $H_1(\partial M; \mathbb{R})$ with $\mathbb{R}^2$. Consider a manifold $T_{M,\epsilon}$ with a complete Riemannian metric g_ϵ defined as follows. $T_{M,\epsilon}$ is the union of two parts. The first part is the complement of the ball of radius 2ϵ centered at the origin in T_M, equipped with the flat metric g. The second part is modeled on a surface of revolution in $\mathbb{R}^3$, as illustrated in Figure 50. The smooth curve that we rotate to get the surface should agree with the x-axis for $x > 2\epsilon$, with the line $x = \epsilon$ for $y > \epsilon$, and should be given by the graph of some smooth function $y = f(x)$ with $f'(x) < 0$ and $f''(x) > 0$ for $x \in (\epsilon, 2\epsilon)$. We will refer to this second part of $T_{M,\epsilon}$ as the *peg*. The constructed surface $T_{M,\epsilon}$ is homeomorphic to T_M, and the conditions on f ensure that g_ϵ is non-positively curved. The flat part of $T_{M,\epsilon}$ has two connected components: one in the interior, and one that is boundary parallel. It is clear from the construction that $T_{M,\epsilon}$ embeds in $\widehat{T}_M \times \mathbb{R}$. Let $p \colon T_{M,\epsilon} \to T_M$ be the projection.

A curve $\gamma \subset T_{M,\epsilon}$ is an ϵ-*geodesic* if it is geodesic for the metric g_ϵ. We summarize the relevant properties of such curves here:

Lemma 7.1. *Any nontrivial free homotopy class of loops in T_M can be represented by an ϵ-geodesic γ. This representative is unique unless γ is entirely contained in the flat part of $T_{M,\epsilon}$, i.e. is either*
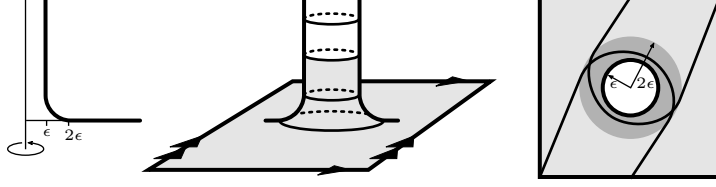
Figure 49. A geometric model for the punctured torus: Attaching an infinite tube of radius ϵ to the flat torus results in a complete metric that is non-positively curved and flat away from the smoothed attaching region. In the centre, we have shaded the compact region used in our appeal to Thorbergsson's work [49]. On the right, we have projected to the torus and illustrated the invariant associated with the right-hand trefoil exterior; compare Figure 49.

boundary parallel or a Euclidean geodesic in the interior flat part. Any two distinct ϵ-geodesics intersect minimally and transversally.

Proof. Consider the compact subset $K \subset T_{M,\epsilon}$ obtained as the complement of the boundary flat part. The only free homotopy classes that have representatives disjoint from K are boundary parallel. Such classes clearly have geodesic representatives—namely a meridional circle around the peg. By a theorem of Thorbegsson [49, Theorem 3.2(i)], any other free homotopy class has a geodesic representative γ. A standard argument using the Gauss–Bonnet theorem shows that if γ, γ' are two geodesic representatives in the same free homotopy class, they are entirely contained in the flat part of $T_{M,\epsilon}$. Such a geodesic is contained in either the interior flat part of the torus, in which case it is a line, or in the boundary flat part, in which case it is boundary parallel. For the last statement, note that two distinct geodesics always intersect transversally, and the theorem of Freedman, Hass, and Scott [13] ensures that they intersect minimally. □

We say that a geodesic contained in the flat part of $T_{M,\epsilon}$ is *loose*; all other geodesics are *tight*. A *corner* of an ϵ-geodesic γ_ϵ is a connected component of the intersections of γ_ϵ with the curved part of $T_{M,\epsilon}$. The property of being loose is equivalent to having no corners.

Given an ϵ-geodesic γ, we will construct another curve $\gamma' \subset T_M$ that is homotopic to $p(\gamma)$ and refer to γ' as an ϵ-*pegboard diagram*. The curve γ' is obtained by modifying the corners of γ, as illustrated in Figure 51. To make this precise, let $C_{2\epsilon}$ be the circle of radius 2ϵ centered on the puncture in T_M, so that $C_{2\epsilon}$ is the boundary of the interior flat part of $T_{M,\epsilon}$. If c is a corner of γ, let p_1 and p_2 be the two points at which c intersects $C_{2\epsilon}$. Let L be the perpendicular bisector of the line between p_1 and p_2, which also intersects the center of $C_{2\epsilon}$. Reflection in L induces an isometry of $T_{M,\epsilon}$ that preserves the endpoints of the geodesic c, and hence c itself. Let L_1 and L_2 be the Euclidean lines in T_M extending the parts of γ in the flat part of $T_{M,\epsilon}$. Reflection in L exchanges L_1 and L_2, so there is a circle C centered at the origin and tangent to L_1 and L_2. Since L_1 and L_2 cut $C_{2\epsilon}$, the radius of C is less than 2ϵ. We can homotope c to obtain a curve c' that runs along L_1 until it reaches C, runs along C for some time, and then exits along L_2.

Let $\theta(c)$ be the angle that L_1 makes with $C_{2\epsilon}$ (by symmetry, this is also the angle that L_2 makes with $C_{2\epsilon}$). There are two important quantities describing c that are determined by $\theta(c)$. The first is $r(c)$, the radius of the circle C. The second is the length of the arc obtained by projecting c radially outward to $C_{2\epsilon}$; we denote this quantity by $\psi(c)$. Observe that r is a strictly decreasing function of θ, while ψ is a strictly increasing function of θ. It follows that r is a strictly decreasing function of ψ.

If c_1 and c_2 are two corners with distinct endpoints, the intersection number $i(c_1, c_2)$, is the minimal number of intersections between transversally intersecting curves in $D_{2\epsilon} \setminus 0$ that are homotopic rel boundary to c_1 and c_2. We say c_1 and c_2 are in *minimal position* if they realize the intersection number. For example, if c_1 and c_2 are corners of ϵ-geodesics, they are in minimal position.
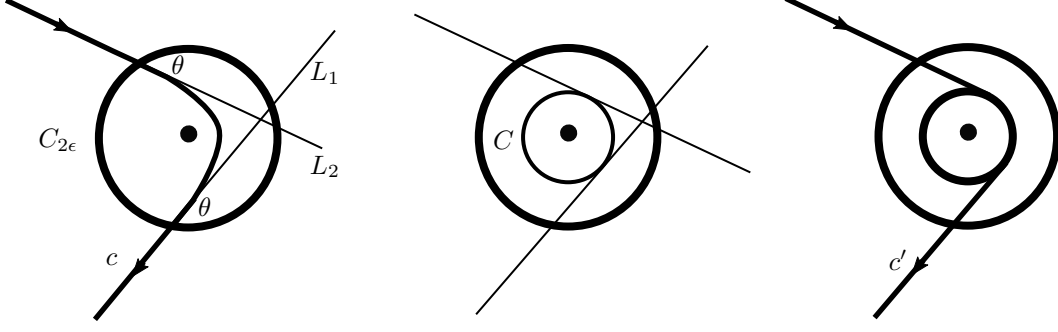
Figure 50. Constructing a pegboard corner c' from a geodesic corner c. The first step shows the original c together with the two lines L_1 and L_2. The second step shows the new circle C, and the last shows the new curve c'.

Lemma 7.2. *Suppose that c_1' and c_2' are two corners constructed as above from geodesic corners c_1 and c_2. If c_1' and c_2' intersect transversally, they are in minimal position.*

Proof. Identify the universal cover of the punctured disk with $\mathbb{R} \times (0, 2\epsilon]$. The curves c_1' and c_2' are in minimal position if $\tilde{c}_1'$ and $\tilde{c}_2'$ are in minimal position for every lift $\tilde{c}_1'$ of c_1' and every lift $\tilde{c}_2'$ of c_2'. The shape of $\tilde{c}_1'$ and $\tilde{c}_2'$ is illustrated in Figure 52. The depth of the trapezoidal shape made by the lift of c_i' is $2\epsilon - r(c_i)$, and its width is $\psi(c_i)$. Since r is a strictly decreasing function of ψ, the lifts have no excess intersection. $\qquad\square$
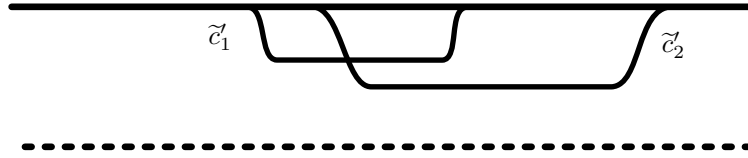
Figure 51. The universal cover of the punctured disk, with lifts of c_1' and c_2'. The key point is that the deeper shape is also wider. To compute the intersection number, we take the intersection of $\tilde{c}_1'$ with the orbit of $\tilde{c}_2'$ under the action of the deck group.

If $r(c_1) = r(c_2)$, c_1' and c_2' will not intersect transversally, but we can perturb them into minimal position by slightly changing $r(c_2)$. Since $r(c_1) = r(c_2)$ implies that $\psi(c_1) = \psi(c_2)$, c_1' and the perturbation of c_2' will be in minimal position as long as c_1 and c_2 have distinct endpoints.

If γ is an ϵ-geodesic, we define the ϵ-pegboard diagram γ' to be the curve obtained by applying the operation above to every corner of γ, so that an ϵ-pegboard diagram is composed of straight line segments together with circular arcs that are tangent to these line segments.

Corollary 7.3. *If γ_1' and γ_2' are the ϵ-pegboard diagrams induced by distinct ϵ-geodesics γ_1 and γ_2, then they are in minimal position.*

Proof. Since γ_1 and γ_2 are distinct ϵ-geodesics, they are in minimal position. By Lemma 7.2 γ_1' and γ_2' have the same number of intersections as γ_1 and γ_2. $\qquad\square$

Finally, we discuss the third notion of peg-board position, which corresponds to letting the radius of the peg go to 0. We use our chosen parametrization of T_M to identify a fundamental domain for T_M with the square $[-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$; the puncture is the point $(\pm\frac{1}{2}, \pm\frac{1}{2})$, the vertical sides are identified with α, and the horizontal sides are identified with β.

Any essential loop in T_M is free homotopic to a curve γ with no backtracking. Such a loop is represented by a cyclic word in $\alpha^{\pm 1}$ and $\beta^{\pm 1}$, which we also call γ. Recall that, by familiar abuse of

notation, the letters of the word γ correspond to the intersections of the curve γ with the horizontal and vertical sides of the square; see Figure 53, below. More precisely, if the ith letter is $\alpha^{\pm 1}$, the corresponding intersection point has coordinates $\mathbf{v}_i = (\pm\frac{1}{2}, x_i)$; if it is $\beta^{\pm 1}$, the intersection has coordinates $\mathbf{v}_i = (x_i, \pm\frac{1}{2})$ for some $x_i \in [-\frac{1}{2}, \frac{1}{2}]$. (Note that with this notation, the generators α and β of π_1 are *dual* to the corresponding parametrizing curves.)

The curve γ is homotopic to a piecewise linear curve obtained by replacing the segment of γ between $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$ with a line segment. Conversely, suppose that the word γ has n letters. Given $\mathbf{x} \in [-\frac{1}{2}, \frac{1}{2}]^n$, we let $\gamma_{\mathbf{x}}$ be the piecewise linear loop in T whose vertices are the points $\mathbf{v}_i$. Consider $F \colon [-\frac{1}{2}, \frac{1}{2}]^n \to \mathbb{R}$ given by $F(\mathbf{x}) = \ell(\gamma_{\mathbf{x}})$, the length of $\gamma_{\mathbf{x}}$. We can express this length as follows. If we cut T_M open to form the square $[-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$, the curve $\gamma_{\mathbf{x}}$ gets cut into a sequence of oriented line segments. Each point $\mathbf{v}_i$ has two preimages $(\pm\frac{1}{2}, x_i)$ in the square. We label these $\mathbf{v}_i^{\mathrm{in}}$, $\mathbf{v}_i^{\mathrm{out}}$ in such a way that the ith segment runs from $\mathbf{v}_i^{\mathrm{out}}$ to $\mathbf{v}_{i+1}^{\mathrm{in}}$. Then $F(\mathbf{x}) = \sum_{i=1}^{n} d(\mathbf{v}_i^{\mathrm{out}}, \mathbf{v}_{i+1}^{\mathrm{in}})$, where d denotes the Euclidean distance, and $\mathbf{v}_{n+1} = \mathbf{v}_1$.

Note that since we allow $x_i \in [-\frac{1}{2}, \frac{1}{2}]$, the points $\mathbf{v}_i^{\mathrm{in}}$, $\mathbf{v}_i^{\mathrm{out}}$ may lie at the corners of the square corresponding to the punctures. If we have two consecutive points $\mathbf{v}_i^{\mathrm{out}}$, $\mathbf{v}_{i+1}^{\mathrm{in}}$ which lie on the same corner of the square, the corresponding line segment can have zero length.

Definition 7.4. If $\mathbf{x}$ is a minimum for the function F we call $\gamma_{\mathbf{x}}$ a *singular pegboard diagram* for γ.

Since the domain of F is compact, every essential loop γ has at least one singular pegboard diagram $\gamma_{\mathbf{x}}$. Note that $\gamma_{\mathbf{x}}$ is not actually a curve in T_M unless $\mathbf{x}$ is in the interior of $[-\frac{1}{2}, \frac{1}{2}]^n$. However $\gamma_{\mathbf{x}}$ is homotopic to γ in $\widehat{T}_M$, so $\gamma_{\mathbf{x}}$ lifts to the covering space $\widehat{\overline{T}}_M$ obtained by filling in the punctures of $\overline{T}_M$. If $\mathbf{v}_i$ is a vertex of $\gamma_{\mathbf{x}}$ for which $x_i = \pm\frac{1}{2}$, we say $\mathbf{v}_i$ is a *corner* of $\gamma_{\mathbf{x}}$.
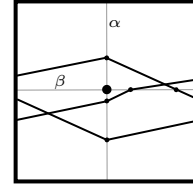
Note that it is possible for multiple points $\mathbf{x}$ to give the same underlying curve $\gamma_{\mathbf{x}}$. This occurs if a segment of $\gamma_{\mathbf{x}}$ containing three or more vertices lies along one of the parametrizing curves, say α. In this case we can move the interior vertices that lie on α freely along α without changing the underlying curve; see the proof of Proposition 7.5, below.

If $\mathbf{x}$ is an unique global minimum for F, we say that $\gamma_{\mathbf{x}}$ is *strongly tight*. If $\mathbf{x}$ is a global minimum for F and every other global minimum gives rise to the same underlying curve $\gamma_{\mathbf{x}}$ (as described in the paragraph above), we say $\gamma_{\mathbf{x}}$ is *weakly tight*. Finally, if $\gamma_{\mathbf{x}}$ is a line in T_M and can be freely deformed to nearby parallel geodesics we say $\gamma_{\mathbf{x}}$ is *loose*. Observe that if γ_{ϵ} is a loose ϵ-pegboard diagram for γ, the same curve is also a loose singular pegboard diagram for γ.

Figure 52. A weakly tight curve is obtained from the word $\alpha\beta^{-1}\alpha^2\beta$. To make the curve easier to see, we've drawn a fundamental domain with the puncture in the middle.

Proposition 7.5. *If $\mathbf{x}$ is a local minimum for F, then $\gamma_{\mathbf{x}}$ is either strongly tight, weakly tight, or loose.*

Proof. Write $F = \sum_{i=1}^{n} f_i$, where either $f_i = \left(1 + (x_i - x_{i+1})^2\right)^{1/2}$ or $f_i = \left((x_i \pm \frac{1}{2})^2 + (x_{i+1} \pm \frac{1}{2})\right)^{1/2}$. Since the functions $f(u) = (1 + u^2)^{1/2}$ and $g(u, v) = (u^2 + v^2)$ are both convex, it follows that f_i and F are convex functions as well. The convexity of F places restrictions on its minima: Any local minimum $\mathbf{x}$ is necessarily a global minimum, and if $\mathbf{x}$ and $\mathbf{x}'$ are two global minima, F is constant along the line segment that joins $\mathbf{x}$ and $\mathbf{x}'$. We refer to such a segment as a *null line* for F.

Suppose that $\mathbf{x} + t\mathbf{w}$, $t \in [0, 1]$, is a null line for F, and let $\gamma_{\mathbf{x}+t\mathbf{w}}$ be the corresponding deformation of $\gamma_{\mathbf{x}}$. We divide the vertices of $\gamma_{\mathbf{x}+t\mathbf{w}}$ into two types: those for which $\mathbf{v}_i(t)$ moves as t varies, and those for which it stays fixed. We claim that if there is one fixed vertex, then the underlying curve $\gamma_{\mathbf{x}+t\mathbf{w}}$ stays fixed as t varies. To see this, suppose we have two fixed vertices $\mathbf{v}_i$ and $\mathbf{v}_j$, separated by some number of moving vertices. (Since the $\mathbf{v}_i$ are cyclically ordered, we can take $\mathbf{v}_i = \mathbf{v}_j$ if

necessary, so one fixed vertex suffices.) Since $\mathbf{v}_{i+1}$ is a moving vertex, we can find some $t \in [0,1]$ for which $\mathbf{v}_{i+1}(t)$ lies in the interior of the interval on which it moves. But for every $t \in [0,1]$, $\mathbf{x} + t\mathbf{w}$ is a global minimum for F. It follows that $\mathbf{v}_i(t), \mathbf{v}_{i+1}(t)$, and $\mathbf{v}_{i+2}(t)$ must lie on a line; otherwise we could shorten $\gamma_{\mathbf{x}+t\mathbf{w}}$ by moving $\mathbf{v}_{i+1}(t)$ and fixing the other $\mathbf{v}_i(t)$. Repeating, we find that $\mathbf{v}_i(t), \mathbf{v}_{i+1}(t), \ldots, \mathbf{v}_j(t)$ are all collinear. But $\mathbf{v}_i$ and $\mathbf{v}_j$ were fixed by hypothesis, so the line they lie on is fixed as well. If this line happens to coincide with an α or β curve, then the intervening points can move along it. In this case $\gamma_{\mathbf{x}}$ is weakly tight. Otherwise, they are fixed as well, and if this is the case for all segments of $\gamma_{\mathbf{x}}$, it is strongly tight. In either case, the underlying curve remains the same.

Finally, suppose that every vertex is moving, so we can find some small t such that all the $\mathbf{v}_i$ lie in the interior of the interval in which they move. The same argument as above shows that the $\mathbf{v}_i$ are all collinear, so the curve is loose. $\qquad\square$

We relate the different notions of pegboard representative by showing that when ϵ is small, an ϵ-geodesic for a curve γ is close to a singular pegboard diagram for γ. We say that curves $\gamma_1, \gamma_2 \subset \widehat{T}_M$ are δ-close if they can be parametrized so that $|\gamma_1(t) - \gamma_2(t)| < \delta$ for all t.

Proposition 7.6. *Suppose $\gamma_{\mathbf{x}}$ is a singular pegboard representative for γ and that we are given $\delta > 0$. Then there exists an $\epsilon_0 > 0$ such that, for every $\epsilon < \epsilon_0$, there is a ϵ-geodesic γ_ϵ representing γ for which $p(\gamma_\epsilon)$ is δ-close to $\gamma_{\mathbf{x}}$.*

Proof. Suppose that $\gamma_{\mathbf{x}}$ is tight. Let X_ϵ be the image of the map $p\colon T_{M,\epsilon} \to T$; it is the complement of an open ball of radius ϵ centered at the origin. There is a unique continuous map $s\colon X_\epsilon \to T_{M,\epsilon}$ satisfying $p \circ s = 1_{X_\epsilon}$. Let γ be a path in $T_{M,\epsilon}$. Since $T_{M,\epsilon}$ embeds in $\widehat{T}_M \times \mathbb{R}$, the length of γ is greater than or equal to the length of $p(\gamma)$.

Let $C(r)$ be the cube $[-r,r]^n$, and let $C^\circ(r)$ be its interior. Define $F_\epsilon\colon C(\frac{1}{2} - \epsilon) \to \mathbb{R}$ by $F_\epsilon(\mathbf{x}) = \sum_{i=1}^n d_\epsilon(s(\mathbf{v}_i), s(\mathbf{v}_{i+1}))$, where d_ϵ denotes the Riemannian distance in $T_{M,\epsilon}$. Said another way, $F_\epsilon(\mathbf{x})$ is the length of the piecewise geodesic arc $\gamma_{\mathbf{x},\epsilon}$ in $T_{M,\epsilon}$ obtained by joining the points $s(\mathbf{v}_i)$ and $s(\mathbf{v}_{i+1})$ by geodesic segments.

F_ϵ has the following properties. First, the fact that $\ell(\gamma) \geq \ell(p(\gamma))$ implies that $F_\epsilon \geq F$. Second, if $|x_i|, |x_{i+1}| < \frac{1}{2} - 4\epsilon$, then the geodesic between $s(\mathbf{v}_i)$ and $s(\mathbf{v}_{i+1})$ is the straight line in T_M. Hence $F = F_\epsilon$ on $C(\frac{1}{2} - 4\epsilon)$. Third, if $\mathbf{x}$ is a local minimum for F_ϵ, then $\gamma_{\mathbf{x},\epsilon}$ is a geodesic. This follows from the usual principle that rounding a corner of a piecewise geodesic segment reduces its length.

Fix some $r > 0$, and consider:

$$N_r = \{\mathbf{z} \in C(1/2) \mid d(\mathbf{x}, \mathbf{z}) \leq r \text{ for some global minimum } \mathbf{x} \text{ for } F\}$$

$$N_r^\circ = \{\mathbf{z} \in C(1/2) \mid d(\mathbf{x}, \mathbf{z}) < r \text{ for some global minimum } \mathbf{x} \text{ for } F\}$$

Set $Y_r = N_r \setminus N_r^\circ$ so that N_r and Y_r are closed subsets of $C(\frac{1}{2})$ while N_r° is an open subset. No point of Y_r is a global minimum, so $F(\mathbf{y}) > F(\mathbf{x})$ for all $\mathbf{y} \in Y_r$. Y_r is compact, so there is some $\eta > 0$ such that $F(\mathbf{y}) \geq F(\mathbf{x}) + \eta$ for $\mathbf{y} \in Y_r$. Since F is continuous, we can find some $\mathbf{z} \in N_r \cap C^\circ(\frac{1}{2})$ with $F(\mathbf{z}) < F(\mathbf{x}) + \eta$. Finally, choose $\epsilon_0 > 0$ so that $\mathbf{z} \in C^\circ(\frac{1}{2} - 4\epsilon_0)$. Suppose that $\epsilon < \epsilon_0$ and let $\mathbf{z}'$ be a global minimum point for F_ϵ on the compact set $N_r \cap C(\frac{1}{2} - \epsilon)$. Using the first and second properties above, we see that for $\mathbf{y} \in Y_r$,

$$F_\epsilon(\mathbf{y}) \geq F(\mathbf{y}) > F(\mathbf{z}) = F_\epsilon(\mathbf{z}).$$

Hence $\mathbf{z}' \in N_r^\circ \cap C(\frac{1}{2} - \epsilon)$, which implies that $\mathbf{z}'$ is a local minimum for F_ϵ. By the third property, $\gamma_{\mathbf{z}',\epsilon}$ is a geodesic. Now if $\epsilon < \delta/4$ it is easy to see that there is some $r > 0$ such that $p(\gamma_{\mathbf{z},\epsilon})$ and $\gamma_{\mathbf{x}}$ are δ-close whenever $\|\mathbf{z} - \mathbf{x}\| < r$. Taking this value of r in the argument above gives the desired statement.

If $\gamma_{\mathbf{x}}$ is loose, the statement is easy: we choose ϵ_0 small enough that there is a line in the flat part of T_{M,ϵ_0} that is δ-close to $\gamma_{\mathbf{x}}$.   □

Corollary 7.7. *Suppose that γ_0 (resp. γ_ϵ) is a singular pegboard diagram (resp. ϵ-pegboard diagram) for γ. If γ_0 is tight, then $p(\gamma_\epsilon) \to \gamma_0$ in the Hausdorff metric as $\epsilon \to 0$.*

Proof. If $\epsilon_i \to 0$, Proposition 7.6 tells us that there is some sequence of ϵ_i-geodesics $\{\gamma_{\epsilon_i}\}$ such that $p(\gamma_{\epsilon_i})$ converges to γ_0 in the Hausdorff metric. As γ_0 is tight, γ_{ϵ_i} is tight for all ϵ_i. Hence γ_{ϵ_i} is the unique g_{ϵ_i} geodesic in the homotopy class of γ. So if $\{\gamma'_{\epsilon_i}\}$ is any sequence of ϵ_i-geodesics for γ, we must have $\{\gamma'_{\epsilon_i}\} = \{\gamma_{\epsilon_i}\}$.   □

If γ_1, γ_2 are ϵ-geodesics (or pegboard diagrams), we can write $i(\gamma_1, \gamma_2) = i^\circ(\gamma_1, \gamma_2) + \sum_{(c_1,c_2)} i(c_1, c_2)$ where $i^\circ(\gamma_1, \gamma_2)$ is the intersection number of the flat parts, of γ_1 and γ_2, and the the sum runs over all pairs of corners (c_1, c_2) for γ_1 and γ_2.

For some arguments it is useful to have a similar formula for singular pegboard diagrams, but when we pass to the limit $\epsilon \to 0$ to obtain a singular pegboard diagram $\overline{\gamma}$, we lose information about the homotopy class of the corners. To be precise, the underlying curve $\overline{\gamma}$ does not depend on the homotopy class of the corners. However, if we remember not just $\overline{\gamma}$, but the minimizer $\mathbf{x}$ that determines it, we can recover the homotopy class from the number of consecutive vertices $\mathbf{v}_i$ with $d(\mathbf{v}_i, \mathbf{v}_{i+1}) = 0$ at the corner. Taking the limit of the formula above as $\epsilon \to 0$, we obtain:

Corollary 7.8. *If $\overline{\gamma}_1, \overline{\gamma}_2$ are singular pegboard diagrams such that no slope of $\overline{\gamma}_1$ is also a slope of $\overline{\gamma}_2$, then*

$$i(\gamma_1, \gamma_2) = i^\circ(\overline{\gamma}_1, \overline{\gamma}_2) + \sum_{(c_1,c_2)} i(c_1, c_2)$$

where $i^\circ(\overline{\gamma}_1, \overline{\gamma}_2)$ is the number of intersections away from the corners, and $i(c_1, c_2)$ is a local intersection number determined by the homotopy data at the corners.   □

To specify the homotopy class at a corner, we choose a small disk D centered at the puncture and project a path realizing the homotopy class out to ∂D. The resulting homotopy class is determined by the total angle ϕ that it covers. (The angle ϕ plays the same role as the quantity ψ for which we used for ϵ-geodesics.) Note that a corner with angle $|\phi| < \pi$ cannot be realized as part of a singular pegboard diagram (since the resulting curve can be shortened). Hence the smallest possible angle always has $\pi \le |\phi| \le 2\pi$. If the angle is $\ge 2\pi$, we say that γ *wraps the peg* at that corner. The boundary case where $\phi = \pi$ often requires special treatment in later arguments: we refer to such a corner as a π-*corner* (or, sometimes, a *straight corner*). When a distinction is required, non-π-corners then will be called *true corners*.

If neither c_1 nor c_2 wraps the peg, the local intersection number $i(c_1, c_2)$ is easily described. Let A_1 and A_2 be the arcs specifying the homotopy classes of c_1 and c_2. Then we have:

$$i(c_1, c_2) = \begin{cases} 0 & \text{if } A_1 \text{ and } A_2 \text{ are nested} \\ 1 & \text{if } A_1 \text{ and } A_2 \text{ are not nested and } A_1 \cap A_2 \text{ has one component} \\ 2 & \text{if } A_1 \text{ and } A_2 \text{ are not nested and } A_1 \cap A_2 \text{ has two components} \end{cases}$$

7.2. Solid torus like curves. Having introduced peg-board diagrams, we pause to discuss two key classes of immersed curves that arise. These are illustrated by the following two examples:

Example 7.9. (a) If M is the solid torus $S^1 \times D^2$, then $\widehat{HF}(M)$ is the longitude $\partial D^2 \times \{\mathrm{pt}\}$ decorated with the trivial 1-dimensional local system.
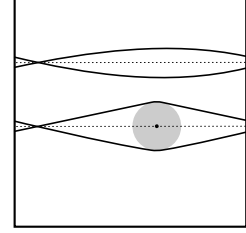
(b) If M is the twisted I-bundle over the Klein bottle, $\widehat{CFD}(M)$ was computed in [10]; see also [23]. In our conventions, as an $\mathcal{A}$-decorated graph, it has two connected components that, with an appropriate choice of bordered structure, can be described as follows:
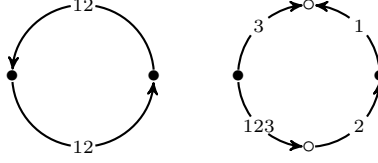
Both components are loop-type, and it is easy to see that $\widehat{HF}(M)$ consists of the two immersed curves shown in Figure 54, each decorated with the trivial local system.

The invariant of the solid torus pulls tight to a singular pegboard diagram lying on a line, and is loose. In contrast, both curves for the twisted I-bundle over the Klein bottle lie on a single line when you pull them tight, but only one of them is a loose curve. The other curve (the lower one in the figure) has two π-corners. The manifolds M for which every component of $\widehat{HF}(M)$ has this form form an interesting class. Consider the following definition, which was first made in [50] (see also [23]):

Figure 53. Pulling the invariant associated with the twisted I-bundle over the Klein bottle tight.

Definition 7.10. Suppose M is a three-manifold with torus boundary, that $\lambda \in H_1(\partial M)$ is the rational longitude, and that $\mu \cdot \lambda = 1$. We say that M is a *Heegaard Floer homology solid torus* if $\widehat{CFD}(M, \mu, \lambda) \cong \widehat{CFD}(M, \mu + \lambda, \lambda)$.

A longer discussion of this class of manifolds is given in [22, Section 1.5]. For our purposes, this definition can be rephrased in terms of curves; compare [22, Theorem 27].

Proposition 7.11. *M is a Heegaard Floer homology solid torus if and only if the immersed multicurve $\widehat{HF}(M)$ is homotopic to a curve that lies in a small neighbourhood of a representative λ running through the puncture.*

Before giving the proof, we discuss the action of Dehn twists on an immersed closed curve. Let (T, α, β) be a marked torus. An immersed closed curve $\gamma \subset T$ determines a reduced cyclic word, which we also denote by γ, in $\pi_1(T) \cong \langle \alpha, \beta \rangle$, where the generators α and β are closed curves dual to the parametrizing arcs. Let $\ell(\gamma)$ be the length of this word.

Lemma 7.12. *If the intersection number $i(\beta, \gamma) \neq 0$ then $\ell(\tau_\beta^n(\gamma)) \to \infty$ as $n \to \infty$.*

Proof. Intersections of γ with the closed curve β are in bijection with cyclic subwords of γ of the form $\alpha\beta^k\alpha$ and $\alpha^{-1}\beta^k\alpha^{-1}$ for $k \in \mathbb{Z}$. The action of τ_β^n replaces such subwords with $\alpha\beta^{k+n}\alpha$ or $\alpha^{-1}\beta^{k-n}\alpha^{-1}$, respectively. The resulting word is still cyclically reduced. As long is there as at least one intersection with β, its length goes to ∞ as n does, since the value of k in each subword is fixed. $\square$

Proof of Proposition 7.11. Let $\tau_\lambda \colon \partial M \to \partial M$ be the Dehn twist along λ. Then $\tau_\lambda(\mu + \lambda) = \mu$ and $\tau_\lambda(\lambda) = \lambda$, so M is a Heegaard Floer solid torus if and only if $\widehat{CFD}(M, \mu, \lambda) \cong \widehat{T}_\lambda \boxtimes \widehat{CFD}(M, \mu, \lambda)$, where $\widehat{T}_\lambda$ is the Dehn twist bimodule considered in Section 5. By the results of that section, this is equivalent to saying that $\widehat{HF}(M)$ is invariant under the action of τ_λ, *i.e.* that up to regular homotopy $\tau_\lambda(\widehat{HF}(M))$ represents the same immersed multicurve as $\widehat{HF}(M)$.

The hypothesis of the proposition is equivalent the statement that $\widehat{HF}(M)$ is homotopic to a curve that is disjoint from the closed curve representing λ. If this is the case, $\widehat{HF}(M)$ is clearly preserved by the action of τ_λ.

Conversely, suppose $\widehat{HF}(M)$ has a component γ for which $i(\gamma, \lambda) \neq 0$. Then by Lemma 7.12 $\ell(\tau_\lambda^n(\gamma)) \to \infty$ as $n \to \infty$ and so, for sufficiently large n, $\tau_\lambda^n(\gamma)$ cannot be a component of $\widehat{HF}(M)$. It follows that $\widehat{HF}(M)$ is not fixed by the action of τ_λ, and M cannot be a Heegaard Floer solid torus. $\qquad\square$

If $\widehat{HF}(M)$ consists of a single loose curve with trivial local system in each Spin^c structure, we say M is solid torus like. Gillespie studied such manifolds and showed that they are boundary compressible [15, Corollary 2.9]. More generally, we say $\widehat{HF}(M)$ is loose if each component of $\widehat{HF}(M)$ is a loose curve (possibly equipped with a nontrivial local system). Since $\widehat{HF}(M)$ lifts to $\overline{T}_M$, all such components must be parallel. We will prove the following generalization of Gillespie's result.

Proposition 7.13. *If $\widehat{HF}(M)$ is loose then M is boundary compressible.*

Before giving the proof, we recall some facts about knot Floer homology and the Thurston norm. Let λ be the rational longitude of M, and suppose that λ has order k in $H_1(M)$. Choose a meridian $\mu \in H_1(\partial M)$ with $\mu \cdot \lambda = 1$. Then μ is k times a primitive element of $H_1(M)$.

Fix a class $x \in H_2(M, \partial M)$ with $\partial x = \lambda$ and define
$$\widehat{HFK}(K_\mu, x, i) = \bigoplus_{\langle c_1(\mathfrak{t}), x \rangle = 2i} \widehat{HFK}(K_\mu, \mathfrak{t})$$
where the sum runs over $\mathfrak{t} \in \mathrm{Spin}^c(M, \partial M)$. Then
$$2 \cdot \max\{i \mid \widehat{HFK}(K, x, i) \neq 0\} = \|x\|_\mu$$
Here $\| \cdot \|_\mu$ is the generalized Thurston norm given by
$$\|x\|_\mu = \min_{[S] = x} c_\mu(S),$$
where the minimum is taken over all properly embedded orientable surfaces representing x. If S is connected, $c_\mu(S) = \max\{-\chi(S) + |\partial S \cdot \mu|, 0\}$; otherwise $c_\mu(S) = \sum c_\mu(S_i)$, where S_i are the connected components of S.

As we described in Section 6.4, the Spin^c decomposition of $\widehat{HFK}(K, \mu)$ can be computed by intersecting $\widehat{HF}(M, \mathfrak{s})$ with lifts of μ to $\overline{T}_{M, \mathfrak{s}}$, where $\mathfrak{s}$ runs over $\mathrm{Spin}^c(M)$. The choice of x determines a height function on $\overline{T}_{M, \mathfrak{s}}$ such that intersections of $\widehat{HF}(M, \mathfrak{s})$ with lifts of the meridian at height i correspond to elements of $\widehat{HFK}(K_\mu, x, i)$.

Similarly, if Y is a closed manifold, and $y \in H_2(Y)$ is a primitive class, we define
$$\widehat{HF}(Y, y, i) = \bigoplus_{\langle c_1(\mathfrak{s}), [\Sigma] \rangle = 2i} \widehat{HF}(Y, \mathfrak{s}).$$
Then $\|y\| = 2 \cdot \max\{i \mid \widehat{HF}(Y, y, i) \neq 0\}$.

Proof of Proposition 7.13. Fix a class x as above, and choose a norm-minimizing surface S realizing x. After tubing to remove excess boundary components, we may assume that ∂S consists of precisely k parallel copies of λ. Since λ has order k in $H_1(M)$, all k boundary components lie on a single component of S. Let S_0 be the component with boundary, and consider the modified complexity $\widetilde{c}(S) = -\chi(S_0) + \sum_{i>0} c(S_i)$, where the sum runs over the other components of S. Note that $\widetilde{c}(S) = c(S)$ unless $k = 1$ and $S_0 = D^2$.

By hypothesis, $\widehat{HF}(M)$ is a union of loose curves in T_M. Each individual curve lifts to some $\overline{T}_{M, \mathfrak{s}}$ as a circle parallel to λ. This circle is at some height $n \in \mathbb{Z}$ with respect to the height function determined by x. Let n_+ and n_- be the heights of the highest and lowest circles in $\widehat{HF}(M)$. Then n_+ is the largest value of i for which $\widehat{HFK}(K_\mu, x, i)$ is nontrivial, and n_- is the lowest. Since knot

Floer homology is conjugation-symmetric, we must have $n_- = -n_+$. The relation between $\widehat{HFK}$ and the Thurston norm implies that $\widetilde{c}(S) = 2n_+ - k$.

Now consider the closed manifold $Y = M \cup_h M$, where $h(\lambda) = -\lambda$ and $h(\mu) = \mu$, and let $y \in H_2(Y)$ be the class obtained by doubling x. Then $\widehat{HF}(Y, y, i)$ is calculated by pairing $\widehat{HF}(M, \mathfrak{s})$ with $\alpha \cdot \overline{h}(\widehat{HF}(M, \mathfrak{s}'))$, where $\mathfrak{s}, \mathfrak{s}'$ run over $\mathrm{Spin}^c(M)$, and α runs over all deck transformations which shift height by i. The largest value of i for which $\widehat{HF}(Y, y, i)$ is potentially nontrivial is obtained by pairing the highest set of curves in $\widehat{HF}(M)$ with the lowest set of curves in $\overline{h}(\widehat{HF}(M))$. The latter is simply the image of the highest set of curves in $\widehat{HF}(M)$ under $\overline{h}$, so the maximal value of i is $n_+ - n_- = 2n_+$. We can view the highest set of curves as a single local system (V, A) on S^1, so

$$\widehat{HF}(Y, y, 2n_+) = HF(S^1_{(V,A)}, S^1_{(V,A)}).$$

The latter group is nonzero by Corollary 4.10. Since $\widehat{HF}(Y, y, 2n_+) \neq 0$, we see that $\|y\| = 4n_+$.

On the other hand, the double DS of S represents y and has complexity $2c(S)$. If M is boundary incompressible, then $S_0 \neq D^2$, so $c(S) = \widetilde{c}(S) = 2n_+ - k$. In this case, we would have $c(DS) = 4n_+ - 2k < \|y\|$, which is impossible. We conclude that $S_0 = D^2$ and M was boundary compressible. $\qquad\square$

Remark 7.14. The proof shows that if $\widehat{HF}(M)$ is loose then it consists of curves that are all at a fixed height (up to overall shift, we may assume that these are all at height 0). Loose manifolds with curves at other heights do not exist. One possible explanation for this would be that $\widehat{HF}(M, \mathfrak{s})$ should represent an exact Lagrangian in $\overline{T}_M$.

7.3. **A dimension inequality for pinches.** Consider the following construction: Let Y be a rational homology sphere of the form $M_0 \cup_h M_1$, so that each M_i is a rational homology solid torus. Denote by λ_i the rational longitude in each ∂M_i; recall that this slope is characterized by the property that some number of like oriented copies of λ_i bounds a properly embedded surface in M_i. The Dehn surgery $Y_0 = M_0(h(\lambda_1))$ is the result of pinching M_1 in Y. Note that $h(\lambda_1)$ is a well defined slope in ∂M_0. In certain settings there is an associated degree n map $Y \to Y_0$, where n is the order of $i_*([\lambda_1])$ in $H_1(M_1)$. In particular, when both M_0 and M_1 are integer homology soild tori, there is always a degree one map associated with a pinch.

Theorem 7.15. *Given a rational homology sphere $Y = M_0 \cup_h M_1$, where M_0 and M_1 are rational homology solid tori, there is an inequality*

$$\dim \widehat{HF}(Y) \geq nm \dim \widehat{HF}(Y_0)$$

where $Y_0 = M_0(h(\lambda_1))$, n is the order of $i_*([\lambda_1])$ in $H_1(M_1)$, and $m = |\mathrm{Spin}^c(M_1)| = |H_1(M_1, \partial M_1)|$.

The proof of Theorem 1.12 follows from a special case of Theorem 7.15 (with $n = m = 1$).

Proof of Theorem 7.15. Let $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$ denote the decorated curves $\widehat{HF}(M_0)$ and $\overline{h}(\widehat{HF}(M_1))$, respectively, in $T_{M_0} = \partial M_0 \setminus z$. By the gluing theorem, $\widehat{HF}(Y)$ is equivalent to the Floer homology $HF(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1)$ and in particular the dimension is at least the intersection number $i(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1)$ between $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$. Let $\frac{p}{q}$ be the slope of the curve in T_{M_0} determined by $h(\lambda_1)$, and let $L_{p,q}$ be the corresponding simple closed curve $p\alpha_0 + q\beta_0$. Since Y is a rational homology sphere we may assume that the slope of λ_0 is not $\frac{p}{q}$; in particular, no component of $\boldsymbol{\gamma}_0$ is commensurable to $L_{p,q}$, and by the gluing theorem $\dim \widehat{HF}(Y_0) = i(\boldsymbol{\gamma}_0, L_{p,q})$.

Let $\hat{\boldsymbol{\gamma}}_1$ be the collection of curves obtained from $\boldsymbol{\gamma}_1$ by replacing any non-trivial local systems with trivial ones of the same dimension and pulling the curves tight in the torus ∂M_0 (ignoring the basepoint). More precisely, after replacing non-trivial local systems, we delete any curve components that are nullhomotopic in the (non-punctured) torus ∂M_0, and we replace the remaining curves with minimal length representatives of their homotopy class in the flat torus. The resulting collection

of curves is loose and represents the same homology class in ∂M_0 as $\boldsymbol{\gamma}_1$; we call $\hat{\boldsymbol{\gamma}}_1$ the *loose representative* of $\boldsymbol{\gamma}_1$. Recall that for each spin^c-structure $\mathfrak{s}$ of M_1, the multicurve $\widehat{HF}(M_1, \mathfrak{s})$ is homologous, ignoring the basepoint, to the homological longitude $n\lambda_1$; see Corollary 6.6. Thus, for each spin^c-structure of M_1, the relevant components of $\boldsymbol{\gamma}_1$ are homologous in ∂M_0 to n copies of $L_{p,q}$. It follows that $\hat{\boldsymbol{\gamma}}_1$ contains at least nm copies of $L_{p,q}$, so that

$$i(\boldsymbol{\gamma}_0, \hat{\boldsymbol{\gamma}}_1) \geq nm \cdot i(\boldsymbol{\gamma}_0, L_{p,q}) = nm \dim \widehat{HF}(Y_0).$$

It remains to show that $i(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) \geq i(\boldsymbol{\gamma}_0, \hat{\boldsymbol{\gamma}}_1)$.

We will fix a small ϵ and work in an ϵ-pegboard diagram. Note that $\hat{\boldsymbol{\gamma}}_1$ can be obtained from $\boldsymbol{\gamma}_1$ by a sequence of the following moves: (i) replacing a local system with a trivial one of the same dimension; (ii) resolving a self-intersection to split off a closed component; (iii) deleting a component, (iv) homotopy in T_{M_0} to put curves in ϵ-pegboard position; and (v) passing the curve through a peg near a corner in the peg board diagram, as in Figure 55. Moves (i) and (ii) clearly do not change the number of intersections with $\boldsymbol{\gamma}_0$ at all, and move (iii) cannot increase the intersection number. Move (iv) also does not increase intersection number, since curves in an an ϵ-pegboard diagram realize the minimal intersection for their homotopy

Figure 54. Sliding a corner over the peg. The corner wraps around a circle of some radius between ϵ and 2ϵ (the inner and outer radii of the shaded annulus). We replace the segment of curve through the annulus with a segment on the opposite side of the peg following the circle $C_{2\epsilon}$.

classes in T. The remaining move, (v), requires some care, since it may introduce new intersection points; see Figure 56.

After removing nullhomologous components, no component of $\boldsymbol{\gamma}_1$ is homotopic to a component of $\boldsymbol{\gamma}_0$; it follows that $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$ intersect minimally and transversely in the ϵ-pegboard diagram. By first applying (ii) as needed, we may assume that at each corner of $\boldsymbol{\gamma}_1$ the curve changes direction by at most an angle of π; that is, there is no peg-wrapping in $\boldsymbol{\gamma}_1$. Now fix a corner c_1 of $\boldsymbol{\gamma}_1$ and consider the effect of applying (v) at this corner; let $\hat{c}_1$ denote the modified corner segment along the circle $C_{2\epsilon}$. Since $\boldsymbol{\gamma}_1$ is unchanged outside the disk of radius 2ϵ, to compute the new intersection number we only need to consider the intersection of $\hat{c}_1$ with the corners of $\boldsymbol{\gamma}_0$. For each corner of $\boldsymbol{\gamma}_0$, replacing c_1 with $\hat{c}_1$ either adds two intersection points, preserves intersection number, or removes two intersection points (see Figure 56). Replacing c_1 with $\hat{c}_1$ will not increase the overall intersection number as long as $\boldsymbol{\gamma}_0$ has at least as many type (d) corners as type (a) corners.

| (a) | (b) | (c) | (d) |
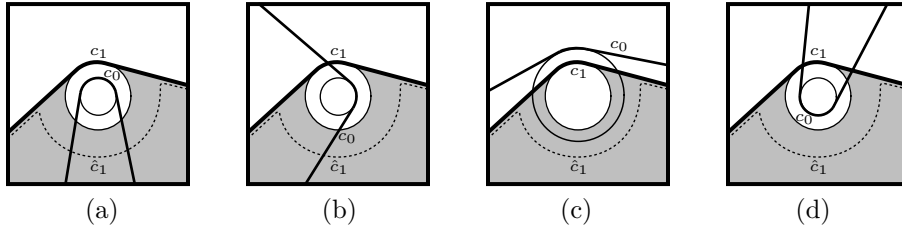
Figure 55. For a fixed corner c_1 of $\boldsymbol{\gamma}_1$, a corner c_0 of $\boldsymbol{\gamma}_0$ has one of four types. The inside of the corner c_1 is shaded, and the dashed line indicates the modified corner $\hat{c}_1$ resulting from move (v).

To analyze this situation, we decompose T_{M_0} into a square tile (obtained by taking the closure of $T_{M_0} \smallsetminus (\alpha_0 \cup \beta_0)$), and consider the component of $\boldsymbol{\gamma}_1$ containing the chosen corner c_1 as in Figure 57. This segment of curve, which we call s, begins on the boundary of the square tile at at s_{in}, wraps some amount around a circle with radius between ϵ and 2ϵ centered at the puncture, then exits the square tile at s_{out}. Note that s divides the square tile into two sides; we call the side that contains the peg the *inside* of the corner c_1. Similarly, any corner c_0 of $\boldsymbol{\gamma}_0$ can be extended to a segment through the square tile; we will call this segment x, with endpoints x_{in} and x_{out} on the boundary of the tile. We can now more precisely classify corners of $\boldsymbol{\gamma}_0$ into the four types shown in
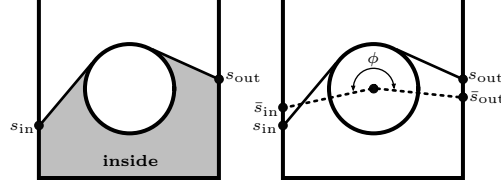
Figure 56. A square tile representing the torus T_{M_0} together with the section s of γ_1 containing a chosen corner c_1. The inside of s is determined by containment of the peg, as shown on the left; on the right, the corresponding segment $\bar{s}$ in a singular pegboard diagram. Note that s lies in a neighborhood of width 4ϵ around $\bar{s}$. We are most interested in the case that the angle ϕ associated with c_1 in the singular pegboard diagram is strictly bigger than π (that is, c_1 is not a π-corner), so that $\bar{s}$ is not a straight line. In this case, the inside of s covers strictly less than half the boundary of the square.

Figure 56 by considering how the segment x interacts with s. We say the corner c_0 is of type (a) if both endpoints lie on the inside of c_1, type (b) if exactly one endpoint lies inside c_1, type (c) if both endpoints lie outside c_1 and c_0 turns the same direction as c_1 with no peg-wrapping, and type (d) if both endpoints lie outside c_1 and c_0 wraps the peg or turns the opposite direction from c_1. Note that for type (a) corners, c_0 necessarily changes direction more than c_1 and so wraps closer to the puncture, giving the nested configuration in the figure. In this case c_0 is disjoint from c_1 but intersects the modified corner $\hat{c}_1$ twice. For type (b) corners, c_0 may turn either more or less tightly than c_1; in either case, c_0 intersects both c_1 and $\hat{c}_1$ once. For type (c) corners, c_1 must be nested inside of c_0, so c_0 is disjoint from both c_1 and $\hat{c}_1$. Finally, type (d) corners have two intersections with c_1 but are disjoint from $\hat{c}_1$. Observe that replacing c_1 with $\hat{c}_1$ will not increase the overall intersection number as long as γ_0 has at least as many type (d) corners as type (a) corners.
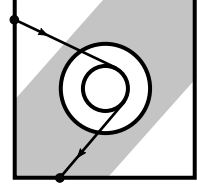
We now need to consider two cases with different behavior, depending on whether the curve changes direction at the corner c_1 in a singular pegboard diagram. Let $\bar{s}$ denote the piecewise linear segment of curve through the square tile and containing the corner c_1 in a singular peg-board diagram for γ_1; in other words, $\bar{s}$ is the limit of s if ϵ is taken to zero. We denote the endpoints of $\bar{s}$ by $\bar{s}_{\text{in}}$ and $\bar{s}_{\text{out}}$. We first consider the case that the singular peg-board representative of γ_1 changes direction at c_1; that is, the segment $\bar{s}$ is not a straight line. In this case $\bar{s}_{\text{in}}$ and $\bar{s}_{\text{out}}$ cut the boundary of the square tile into two unequal pieces, with the smaller piece corresponding to the inside of c_1. Note that s_{in} and s_{out} are close to $\bar{s}_{\text{in}}$ and $\bar{s}_{\text{out}}$, respectively,

Figure 57. Labelling endpoints for corners of γ_0.

where close means within a small multiple of ϵ; in particular, since s lies in the 2ϵ-neighborhood of $\bar{s}$ and since $\bar{s}$ meets the boundary of the square in an angle of at least $\frac{\pi}{4}$, the distances are bounded by $2\sqrt{2}\epsilon$. Since we can take ϵ to be arbitrarily small, it follows that the inside of c_1 covers strictly less than half the boundary of the square tile. We now run through all the corners $\{c_0^i\}$ of γ_0, labeling the endpoints of the corresponding curve segments in the square tile by x_{in}^i and x_{out}^i, as in Figure 58 (the index i should be interpreted cyclically, with the corners ordered as they appear traversing the curve). Note that the curve γ_0 may pass through the square tile (without interacting with the peg) many times between x_{out}^i and x_{in}^{i+1}; we ignore these portions of the curve. Since γ_0 does not change direction between corners, the slope of the segment leaving the tile at x_{out}^i agrees with the slope of the segment entering the tile at x_{in}^{i+1}. It follows that x_{in}^{i+1} is close to the antipodal point to x_{out}^i on the boundary of the square (i.e. the point obtained by a half-rotation about the center of the square), where again close means within a small fixed multiple of ϵ. More precisely, if we consider the analogous points $\bar{x}_{\text{out}}^i$ and $\bar{x}_{\text{in}}^{i+1}$ coming from a singular pegboard representative of γ_1, we observe that these are antipodal points since the segments connecting them to the center of the square have the same slope. Because γ_0 lies in a 2ϵ-neighborhood of the singular peg-board representative, x_{out}^i is within $2\sqrt{2}\epsilon$ of $\bar{x}_{\text{out}}^i$, x_{in}^{i+1} is within $2\sqrt{2}\epsilon$ of $\bar{x}_{\text{in}}^{i+1}$, and x_{in}^{i+1} is a distance of less than $4\sqrt{2}\epsilon$ from the antipodal point to x_{out}^i. For sufficiently small ϵ, it follows that if x_{out}^i is on the inside of c_1 then x_{in}^{i+1} must be on the outside of c_1 and not close to s_{in} or s_{out} in the aforementioned
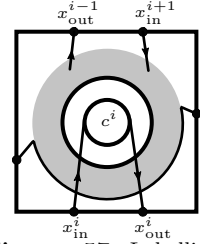
sense. We also observe that if c_0^i is a type (c) corner and x_{in}^i is not close to s_{in} or s_{out}, then x_{in}^{i+1} is on the outside of c_1 and not close to s_{in} or s_{out}. This follows from the fact that c_1 lies entirely inside c_0 and the antipodal point to x_{out}^i lies either outside c_0 or close to x_{in}^i. In particular, if c_0^i is type (a), then x_{in}^{i+1} must lie outside of c_1 and c_0^{i+1} must be type (b), (c), or (d). Moreover, if it is type (b) or (c) then x_{in}^{i+2} lies outside c_1, and we can repeat the argument with c_0^{i+2}; we must reach a type (d) corner before reaching another type (a) corner. Thus there are at least as many type (d) corners as type (a) corners in $\boldsymbol{\gamma}_0$, and applying move (v) at the corner c_1 does not increase the total intersection number.

It remains to deal with the case that the corner c_1 is straight in a singular pegboard diagram. The argument above breaks down in this case, since the inside of the corner may cover more than half of the boundary of the square tile, and indeed applying move (v) to a single corner that is straight in the singular diagram may increase the total intersection number. By applying (v) at all corners that are not straight in the singular diagram, along with moves (i)-(iv), we can reduce to the case that all corners of $\boldsymbol{\gamma}_1$ are straight; that is, $\boldsymbol{\gamma}_1$ is a Heegaard Floer solid torus curve. We now argue directly that for such a curve the loose representative $\hat{\boldsymbol{\gamma}}_1$ realizes the minimal intersection with $\boldsymbol{\gamma}_0$. To see this consider an annular neighbourhood of $\boldsymbol{\gamma}_1$ containing the radius 2ϵ circle around the peg; see Figure 59. Now decomposing $\boldsymbol{\gamma}_0$ into

Figure 58. Segments of $\boldsymbol{\gamma}_0$ interact with a neighbourhood of $\boldsymbol{\gamma}_1$ in two distinct ways.

segments connecting the corners $\{c_0^i\}$ shows two distinct behaviours. We may ignore segments of $\boldsymbol{\gamma}_0$ that are contained in the neighbourhood of $\boldsymbol{\gamma}_1$, since the loose representative $\hat{\boldsymbol{\gamma}}_1$ does not intersect these segments of $\boldsymbol{\gamma}_0$. We consider then only the intersections with those segments that leave the neighbourhood of $\boldsymbol{\gamma}_1$, and observe that the original curve $\boldsymbol{\gamma}_1$ and the loose representative $\hat{\boldsymbol{\gamma}}_1$ have the same number of intersections with these segments. As a result, we can replace $\boldsymbol{\gamma}_1$ with $\hat{\boldsymbol{\gamma}}_1$ without increasing the number of intersection points. $\qquad\square$

7.4. **Heegaard Floer homology for toroidal manifolds.** As another immediate consequence of the geometric interpretation of bordered Floer invariants, we will now prove Theorem 1.8, which states that if Y contains an essential torus that is separating then $\dim \widehat{HF}(Y) \geq 5$.

Proof of Theorem 1.8. By cutting along the torus, realize Y as $M_0 \cup_h M_1$ where M_0 and M_1 are manifolds with incompressible torus boundary and $h\colon \partial M_1 \to \partial M_0$ is a diffeomorphism. Let $\boldsymbol{\gamma_0} = \widehat{HF}(M_0)$ and $\boldsymbol{\gamma_1} = \bar{h}(\widehat{HF}(M_1))$ be the relevant decorated immersed multicurves in $\partial M_0 \setminus z_0$.

We first set some homological prerequisites in place. A manifold with torus boundary M has rational longitude λ and homological longitude $p\lambda$. In particular, $p\lambda$ bounds a properly embedded surface in M; and $\langle\lambda\rangle$ generates a $\mathbb{Z}/p\mathbb{Z}$ subgroup in $H_1(M;\mathbb{Z})$. Applying universal coefficients and Poincaré duality then gives a $\mathbb{Z}/p\mathbb{Z}$ subgroup in $H^2(M;\mathbb{Z})$, which in turn guarantees p torsion spin^c-structures on M that we denote $\{\mathfrak{s}_i\}_{i=0}^{p-1}$. Finally, observe that for any choice of meridian μ, that is, a slope at distance one from λ, the Dehn filling $M(\mu)$ inherits $\mathbb{Z}/p\mathbb{Z} \subset H^2(M(\mu);\mathbb{Z})$; by abuse of notation, we denote the associated p torsion spin^c-structures on $M(\mu)$ by $\{\mathfrak{s}_i\}_{i=0}^{p-1}$ as well. With this homological data in hand, we turn to Heegaard Floer homology. In particular, note that $\widehat{HF}(M(\mu),\mathfrak{s}_i)$ is non-trivial for each $i = 0, \ldots, p - 1$. This appears to be known to experts, and appeals to work of Lidman [36]; an argument, due to the second author, is written down in [32, Proposition 8.9]. In particular, by applying our pairing theorem, this non-vanishing result guarantees that there is at least one curve in $\widehat{HF}(M,\mathfrak{s}_i)$ for each $i = 0, \ldots, p - 1$. Moreover, if we forget the punctures, this curve is homologous to $p\lambda$; compare Corollary 6.6.

As in the proof of Theorem 7.15, there are two somewhat distinct cases to consider, with Heegaard Floer solid torus type curves requiring a different approach. Recall that a corner of $\boldsymbol{\gamma_i}$ is a π-corner if the associated angle in a singular pegboard diagram is exactly π, so that the corner pulls tight to a straight line, and that we call all other corners true corners. If $\boldsymbol{\gamma_i}$ has no true corners then M_i is

a Heegaard Floer solid torus. Note that if $\boldsymbol{\gamma_i}$ has one true corner, then it must have at least two true corners. This is because if $\boldsymbol{\gamma_i}$, a closed curve, changes direction at one corner it must change direction again to get back to its initial slope. The one potential exception is if $\boldsymbol{\gamma_i}$ contains a single true corner wrapping the peg an integer number of full wraps so that the curve does not change direction; however, it follows from Corollary 7.19 in the next section that such curves do not arise.

A key observation is that, if $\boldsymbol{\gamma_0}$ is not a Heegaard Floer solid torus curve, the minimal intersection number between $\boldsymbol{\gamma_0}$ and $\boldsymbol{\gamma_1}$ satisfies

$$(5) \qquad i(\boldsymbol{\gamma_0}, \boldsymbol{\gamma_1}) \geq 2\#\{\text{true corners of } \boldsymbol{\gamma_1}\} + \#\{\pi\text{-corners of } \boldsymbol{\gamma_1}\} + \#\{\text{loose components of } \boldsymbol{\gamma_1}\}.$$

To show this, we first consider a fixed true corner c_1 of $\boldsymbol{\gamma_1}$ and argue that the combined intersection number of c_1 with all the corners $\{c_0^i\}$ of $\boldsymbol{\gamma_0}$ is at least two. As in the proof of Theorem 7.15, we assign the corners of $\boldsymbol{\gamma_0}$ types as in Figure 56, and we extend each corner c_0^i to a segment x^i through the square tile with endpoints x_{in}^i and x_{out}^i. If $\boldsymbol{\gamma_0}$ has a type (d) corner, we are done since this corner alone contributes two intersection points with c_1. If $\boldsymbol{\gamma_0}$ has a type (a) corner, then the argument in the proof of Theorem 7.15 shows there must also be a type (d) corner. By assumption $\boldsymbol{\gamma_0}$ has at least one true corner. If $\boldsymbol{\gamma_0}$ has a true corner of type (b), take this to be c_0^i and orient the curve so that x_{out}^i is on the inside of c_1. We then have that x_{in}^{i+1}, being within a small neighborhood of the antipodal point of x_{out}^i, lies outside of c_1 so that the next corner c_0^{i+1} must have type (b), (c), or (d). If c_0^{i+1} is type (b) or (d) we are done, since this contributes at least one more intersection with c_1 in addition to the one intersection from c_0^i. The only way that c_0^{i+1} can have type (c) is if the segment from c_0^i to x_{out}^i coincides with one of the segments defining c_1 when we pass to a singular peg-board diagram and c_0^{i+1} is a π-corner. In this case, the same reasoning shows that c_0^{i+1} is type (b), (c), or (d) and can only be type (c) under the same conditions. Since we must eventually reach a true corner, we eventually reach a type (b) or (d) corner, contributing at least the second intersection with c_1 (note that if we reach a type (b) corner, it can not be the corner c_0^i we started with since the incoming segment has the same slope as the outgoing segment of c_0^i, and c_0^i was not a π-corner). Finally, if $\boldsymbol{\gamma_0}$ has a true corner of type (c) then there is a straight line segment through the puncture in the square tile such that both endpoints lie strictly outside of this corner. Up to reparametrization we may assume this line is horizontal, so that the corner represents a local maximum of $\boldsymbol{\gamma_0}$. Being a closed curve, $\boldsymbol{\gamma_0}$ must also have a local minimum, and a corner at which a local minimum occurs is necessarily type (d). Thus the true corner c_1 contributes at least two intersections to $i(\boldsymbol{\gamma_0}, \boldsymbol{\gamma_1})$.

If the fixed corner c_1 of $\boldsymbol{\gamma_1}$ is a π-corner, a similar argument shows the contribution is at least one. Note that any type (b) or (d) corner in $\boldsymbol{\gamma_0}$ contributes at least one intersection point with c_1. Any corner that is type (c) with respect to c_1 must be a π-corner, and by assumption $\boldsymbol{\gamma_0}$ has at least one true corner, so we are left with considering a true type (a) corner c_0^i. In this case, the endpoints x_{in}^i or x_{out}^i lie on the inside of c_1, and at least one of them is strictly on the inside c_1 (in the sense that it is at least a distance of $4\sqrt{2}\epsilon$ from the points s_{in} or s_{out}). Suppose x_{out}^i lies strictly inside of c_1; it follows that x_{in}^{i+1} lies strictly on the outside of c_1 and c_0^{i+1} is type (b) or (d), contributing the needed intersection point.

Finally, we show that that a loose component of $\boldsymbol{\gamma_1}$ contributes at least one intersection. Indeed, if it did not then $\boldsymbol{\gamma_0}$ would have to lie in a neighborhood of a line with the same slope as the loose component, but this would imply that $\boldsymbol{\gamma_0}$ is a Heegaard Floer solid torus. This completes the proof of equation (5).

We first consider the case that neither M_0 nor M_1 is a Heegaard Floer solid torus. Equation (5) then implies that $i(\boldsymbol{\gamma_0}, \boldsymbol{\gamma_1}) \geq 4$. Suppose that $i(\boldsymbol{\gamma_0}, \boldsymbol{\gamma_1}) = 4$. By Equation (5) we must have that $\boldsymbol{\gamma_1}$ contains a single component with exactly two corners, and thus exactly two segments in a singular pegboard diagram. Interchanging the roles of M_0 and M_1 we see that the same is true for $\boldsymbol{\gamma_0}$. It is also clear that neither curve can carry a non-trivial local system, since this would multiply the number of intersection points by the dimension of the local system. Let p_1 and p_2 be the slopes of the two segments in a singular peg-board diagram for γ_0, and let q_1 and q_2 be the slopes of the
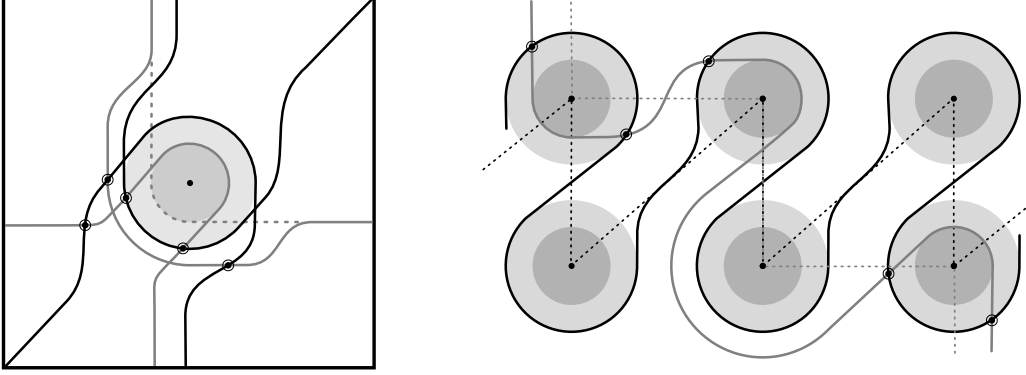
Figure 59. The pairing of γ_0 (gray) consisting of two segments of slopes 0 and ∞ with γ_1 (black) consisting of two segments of slopes 1 and ∞, shown both in the torus and in the covering space $\mathbb{R}^2 \setminus \mathbb{Z}^2$. This corresponds to a gluing of two trefoil complements. The result of the gluing is an integer homology sphere, since there is ± 1 intersection point counted with sign. Note that, for the purposes of illustration, neither curve is pulled tight.

two segments for γ_1. Observe that, since neither curve is a Heegaard Floer solid torus, $p_1 \neq p_2$ and $q_1 \neq q_2$.
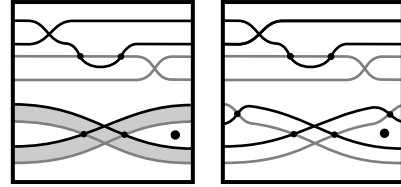
The intersection points counted so far all occur at the corners; as ϵ goes to zero in an ϵ-pegboard diagram, these intersection points approach the puncture. We will now consider intersections away from the corners. We observe that any two segments of slope r_1 and r_2 in the singular pegboard diagram intersect d times, where $d = \Delta(r_1, r_2)$ is the distance between the slopes (calculated by taking minimal geometric intersection number). One of these intersection points is at the endpoints, but the remaining $d - 1$ intersection points occur in the interior of the segments; these intersections remain in an ϵ-pegboard diagram and are a finite distance from the peg. It follows that if $i(\gamma_0, \gamma_1) = 4$ then $\Delta(p_1, q_1)$, $\Delta(p_1, q_2)$, $\Delta(p_2, q_1)$, and $\Delta(p_2, q_2)$ are all at most 1. First suppose that the four slopes p_1, p_2, q_1, and q_2 are all distinct. Up to reparametrization, we may assume $p_1 = \infty$; it follows that q_1 and q_2 are distinct integers. Since p_2 is distance 1 from each of these and distinct from ∞, we must have $p_2 = n$, $q_1 = n - 1$ and $q_2 = n + 1$. This case is ruled out, because it would imply that the rational longitude λ_1 of M_1 has order two, but as above this implies the existence of another component and gives $i(\gamma_0, \gamma_1) > 4$. Now instead suppose both curves have a segment of the same slope; up to reparametrization and reindexing, we may assume $p_1 = q_1 = \infty$. It follows that p_2 and q_2 are integers, with $|p_2 - q_2| \leq 1$. Up to reparametrization and switching the roles of γ_0 and γ_1, we may assume that $p_2 = 0$ and q_2 is either 0 or 1. In the first case, the curves γ_0 and γ_1 coincide. This is the one setting in which an ϵ-pegboard diagram does not give transverse intersection but, perturbing slightly from an ϵ-pegboard diagram, it is a simple exercise to check that the minimal intersection $i(\gamma_0, \gamma_1)$ is 4. However, in this minimal position there is an immersed annulus, so achieving admissibility requires adding two additional intersection points and $\dim HF(\gamma_0, \gamma_1) = 6$. The remaining case is depicted in Figure 60 and has 5 intersection points.

We next consider the case that M_1 is a Heegaard Floer solid torus but M_0 is not. If γ_1 has true corners, the argument above applies so that $i(\gamma_0, \gamma_1) > 4$ unless γ_1 has a single component and exactly two corners. It is easy to see that such a curve cannot be the invariant of a Heegaard Floer solid torus. If both segments in a singular peg-board diagram have the same slope then either the curve is nullhomologous or it is homologous to twice a primitive curve; in the first case there must be another component since γ_1 is homologous to the rational longitude λ_1 of M_1, and in the second case there must be another component because if λ_1 has order two there are at least two spin^c structures of M_1. We conclude that γ_1 has only π-corners. It is still the case that each π-corner of γ_1 and each loose component of γ_1 contributes one to $i(\gamma_0, \gamma_1)$; in fact, a loose component that is p times a primitive curve contributes at least p intersection points. If p is the order of the rational longitude of M_1, then γ_1 has at least p components which each contribute at least p intersection

points, so $i(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) \geq p^2$. If $i(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) < 5$ then $p = 2$ and $\boldsymbol{\gamma}_1$ has exactly two components, each of which is one of the two curves associated with the twisted I-bundle (see Figure 54). Since M_1 has incompressible boundary, at least one of the components of $\boldsymbol{\gamma}_1$ is tight. It is now a simple exercise that $i(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) \geq 6$. The lower bound of four given previously records two intersections with a loose component and two intersections near corners in a tight component. However, if $\boldsymbol{\gamma}_0$ has two segments with slope different from the slope of λ_1 then there will be two type (d) corners, and if one segment of $\boldsymbol{\gamma}_0$ has the same slope as λ_1 this segment will have two additional intersection points, so the tight component of $\boldsymbol{\gamma}_0$ contributes at least four intersections.

The final case to consider is when both M_0 and M_1 are Heegaard Floer solid tori; that is, each multicurve pulls tight to a straight line in a singular pegboard diagram. These curve-sets may be divided into sets of loose and tight components. By irreducibility and Proposition 7.13, both $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$ must have at least one tight component. We claim that each multicurve must also have at least one loose component. In fact, each must have a loose component for which the associated local system has 1 as an eigenvalue (note that when discussing local systems on loose components we will assume the underlying curve is primitive by applying Lemma 4.8 if needed). To see this, consider pairing with L_i, a simple closed curve in ∂M_i representing the rational longitude λ_i of M_i. The loose curve L_i pairs trivially with any tight component of $\boldsymbol{\gamma}_i$, since the curves can be made disjoint without any immersed annuli. Any loose component of $\boldsymbol{\gamma}_i$ decorated with a local system A pairs with L_i to give Floer homology of dimension $2\dim(\ker(I + A))$; if $I + A$ is invertible (equivalently if 1 is not an eigenvalue of A) then this contribution is again trivial. Since the Floer homology $HF(\boldsymbol{\gamma}_i, L_i)$ computes $\widehat{HF}$ of the Dehn filling $M_i(\lambda_i)$, which must be nontrivial, there must be at least one component of $\boldsymbol{\gamma}_i$ which is loose and with 1 as an eigenvalue of the local system.

Let p_i be the order of the rational longitude λ_i of M_i. Recall that $\boldsymbol{\gamma}_i$ has at least p_i components, each homologous to at least p_i copies of λ_i. From this it easily follows that if Y is a rational homology sphere, $i(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) \geq p_0^2 p_1^2$. Moreover, if a tight component of $\boldsymbol{\gamma}_i$ has only π-corners then p_i must be at least two, and if this holds for both curves then $i(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) \geq 16$ for a rational homology sphere gluing. There are a few pathological examples to consider (which likely do not arise from manifolds) of tight Heegaard Floer solid torus curves with $p_i = 1$, but such curves traverse the underlying primitive curve at least three times and still give a lower bound of 9 for rational homology sphere

Figure 60. Pairing two twisted I-bundles over the Klein bottle: At left, the loose components contribute two intersections and the tight components contribute two intersections. Imposing admissibility to the tight components as on the right, brings the intersection number to 6.

gluings. The more interesting case is when $h(\lambda_1) = \lambda_0$; we will reparametrize so that both curves are horizontal. In this setting, any pair of tight components contribute at least two intersection points to Floer homology; to see this note that each curve must pass above the peg at least once and below the peg at least once, and a (rightward moving) segment moving from a corner below the peg to a corner above the peg always intersects a segment moving from a corner above the peg to a corner below the peg. Any pair of loose curves with local systems A_0 and A_1 for which 1 as an eigenvalue also have Floer homology of dimension at least two. Although the minimal intersection number is zero, by Corollary 4.10 the dimension of Floer homology is given by $2\ker(I + A_0 \otimes A_1)$; since 1 is an eigenvalue of both A_0 and A_1, it is also an eigenvalue of $A_0 \otimes A_1$ and so $I + A_0 \otimes A_1$ is singular. To summarize, if x_i is the number of tight components of $\boldsymbol{\gamma}_i$ and y_i is the number of loose components of $\boldsymbol{\gamma}_i$ with 1 as an eigenvalue of the local system, then $\dim HF(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1)$ is bounded below by $2x_0 x_1 + 2y_0 y_1$. The discussion above ensures that x_i and y_i are both nonzero for $i = 0, 1$, and so this bound is at least 4. Moreover, the bound is only 4 if $x_0 = x_1 = y_0 = y_1 = 1$. In this case, both rational longitudes must have order two (since each multicurve has only two components), and there is a unique tight curve which wraps twice around the torus following the horizontal line. Since the tight component of $\boldsymbol{\gamma}_0$ is commensurate with the tight component of $\boldsymbol{\gamma}_1$, one can check that pairing these curves in fact contributes at least 4 to $\dim HF(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1)$, raising the lower bound in this case to

6. Note that this bound can be realized by letting M_0 and M_1 both be twisted I-bundles over the Klein bottle; this example is shown in Figure 61.   □

The case in Figure 60 is realized by gluing two trefoil complements. For instance, it arises from gluing the complement of two right handed trefoil complements by identifying 0-framed longitude with meridian and meridian with 1-framed longitude. This provides an example of a toroidal integer homology sphere Y realizing $\dim \widehat{HF}(Y) = 5$.

With this example in hand, consider the function f taking values in the natural numbers defined by

$$f(n) = \min\{\dim \widehat{HF}(Y) \,|\, Y \text{ is a rational homology sphere with } n \text{ separating JSJ tori}\}$$

and note that $f(0) = 1$ (realized by the three-sphere or the Poincaré homology sphere) and $f(1) = 5$ (realized by gluing two trefoil exteriors as in Theorem 7.20, see Figure 60). For Heegaard Floer aficionados, we propose:

Question 7.16. Is it possible to determine f or describe properties of f as n grows?

If we further assume that Y is an L-space, we can in fact strengthen the result of Theorem 1.8 by enumerating all examples with $\dim \widehat{HF}(Y) < 7$; see Theorem 7.20. This will make use of a result specific to L-spaces. Toward establishing this result, we now turn our attention to properties of the immersed curve invariants relevant to L-spaces.

7.5. **Characterizing L-space slopes.** Given a manifold with torus boundary M, let $\mathcal{L}_M$ denote the set of L-space slopes, that is, the subset of slopes giving rise to L-spaces on Dehn filling, and let $\mathcal{L}_M^\circ$ denote the interior of $\mathcal{L}_M$. This section gives a characterization of the set $\mathcal{L}_M^\circ$ in terms of the collection of curves $\widehat{HF}(M)$.

Given a spin^c structure $\mathfrak{s}$ on M, we will define a set of slopes $S(M, \mathfrak{s}) \subset \mathbb{R}P^1$ associated with the multicurve $\widehat{HF}(M, \mathfrak{s})$. For each $\epsilon > 0$, we consider an ϵ-pegboard diagram for the curves $\widehat{HF}(M, \mathfrak{s})$. If $\widehat{HF}(M, \mathfrak{s})$ has single component equipped with the trivial local system we define $S_\epsilon(M, \mathfrak{s})$ as the set of tangent slopes of the ϵ-pegboard representative of $\widehat{HF}(M, \mathfrak{s})$. Otherwise, if $\widehat{HF}(M, \mathfrak{s})$ is disconnected or has a non-trivial local system we define $S_\epsilon(M, \mathfrak{s})$ to be all of $\mathbb{R}P^1$. Note that when computing the set of tangent slopes in a ϵ-pegboard diagram, it is convenient to lift to the covering space $\overline{T}_M$. We now define $S(M, \mathfrak{s})$ by taking $\epsilon \to 0$, in the sense that $S(M, \mathfrak{s}) = \bigcap_{n=N}^{\infty} S_{\frac{1}{n}}(M, \mathfrak{s})$. We define $S(M)$ to be the union of $S(M, \mathfrak{s})$ over all spin^c structures.
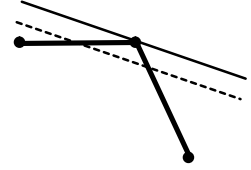
We can define a related set $S^{sing}(M, \mathfrak{s})$ using a singular pegboard diagram. If $\widehat{HF}(M, \mathfrak{s})$ has a single component carrying a trivial local system, $S^{sing}(M, \mathfrak{s})$ is the set of slopes of tangent lines to the singular peg-board representative, where we say a line L through a corner is a tangent line if L coincides with one of the two segments at that corner or if both segments lie on the same side of L. As with $S(M, \mathfrak{s})$, we set $S^{sing}(M, \mathfrak{s}) = \mathbb{R}P^1$ if for some Spin^c structure $\mathfrak{s}$, $\widehat{HF}(M, \mathfrak{s})$ has more than one curve or a non-trivial local system. It is straightforward to see that $S(M) = S^{sing}(M)$ when $\widehat{HF}(M)$ does not wrap any pegs. *A priori*, it seems like the two sets may differ in the presence of peg-wrapping, but we will prove in Corollary 7.19 below that if there is any peg-wrapping, we must have $S^{sing}(M) = S(M) = \mathbb{R}P^1$.

Using tangent slopes in peg-board diagrams, we can give a new interpretation of L-space slopes for a loop-type manifold.

Theorem 7.17. *If M is a manifold with torus boundary then $\mathcal{L}_M^\circ$ is the complement in $\mathbb{Q}P^1$ of the set $S_{\mathbb{Q}}^{sing}(M) = S^{sing}(M) \cap \mathbb{Q}P^1$.*

Proof. By the pairing theorem, $\widehat{HF}(M(\alpha))$ is equivalent to the intersection Floer homology of $\widehat{HF}(M)$ and a straight line with slope α, which we denote L_α. Suppose first that α is a non-L-space slope, so that the Dehn filling $M(\alpha)$ is not an L-space and there is some spin^c-structure $\bar{\mathfrak{s}}$ of $M(\alpha)$ so that $\dim \widehat{HF}(M(\alpha); \bar{\mathfrak{s}}) > 1$. It follows that (for any homotopy representative of $\widehat{HF}(M)$) there are two intersection points x and y between $\widehat{HF}(M)$ and the line L_α with the same spin^c-grading. This means that x and y correspond to intersection points $\tilde{x}$ and $\tilde{y}$ between (a lift of) L_α and $\widehat{HF}(M, \mathfrak{s})$ in the covering space $\mathbb{R}^2 \setminus \mathbb{Z}^2$, for some spin^c-structure $\mathfrak{s}$ of M. If $\widehat{HF}(M, \mathfrak{s})$ contains more than one curve then $S^{sing}(M) = \mathbb{R}P^1$ by definition and $\alpha \in S_{\mathbb{Q}}^{sing}(M)$, so suppose that $\widehat{HF}(M, \mathfrak{s})$ contains a single curve. The intersection points $\tilde{x}$ and $\tilde{y}$ exist for any homotopy representative of $\widehat{HF}(M, \mathfrak{s})$ (in particular the pegboard representatives for arbitrarily small ϵ) and since L_α can be taken to be some finite distance away from every peg, we can in fact take $\tilde{x}$ and $\tilde{y}$ to be intersections of the singular peg-board representative of $\widehat{HF}(M, \mathfrak{s})$ with a lift of L_α. Note that the corners of the singular representative for $\widehat{HF}(M, \mathfrak{s})$ can be smoothed without affecting intersections with L_α and that $S^{sing}(M)$ is precisely the set of tangent slopes after this smoothing. By the (extended) mean value theorem, there is a point on the smoothing of the singular representative of $\widehat{HF}(M, \mathfrak{s})$ for which the tangent line has slope α. We conclude that $\alpha \in S_{\mathbb{Q}}^{sing}(M)$ as claimed. We have proved that $\mathcal{L}_M^c \subset S_{\mathbb{Q}}^{sing}(M)$; in fact, since $S_{\mathbb{Q}}^{sing}(M)$ is closed (as a subset of $\mathbb{Q}P^1$), we have $\overline{\mathcal{L}_M^c} = (\mathcal{L}_M^\circ)^c \subset S_{\mathbb{Q}}^{sing}(M)$.

Conversely consider a slope α in $S_{\mathbb{Q}}^{sing}(M)$. We will work with the singular peg-board representative of $\widehat{HF}(M)$. If $S^{sing}(M) = \{\alpha\}$, then α is the rational longitude of M and thus not an L-space slope. Otherwise, α is in the interval of tangent slopes determined by some corner c; let s_1 and s_2 denote the two line segments meeting at c. First suppose α is in the interior of the interval for the corner c. Let L_α be the line of slope α through c; since L_α is a tangent line, s_1 and s_2 both lie on one side of line L_α. Let L_α' be a small pushoff of L_α that intersects the segments s_1 and s_2 and is disjoint from all lattice points. For sufficiently small ϵ (smaller than

Figure 61. Intersection points for a slope in $S^{sing}(M)$.

the minimum distance between L_α' and any lattice point), replacing the singular representative for $\widehat{HF}(M)$ with the radius ϵ peg-board representative preserves these two intersection points. These points give two generators with the same spin^c grading in the intersection Floer chain complex of $\widehat{HF}(M)$ with L'. Since both L_α' and $\widehat{HF}(M)$ are ϵ-pegboard representatives, they are in minimal position and both generators survive in homology. It follows that $M(\alpha)$ is not an L-space, that is, $\alpha \notin \mathcal{L}_M$. Now suppose α is a boundary of the interval of slopes determined by the corner c; there are slopes arbitrarily close to α in the interior of this interval, and these slopes are not in $\mathcal{L}_M$, so $\alpha \notin \mathcal{L}_M^\circ$. $\square$

Theorem 7.17 is in fact true for arbitrary immersed curves in T (using an appropriate definition of the set of L-space slopes for a curve), not just for the collections of curves associated with manifolds. For curves arising from manifolds, however, we can replace $S^{sing}(M)$ with $S(M)$. To prove this, we first make the following observation about Floer simple manifolds:

Lemma 7.18. *If M is Floer simple then $S(M) \neq \mathbb{R}P^1$.*

Proof. If M is Floer simple, then by [21, Proposition 6] M has simple loop type, that is, for some parametrization (α, β) of ∂M, $\widehat{CFD}(M, \alpha, \beta)$ is a collection of loops (one for each Spin^c structure on M) consisting only of ρ_1, ρ_3 and ρ_{23} arrows. It follows that the corresponding curves in the plane travel only up and right. Clearly $S(M)$ contains only positive slopes (relative to the chosen parametrization), and so $S(M) \neq \mathbb{R}P^1$. Note that reparametrizing the boundary changes this interval but does not change whether or not it is all of $\mathbb{R}P^1$. $\square$

Corollary 7.19. *For any manifold M, $S(M) = S^{sing}(M)$.*

Proof. First observe that if M is not loop type (that is, the associated invariant carries a nontrivial local system) then $S^{sing}(M) = S(M) = \mathbb{R}P^1$ by definition. Consider then the case when M is loop type. In the absence of peg-wrapping $S(M) = S^{sing}(M)$. If there is peg-wrapping, then $S(M) = \mathbb{R}P^1$, and by Lemma 7.18, M is not Floer simple. Equivalently, $\mathcal{L}_M^\circ = \emptyset$ and by Theorem 7.17 $S^{sing}(M) = \mathbb{R}P^1$. $\square$

We remark that, as an immediate corollary of Theorem 7.17, the set $\mathcal{L}_M^\circ$ is either empty, an open interval with rational endpoints, or all of $\mathbb{Q}P^1$ but the rational longitude λ; this is because $S(M)$ contains at least λ and is a closed interval. More generally, it is possible to characterize the set $\mathcal{L}_M$ in terms of the tangent slopes in a singular peg-board diagram: roughly speaking, $\mathcal{L}_M$ is the complement of the set $S'(M)$ obtained by taking the union over all corners of the *interior* of the set of tangent slopes determined by the corner, with the convention that a degenerate corner with two segments of slope α contributes $\mathbb{R}P^1 \setminus \{\alpha\}$, along with the slope of any segment such that $\widehat{HF}(M)$ turns left at both of the corners connected by the segment or turns right at both corners. It follows that $\mathcal{L}_M$ is either empty, a single point, a closed interval with rational endpoints, or $\mathbb{Q}P^1 \setminus \{\lambda\}$; this recovers [46, Theorem 1.6] (see also [23, Theorem 1.2]).

7.6. **The L-space gluing theorem.** We now turn to the proof of Theorem 1.14. In previous works [23, 46, 21] the authors and S. Rasmussen prove this L-space gluing criterion when the manifolds have simple loop type (equivalently, when the manifolds are Floer simple). Interpreting bordered Floer homology as immersed curves with local systems allows us to give an elegant proof of this gluing theorem without requiring the simple loop type hypothesis on the manifolds.

Proof of Theorem 1.14. By the pairing theorem, $M_0 \cup_h M_1$ is equivalent to the intersection Floer homology of $\boldsymbol{\gamma}_0 = \widehat{HF}(M_0)$ and $\boldsymbol{\gamma}_1 = \bar{h}(\widehat{HF}(M_1))$. Suppose first that $M_0 \cup_h M_1$ is not an L-space, implying that there is some $\mathfrak{s}$ so that $\dim \widehat{HF}(M_0 \cup_h M_1, \mathfrak{s}) > 1$. In other words, there are two intersection points x and y between $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$ with the same spin^c grading. This means, in particular, that x and y are intersection points of $\boldsymbol{\gamma}_{0,\mathfrak{s}_0} = \widehat{HF}(M_0, \mathfrak{s}_0)$ and $\boldsymbol{\gamma}_{1,\mathfrak{s}_1} = \bar{h}(\widehat{HF}(M_1, \mathfrak{s}_1))$ for some $\mathfrak{s}_0 \in \mathrm{Spin}^c(M_0)$ and $\mathfrak{s}_1 \in \mathrm{Spin}^c(M_1)$. We will assume for $i \in \{0, 1\}$ that $\boldsymbol{\gamma}_{i,\mathfrak{s}_i}$ contains only one curve with a trivial local system, since otherwise $\mathcal{L}_{M_i}^\circ$ is empty and $\mathcal{L}_{M_0}^\circ \cup h(\mathcal{L}_{M_1}^\circ) \neq \mathbb{Q}P^1$. Let c_0 be the path from x to y in $\boldsymbol{\gamma}_{0,\mathfrak{s}_0}$ and let c_1 be the path from x to y in $\boldsymbol{\gamma}_{1,\mathfrak{s}_1}$. The fact that x and y have the same spin^c grading implies that $[c_0 - c_1] = 0 \in H_1(T)$, or equivalently that the paths lift to form a bigon in $\mathbb{R}^2$, the corners of which are lifts $\tilde{x}$ and $\tilde{y}$ of x and y. Let α be the slope of the line segment connecting $\tilde{x}$ and $\tilde{y}$. By the (extended) mean value theorem, $\tilde{c}_0$ and $\tilde{c}_1$ each contain a point with slope α.

This argument applies for any homotopy representative of $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$. In particular, if we take ϵ-pegboard diagrams relative to peg radius $\epsilon = \frac{1}{n}$ we find a slope α_n that is a tangent slope to both curves. It follows that $\alpha_n \in S_{\frac{1}{n}}(M_0) \cap h(S_{\frac{1}{n}}(M_1))$. Here, by abuse of notation, h refers to the map on slopes induced by h; note that, as maps on slopes, $\bar{h}$ and h agree since the elliptic involution preserves slopes. Since $S_{\frac{1}{n}}(M_0) \cap h(S_{\frac{1}{n}}(M_1))$ for positive integers n gives a nested sequence of nonempty closed sets there is some slope α in the total intersection, which is simply $S(M_0) \cap h(S(M_1))$. Moreover, we can take this α to be rational, since $S(M_0)$ and $h(S(M_1))$ have rational endpoints. By Theorem 7.17 and Corollary 7.19, $S_{\mathbb{Q}}(M_0) = (\mathcal{L}_{M_0}^\circ)^c$ and $S_{\mathbb{Q}}(M_1) = (\mathcal{L}_{M_1}^\circ)^c$. Thus α is not in $\mathcal{L}_{M_0}^\circ \cup h(\mathcal{L}_{M_1}^\circ)$.

Conversely, suppose that there is a slope α not in $\mathcal{L}_{M_0}^\circ \cup h(\mathcal{L}_{M_1}^\circ)$. Equivalently, α is in both $S_{\mathbb{Q}}^{sing}(M_0)$ and $h(S_{\mathbb{Q}}^{sing}(M_1))$, so α is a tangent slope of both $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$ in a singular pegboard diagram. Let c_0 and c_1 be components of $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$, respectively, for which α is a tangent slope to the singular pegboard representative. We may assume that c_0 and c_1 are not solid torus like components. If, for example, c_0 were solid torus like then α would be the rational longitude of M_0 and every component of $\boldsymbol{\gamma}_0$ would have a tangent line of slope α; we could then replace c_0 by another component, at least

one of which is not solid torus like because M_0 is not solid torus like. By reparametrizing ∂M_0 and ∂M_1, we may assume that $\alpha = 0$.

Fixing a small peg radius ϵ, consider ϵ-pegboard representatives of c_0 and c_1 with a chosen orientation on each curve. Recall that each corner wraps around a circle of different radius between ϵ and 2ϵ, with corners that change direction more wrapping closer to the puncture (as usual, if two radii agree we perturb one of them slightly to achieve transversality). Since c_0 and c_1 are in minimal position, to show that $M_0 \cup_h M_1$ is not an L-space it is enough to find two intersection points between c_0 and c_1 with opposite $\mathbb{Z}/2\mathbb{Z}$-grading. It is easy to find two such generators if both curves have peg-wrapping. In this case the curves c_0 and c_1 each contain a portion of sucrve bounding a teardrop enclosing the puncture. Suppose the relevant corners of c_0 and c_1 wrap around circles radius ϵ_0 and ϵ_1 around the puncture. Suppose, without loss of generality, that $\epsilon_0 < \epsilon_1$, so that the circle of radius ϵ_0 is contained in the teardrop bounded by c_1. The curve c_0 also contains two line segments tangent at their ends to the circle of radius ϵ_0, one oriented toward the circle and one oriented away from it. These two segments clearly intersect the teardrop portion of c_1 with opposite sign, and the two corresponding generators of $HF(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1)$ have opposite $\mathbb{Z}/2\mathbb{Z}$ grading. Thus we may assume that there is no peg-wrapping in c_0.

In the singular peg-board diagram for c_0, the line segments connecting corners can be classified as moving upwards, downwards, or horizontally (once we have fixed an orientation on c_0). The segments can not be all upward or all downward, since then $\alpha = 0$ would not be a tangent slope. If the slopes are not all horizontal, we can choose a corner so that the segments preceding and following the corner are up and down, up and horizontal, or horizontal and down. Note that the corresponding corner in an ϵ-pegboard diagram for c_0, which we suppose wraps around a circle of radius ϵ_0, has a point of horizontal tangency on the top side of that circle at the point $(0, \epsilon_0)$. If every line segment in c_0 is horizontal, then every corner has horizontal tangency to the peg, and we can chose the corner so that this point of tangency is on the top side of the peg. Moving away from this point of tangency in either direction, c_0 moves either horizontally or downwards. If it moves horizontally it wraps once around the torus and returns to the peg at another horizontal tangency on the top of the peg; after some number of horizontal segments like this, c_0 must move downwards. Thus, c_0 contains an upward moving piece c_0^I oriented toward the point $(0, \epsilon_0)$ and a downward moving piece c_0^O oriented away from $(0, \epsilon_0')$, where $\epsilon_0' = \epsilon_0$ if there are no horizontal segments in between. Combining c_0^O and c_0^I gives a curve segment as in Figure 63(a); note that if there are horizontal segments in between c_0^O and c_0^I we ignore them—in this case the endpoints do not match up perfectly since $\epsilon_0' \neq \epsilon_0$, but by extending one segment until it crosses the other one we get a piecewise smooth version of the segment in Figure 63(a). Similarly, we can choose a corner of c_1 such that the preceding segment in the singular pegboard diagram moves downward or horizontally and the following segment moves horizontally or upward, and such that in an ϵ-pegboard diagram there is a horizontal tangency to the peg at $(0, -\epsilon_1)$. Let c_1^I be the horizontal or downward piece of c_1 oriented toward $(0, -\epsilon_1)$ and let c_1^O be the horizontal or upward piece oriented away from $(0, -\epsilon_1)$ (or, potentially, $(0, -\epsilon_1')$), as in Figure 63(b). Note that we ignore any full wraps around the peg as well as any horizontal segments between c_1^I and c_1^O. We consider these curves in one fundamental domain of T, the square $[-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$. The intersection of c_0^I and c_0^O with the boundary of this square can be any point with non-positive y-coordinate; the intersection of c_1^I and c_1^O with the boundary of this square can have any non-negative y-coordinate or y-coordinate $-\epsilon_1$. It is clear that if both c_0^I and c_0^O intersect the boundary of the square $[-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$ with strictly negative y-value or if both c_1^I and c_1^O intersect the boundary of the square with strictly positive y-value (equivalently, if the corresponding segments in the singular diagram are not horizontal), then there are two intersections points with opposite $\mathbb{Z}/2\mathbb{Z}$ grading (see Figure 63(c)).

The remaining case is when one of the specified segments in both c_0 and c_1 correspond to horizontal segments in a singular diagram. We consider the case that c_0^O and c_1^O are both horizontal moving to the right (see Figure 64); cases when the incoming segments or both segments are horizontal are similar. In this case, we do not get two intersection points of $c_0^I \cup c_0^O$ and $c_1^I \cup c_1^O$ if we restrict to
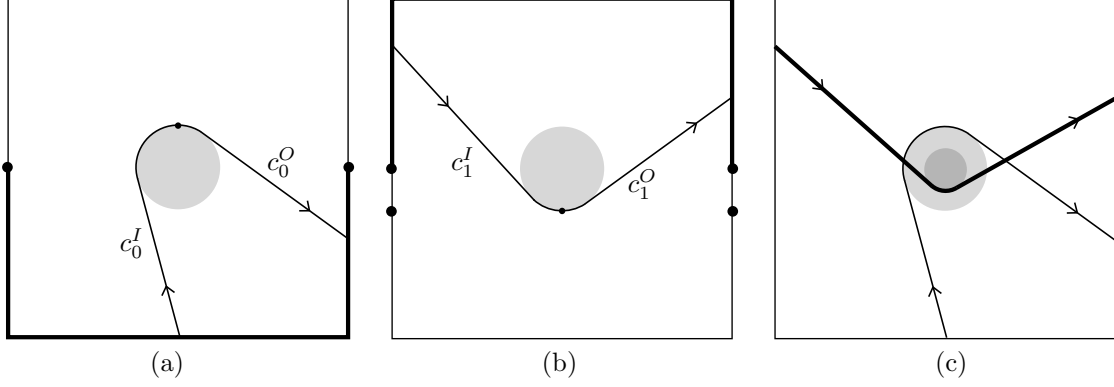
Figure 62. (a) The segments c_0^I and c_0^O in c_0; these segments can enter/leave the square anywhere in the bottom half of its boundary. (b) The segments c_1^I and c_1^O in c_1; these segments can enter/leave the square anywhere in the top half of its boundary or though the points $(\pm\frac{1}{2}, -\epsilon_1)$. (c) The combined curve segments $c_0^I \cup c_0^O$ and $c_1^I \cup c_1^O$ have two intersections of opposite signs if the endpoints of c_0^I and c_0^O are strictly in the bottom half of the square or if the endpoints of c_1^I and c_1^O are strictly in the top half of the square.
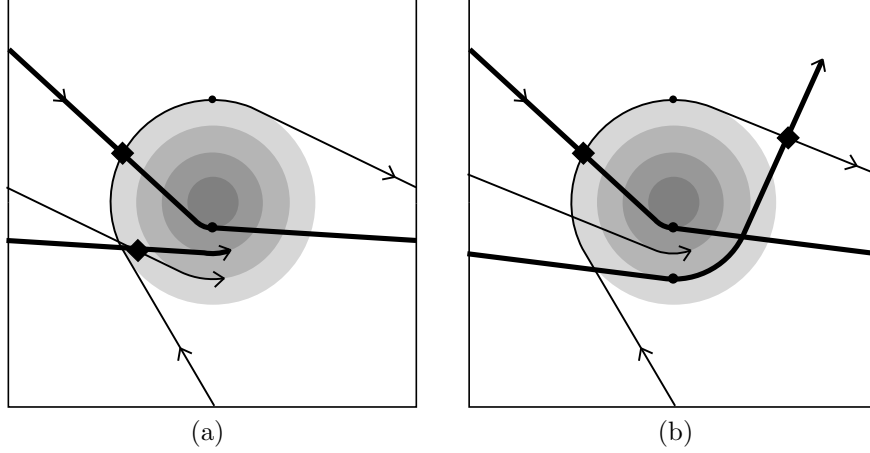
Figure 63. The combined segments $c_0^I \cup c_0^O$ and $c_1^I \cup c_1^O$ in an ϵ-pegboard diagram when c_0^O an c_1^O both correspond to horizontal segments in singular diagrams for c_0 and c_1. There are two cases, depending on which curve wraps with smaller radius at the next corner; in either case there are two intersection points between c_0 and c_1 with opposite sign.

one fundamental domain of T. However, c_0^O and c_1^O both wrap once around T horizontally, leading to new corners of c_0 and c_1, respectively. Let ϵ_0'' and ϵ_1'' be the peg radii associated with these two corners. If $\epsilon_1'' < \epsilon_0''$, then c_0^O intersects c_1^O before reaching the next corner (Figure 64(a)). If $\epsilon_1'' > \epsilon_0''$, then c_1 must not have peg-wrapping at this corner. It follows that c_1 must continue by leaving the peg horizontally or upwards. If it leaves the peg moving upwards, it will intersect with $c_0^I \cup c_0^O$ (Figure 64(b)). If it leaves the peg horizontally, it wraps around T once more to a new corner with some radius ϵ_1''', and we repeat the argument using this corner in place of the corner with radius ϵ_1''. In either case, we find two intersection points between c_0 and c_1 with the opposite sign, as desired. □

7.7. Enumerating toroidal L-spaces.
As our first application of the L-space gluing theorem, we classify small toroidal L-spaces.

Theorem 7.20. *If Y is a prime toroidal L-space then $|H_1(Y)| = 5$ or $|H_1(Y)| \geq 7$. Moreover, if $|H_1(Y)| = 5$ then Y is one of precisely 4 manifolds obtained by gluing trefoil exteriors.*

Proof. We begin by setting conventions. Consider manifolds M_0 and M_1 with torus boundary, and let $h\colon \partial M_1 \to \partial M_0$ be an orientation reversing homeomorphism. Since we are interested in rational homology spheres $Y = M_0 \cup_h M_1$, we may assume that both M_0 and M_1 are rational homology solid tori, with slopes $\lambda_i \in \partial M_i$ representing the rational longitude for $i = 0, 1$. Recall that

$$|H_1(M_0 \cup_h M_1; \mathbb{Z})| = t_0 t_1 |\lambda_0||\lambda_1|\Delta(\lambda_0, h(\lambda_1))$$

where t_i is the order of the torsion subgroup of $H_1(M_i; \mathbb{Z})$ and $|\lambda_i|$ is the order of λ_i in $H_1(M_i; \mathbb{Z})$. The integer $\Delta(\lambda_0, h(\lambda_1))$ is the distance between the relevant slopes, that is, the minimal geometric intersection in ∂M_0. If we fix Spin^c structures $\mathfrak{s}_i$ on M_i, the quantity $|\lambda_0||\lambda_1|\Delta(\lambda_0, h(\lambda_1))$ is the number of Spin^c structures on Y that restrict to $\mathfrak{s}_0$ on M_0 and $\mathfrak{s}_1$ on M_1.

Since M_i is boundary incompressible, there is at least one $\mathfrak{s}_i \in \text{Spin}^c(M_i)$ for which $\gamma_i = \widehat{HF}(M_i, \mathfrak{s}_i)$ is not loose. The argument used in the proof of Theorem 1.8 shows that the curves representing γ_0 and γ_1 intersect in at least four points. Since Y is an L-space, this implies that $|\lambda_0||\lambda_1|\Delta(\lambda_0, h(\lambda_1)) \geq 4$. To have $|H_1(Y; \mathbb{Z})| \leq 6$, we must have $t_0 = t_1 = 1$, *i.e.* both M_i are integer homology solid tori.

By abuse of notation we identify h with a matrix $\left(\begin{smallmatrix} s & r \\ q & p \end{smallmatrix}\right)$, acting on the right, where $(1, 0)$ is identified with λ_1 and $(0, 1)$ represents a meridian μ_1. With this notation, one checks that $|r| = \Delta(\lambda_0, h(\lambda_1))$. Towards minimizing $|H_1(M_0 \cup_h M_1; \mathbb{Z})|$, we may assume that the M_i are integer homology solid tori, so that $|\lambda_0| = |\lambda_1| = 1$ and $|H_1(M_0 \cup_h M_1; \mathbb{Z})| = |r|$.

Now suppose that Y is an L-space; the intervals $\mathcal{L}_i = \mathcal{L}_{M_i}$ are necessarily non-empty. Since $H_1(M_i) = \mathbb{Z}$, the M_i cannot be Floer homology solid tori. Let $b_i\lambda_i + a_i\mu_i$ be an endpoint of $\mathcal{L}_i$, where without loss of generality we assume $a_i > 0$. The characterization of $\mathcal{L}_i$ proved in [46] implies that $\mathcal{L}_i$ is contained either in the interval $[a_i/b_i, a_i/(b_i + 1)]$ or in the interval $[a_i/(b_i - 1), a_i/b_i)$. (Here our notation denotes the cyclic interval not containing 0 with the given endpoints.) By choosing μ_i appropriately, we may assume that $0 > b_i > -a_i$ in the first case, and that $0 < b_i < a_i$ in the second, so that $\mathcal{L}_i \subset [-\infty, -1]$ or $\mathcal{L}_i \subset [1, \infty]$. Finally, by reversing the global orientation on Y, we may assume that $\mathcal{L}_1 \subset [1, \infty]$.

We consider the images $h(0, 1) = (q, p)$, $h(1, 0) = (s, r)$, and $h(1, 1) = (q + s, p + r)$ relative to the interval of slopes $\mathcal{L}_0$. If Y is an L-space then we must have $1 < h(\alpha) < \infty$ for each of these slopes or $-\infty < h(\alpha) < -1$ for each of these slopes.

First consider the case where $\mathcal{L}_0 \subset [1, \infty]$. Then

$$1 < \tfrac{p}{q} < \infty \Rightarrow 1 < |q| < |p|$$
$$1 < \tfrac{r}{s} < \infty \Rightarrow 1 < |s| < |r|$$

where p and q have the same sign and r and s have the same sign. Similarly, considering the inverse homeomorphism $h^{-1} = \left(\begin{smallmatrix} -p & r \\ q & -s \end{smallmatrix}\right)$,

$$1 < -\tfrac{s}{q} < \infty \Rightarrow 1 < |q| < |s|$$
$$1 < -\tfrac{r}{p} < \infty \Rightarrow 1 < |p| < |r|$$

so that p and r have opposite signs and q and s have opposite signs. Additionally, note that $\det(h) = ps - qr = -1$ since h reverses orientation. Without loss of generality, we may assume that $r > 0$; one checks that there are no matrices with $r = 1, 2$ satisfying this list of constraints.

When $r = 3, 4$ the only possible matrices are $\left(\begin{smallmatrix} 2 & 3 \\ -1 & -2 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 3 & 4 \\ -2 & -3 \end{smallmatrix}\right)$. However a homeomorphism of the form $\left(\begin{smallmatrix} r-1 & r \\ 2-r & 1-r \end{smallmatrix}\right)$ is always ruled out since $h(1, 1) = (1, 1)$ in this case (so a boundary L-space slope is not mapped to an interior L-space slope, as required). The same argument applies in the case $r = 6$

From this discussion we observe that in fact we seek to list matrices with determinant -1 satisfying $1 < \frac{p+r}{q+s} < \frac{r}{s} < \frac{p}{q} < \infty$. For $r = 5$, this leaves $\left(\begin{smallmatrix} 3 & 5 \\ -1 & -2 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 2 & 5 \\ -1 & -3 \end{smallmatrix}\right)$ as a complete list of possible L-space gluings in this case.

The case where $\mathcal{L}_0 \subset [-\infty, -1]$ is similar. Potential examples of L-spaces for which $r < 5$ or $r = 6$ are ruled out, by the requirement that $-\infty < \frac{p+r}{q+s} < \frac{r}{s} < \frac{p}{q} < -1$, leaving $\left(\begin{smallmatrix} -3 & 5 \\ 2 & -3 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} -2 & 5 \\ 1 & -2 \end{smallmatrix}\right)$ as a complete list of L-space gluings for $r = 5$.

Next, we claim that in each of these four cases, we must have $\mathcal{L}_1 = [1, \infty]$. We provide the argument when h is given by $\left(\begin{smallmatrix} s & r \\ q & p \end{smallmatrix}\right) = \left(\begin{smallmatrix} 2 & 5 \\ -1 & -3 \end{smallmatrix}\right)$; the other cases are very similar. As observed above, if a/b is the right endpoint of $\mathcal{L}_1$, then $\mathcal{L}_1 \subset [a/(b+1), a/b]$. Since neither μ_0 nor $\lambda_0 + \mu_0$ are elements of $\mathcal{L}_0^\circ$ both $h^{-1}(\mu_0)$ and $h^{-1}(\lambda_0 + \mu_0)$ must be contained in $\mathcal{L}_1^\circ$. We calculate that this includes the slopes $-\frac{s}{q} = 2$ and $\frac{r-s}{q-p} = \frac{3}{2}$, respectively, in the set $\mathcal{L}_1^\circ$, and deduce that $\frac{a}{b+1} < \frac{3}{2}$ and $\frac{a}{b} > 2$ or, equivalently, $\frac{b}{a} + \frac{1}{a} > \frac{2}{3}$ and $\frac{b}{a} < \frac{1}{2}$. It follows that $a < 6$, and a direct check shows that there are no open intervals $(a/(b+1), a/b)$ with $0 < b < a < 6$ containing both $\frac{3}{2}$ and 2. Thus the right-hand endpoint of $\mathcal{L}_1$ is ∞, which implies that M_1 has an integer-homology sphere L-space filling. In turn, this implies that $\mathcal{L}_1 = [2g-1, \infty]$, where g is the genus of M_1. Since $\frac{3}{2}, 2 \in \mathcal{L}_1^\circ$ it must be that $g = 1$.

Reversing the roles of M_0 and M_1, we see that $\mathcal{L}_0 = [1, \infty]$ or $\mathcal{L}_1 = [-\infty, -1]$. In either case, both M_0 and M_1 are prime, boundary incompressible and have two different L-space homology sphere fillings. It follows from work of Ghiggini [14] that they are both homeomorphic to trefoil complements. □

We expect that there will only be finitely many toroidal L-spaces with $|H_1(Y)| = 7$, but cannot prove it, since we can't classify Floer simple manifolds with the same Alexander polynomial as $T(2,5)$. In contrast, it is possible to obtain infinitely many toroidal L-spaces with $|H_1(Y)| = 8$ by gluing the trefoil complement to the twisted I-bundle over the Klein bottle.

7.8. **Satellite L-space knots.** Given a pattern knot P and a companion knot C, both in the three-sphere, denote by $P(C)$ the result of forming a satellite knot (note that this depends on some additional choices). The following is a conjecture of Hom, Lidman, and Vafaee [28, Conjecture 1.7]:

Conjecture 7.21. *If $P(C)$ is an L-space knot, then P and C are L-space knots as well.*

To conclude the section we prove that this conjecture holds.

Proof of Theorem 1.15. Consider the toroidal L-space Y resulting from surgery on a satellite knot K and write $Y = M_0 \cup_h M_1$. By Theorem 1.14, $\mathcal{L}_{M_0}^\circ \cup h(\mathcal{L}_{M_1}^\circ) = \mathbb{Q}P^1$ so that the manifolds M_0 and M_1 are, in fact, Floer simple manifolds (that is, the sets $\mathcal{L}_{M_i}^\circ$ are non-empty). In particular, M_0 the complement of an L-space knot C in S^3 (the companion knot). Since the Seifert longitude λ_C of C is not an L-space slope, it must be that $\lambda_C \in h(\mathcal{L}_{M_1}^\circ)$ so that $\alpha_P = h^{-1}(\lambda_C)$ gives an L-space $M_1(\alpha_P)$.

Now consider the pattern knot P in $D^2 \times S^1$, obtained from $M_K \setminus M_0$, where the boundary of $D^2 \times S^1$ is framed so that $\alpha_P \simeq \{\mathrm{pt}\} \times S^1$. Note that filling this manifold along α_P gives a knot in S^3 (the pattern knot), which we will also denote by P. Further, P must be an L-space knot since there is more than one choice α, which existed by Floer simplicity of M_K (and was required for the construction of Y). This establishes the conjecture. □

There is a converse to this statement that only requires a weaker version of the L-space gluing theorem from [21]; see Hom [27].

Appendix A. Three-manifold invariants are extendable

We outline the proof of Theorem 1.4, which is mainly a matter of assembling the correct elements from [39] and checking that they still apply in our situation. We will assume the reader is familiar with the notation and terminology of [39].

Suppose that (M, α_1, α_2) is a compact oriented three-manifold with parametrized T^2 boundary, and that $\mathcal{H} = (\Sigma, \alpha, \beta)$ is a bordered Heegaard diagram representing it. (There is a notation clash here with our previous use of α, β for the parametrization.)

Our first step is to describe the moduli spaces of pseudo-holomorphic maps used to defined the generalized coefficient maps D_I. The extended torus algebra $\widetilde{A}$ has a natural basis $\langle \rho_I \rangle$, where each ρ_I corresponds to a Reeb chord on the ideal contact boundary $\mathcal{Z} = \partial \overline{\Sigma}$. Suppose that x, y are generators for $\mathcal{H}$, that $B \in \widetilde{\pi}_2(x, y)$, and that $\vec{\rho}$ is a sequence of Reeb chords whose product is equal to ρ_I. (Here, the notation $\widetilde{\pi}_2$ indicates that we consider domains which may have nonzero multiplicity at z, as in [39, Section 10.2]).

We consider decorated sources as in [39, Definition 5.2], but with boundary punctures labeled by arbitrary Reeb chords ρ_I on $\mathcal{Z}$, including those which pass through the basepoint z. (Sources of this type are not sufficient for a bordered theory of HFK^-, since they lack interior punctures, but they are good enough for our needs.) Given a decorated source $S^{\triangleright}$ we can consider the moduli space $\widetilde{\mathcal{M}}^B(x, y; S^{\triangleright})$ of pseudoholomorphic maps $U \colon S \to \Sigma \times [0, 1] \times \mathbb{R}$ which represent the homology class B as in [39, Definition 5.3], but omitting axiom (M-9) (the requirement that S has 0 multiplicity near the basepoint). Similarly, if $\vec{P}$ is an ordering of the punctures of S, we can consider the open subset $\widetilde{\mathcal{M}}^B(x, y; S^{\triangleright}; \vec{P})$ for which projection to the $\mathbb{R}$ direction induces the given ordering $\vec{P}$. Finally, we let $\mathcal{M}^B(x, y; S^{\triangleright}; \vec{P})$ be the corresponding reduced moduli space.

Most of the analysis of pseudo-holomorphic curves in [39, Sections 5.2-5.6] carries over unchanged to this context, with the notable exception of the fact that boundary degenerations (see [39, Lemma 5.48]) can (and do) occur. More specifically, the proof of [39, Lemma 5.48] relies on the fact that $\pi_2(\Sigma, \alpha) = 0$. Since we are allowing sources whose boundary maps to ρ_0, we must instead consider $\widetilde{\pi}_2(\Sigma, \alpha) = \mathbb{Z}$. The generator of this group is the domain $[\Sigma]$, whose boundary is $\sum_{i=0}^3 \rho_i$. (Notice that since we do not allow sources with interior punctures, $\widetilde{\pi}_2(\Sigma, \beta) = 0$.)

Now suppose we are given a sequence of Reeb chords $\vec{\rho}$ and that B is compatible with $\vec{\rho}$ in the sense of [39, Definition 5.68]. Following [39, Definition 5.68], we let $\mathcal{M}^B(x, y; \vec{\rho})$ be the union of those moduli spaces $\mathcal{M}^B(x, y; S^{\triangleright})$ whose constituent pseudo-holomorphic curves are *embedded* and for which $[\vec{P}] = \vec{\rho}$.

Finally, following [39, Definition 6.3], we let

$$D_I(x) = \sum_y \sum_{\{\vec{\rho} | a(-\vec{\rho}) = \rho_I\}} \sum_{\{B | \mathrm{ind}(B, \vec{\rho}) = 1\}} \#(\mathcal{M}^B(x, y; \vec{\rho}))y$$

and define $\widetilde{\partial}(x) = \sum_I \rho_I \otimes D_I(x)$, where the sum runs over all indices I which contain at most one 0.

Proof of Theorem 1.4. We must show that $\widetilde{\partial}^2(x) = W \otimes x$ modulo terms of the form $\rho_I \otimes y$, where I contains at least two 0's. To do so, we argue as in the proofs of [39, Proposition 6.7] and [39, Proposition 11.30]. Write

$$\widetilde{\partial}^2(x) = \sum_I \sum_y n_{xy, I} \rho_I y.$$

Fix y and I which contains at most one 0, and consider the moduli space

$$\mathcal{M}(x,y;\rho_I) = \bigcup_{\{\vec{\rho}\,|\,a(-\vec{\rho})=\rho_I\}} \;\bigcup_{\{B\,|\,\text{ind}(B,\vec{\rho})=2\}} \mathcal{M}^B(x,y;\vec{\rho})$$

As in the proof of [39, Theorem 6.7], $n_{xy,I}$ is the mod 2 number of ends $\mathcal{M}(x,y;\rho_I)$ which do not correspond to boundary degenerations.

Suppose that boundary degenerations appear as ends of $\mathcal{M}(x,y;\rho_I)$. Index considerations dictate that such a degeneration must be of the form $\varphi_1 \vee \varphi_2$, where $\varphi_1 \in \mathcal{M}^{B_1}$ for some $B_1 \in \pi_2(\Sigma,\alpha)$ with index k, and $\varphi_2 \in \mathcal{M}^{B_2}$, where $B_2 \in \pi_2(x,y)$ has index $-k$. Assuming we have chosen a generic almost complex structure, all moduli spaces have their expected dimensions, and $k = 0$. Thus φ_2 is a union of trivial strips. The domain of B_1 is some multiple of the generator of $\pi_2(\Sigma,\alpha)$. Since 0 occurs in I at most once, the domain of B_1 must be Σ, which implies that $|I| = 4$. We conclude that if $|I| \neq 4$ or or $x \neq y$, there are no boundary degenerations and $n_{xy,I} = 0$. The case $|I| = 4$ and $x = y$ was studied by Lipshitz, Ozsváth, and Thurston in [39, Proposition 11.30]. They showed that the number of boundary degenerations in this case is equal to 1 mod 2, so $n_{xy,I} = 1$. $\square$

References

[1] Mohammed Abouzaid. On the Fukaya categories of higher genus surfaces. *Adv. Math.*, 217(3):1192–1235, 2008.

[2] Akram Alishahi and Robert Lipshitz. Bordered Floer homology and incompressible surfaces. *Ann. Inst. Fourier (Grenoble)*, 69(4):1525–1573, 2019.

[3] Denis Auroux. Fukaya categories and bordered Heegaard-Floer homology. In *Proceedings of the International Congress of Mathematicians. Volume II*, pages 917–941. Hindustan Book Agency, New Delhi, 2010.

[4] Denis Auroux. Fukaya categories of symmetric products and bordered Heegaard-Floer homology. *J. Gökova Geom. Topol. GGT*, 4:1–54, 2010.

[5] Denis Auroux and Ivan Smith. Fukaya categories of surfaces, spherical objects and mapping class groups. *Forum Math. Sigma*, 9:Paper No. e26, 50, 2021.

[6] Kenneth L. Baker and Kimihiko Motegi. Seifert vs. slice genera of knots in twist families and a characterization of braid axes. *Proc. Lond. Math. Soc. (3)*, 119(6):1493–1530, 2019.

[7] John A. Baldwin and Steven Sivek. Khovanov homology detects the trefoils. *Duke Math. J.*, 171(4):885–956, 2022.

[8] Michel Boileau, Steven Boyer, and Cameron McA. Gordon. Branched covers of quasi-positive links and L-spaces. *J. Topol.*, 12(2):536–576, 2019.

[9] Steven Boyer and Adam Clay. Foliations, orders, representations, L-spaces and graph manifolds. *Adv. Math.*, 310:159–234, 2017.

[10] Steven Boyer, Cameron McA. Gordon, and Liam Watson. On L-spaces and left-orderable fundamental groups. *Math. Ann.*, 356(4):1213–1245, 2013.

[11] Vin de Silva, Joel W. Robbin, and Dietmar A. Salamon. Combinatorial Floer homology. *Mem. Amer. Math. Soc.*, 230(1080):v+114, 2014.

[12] Eaman Eftekhary. Bordered Floer homology and existence of incompressible tori in homology spheres. *Compos. Math.*, 154(6):1222–1268, 2018.

[13] Michael Freedman, Joel Hass, and Peter Scott. Closed geodesics on surfaces. *Bull. London Math. Soc.*, 14(5):385–391, 1982.

[14] Paolo Ghiggini. Knot Floer homology detects genus-one fibred knots. *Amer. J. Math.*, 130(5):1151–1169, 2008.

[15] Thomas Gillespie. L-space fillings and generalized solid tori. Preprint, arXiv:1603.05016.

[16] Cameron Gordon and Tye Lidman. Corrigendum to "Taut foliations, left-orderability, and cyclic branched covers". *Acta Mathematica Vietnamica*, Jun 2017.

[17] Cameron McA. Gordon and Tye Lidman. Taut foliations, left-orderability, and cyclic branched covers. *Acta Math. Vietnam.*, 39(4):599–635, 2014.

[18] F. Haiden, L. Katzarkov, and M. Kontsevich. Flat surfaces and stability structures. *Publ. Math. Inst. Hautes Études Sci.*, 126:247–318, 2017.

[19] Jonathan Hanselman. Bordered Heegaard Floer homology and graph manifolds. *Algebr. Geom. Topol.*, 16(6):3103–3166, 2016.

[20] Jonathan Hanselman. Splicing integer framed knot complements and bordered Heegaard Floer homology. *Quantum Topol.*, 8(4):715–748, 2017.

[21] Jonathan Hanselman, Jacob Rasmussen, Sarah Dean Rasmussen, and Liam Watson. L-spaces, taut foliations, and graph manifolds. *Compos. Math.*, 156(3):604–612, 2020.

[22] Jonathan Hanselman, Jacob Rasmussen, and Liam Watson. Heegaard Floer homology for manifolds with torus boundary: properties and examples. Preprint, arXiv:1810.10355.

[23] Jonathan Hanselman and Liam Watson. A calculus for bordered Floer homology. *Geom. Topol.*, to appear.

[24] Matthew Hedden and Adam Simon Levine. Splicing knot complements and bordered Floer homology. *J. Reine Angew. Math.*, 2016:129–154, 2014.

[25] Matthew Hedden and Yi Ni. Manifolds with small Heegaard Floer ranks. *Geom. Topol.*, 14(3):1479–1501, 2010.

[26] Matthew Hedden and Yi Ni. Khovanov module and the detection of unlinks. *Geom. Topol.*, 17(5):3027–3076, 2013.

[27] Jennifer Hom. Satellite knots and L-space surgeries. *Bull. Lond. Math. Soc.*, 48(5):771–778, 2016.

[28] Jennifer Hom, Tye Lidman, and Faramarz Vafaee. Berge-Gabai knots and L-space satellite operations. *Algebr. Geom. Topol.*, 14(6):3745–3763, 2014.

[29] Jennifer Hom, Tye Lidman, and Liam Watson. The Alexander module, Seifert forms, and categorification. *Journal of Topology*, 10(1):22–100, 2017.

[30] András Juhász. Holomorphic discs and sutured manifolds. *Algebr. Geom. Topol.*, 6:1429–1457, 2006.

[31] Artem Kotelskiy. Bordered theory for pillowcase homology. *Math. Res. Lett.*, 26(5):1467–1516, 2019.

[32] Artem Kotelskiy, Liam Watson, and Claudius Zibrowius. Thin links and Conway spheres. Preprint, arXiv:2105.06308.

[33] P. B. Kronheimer and T. S. Mrowka. Khovanov homology is an unknot-detector. *Publ. Math. Inst. Hautes Études Sci.*, (113):97–208, 2011.

[34] Yankı Lekili and Timothy Perutz. Fukaya categories of the torus and Dehn surgery. *Proc. Natl. Acad. Sci. USA*, 108(20):8106–8113, 2011.

[35] Adam Simon Levine. Knot doubling operators and bordered Heegaard Floer homology. *J. Topol.*, 5(3):651–712, 2012.

[36] Tye Lidman. Heegaard Floer homology and triple cup products. Preprint, arXiv:1011.4277.

[37] Robert Lipshitz. *A Heegaard Floer invariant of bordered three-manifolds*. PhD thesis, Stanford University, 2006.

[38] Robert Lipshitz, Peter S. Ozsváth, and Dylan P. Thurston. Bimodules in bordered Heegaard Floer homology. *Geom. Topol.*, 19(2):525–724, 2015.

[39] Robert Lipshitz, Peter S. Ozsvath, and Dylan P. Thurston. Bordered Heegaard Floer homology. *Mem. Amer. Math. Soc.*, 254(1216):viii+279, 2018.

[40] Lee Mosher. What is a train track? *Notices of the AMS*, 50(3):354–355, 2003.

[41] András Némethi. Links of rational singularities, L-spaces and LO fundamental groups. *Invent. Math.*, 210(1):69–83, 2017.

[42] Peter Ozsváth and Zoltán Szabó. Heegaard diagrams and holomorphic disks. In *Different faces of geometry*, volume 3 of *Int. Math. Ser. (N. Y.)*, pages 301–348. Kluwer/Plenum, New York, 2004.

[43] Peter Ozsváth and Zoltán Szabó. Holomorphic disks and genus bounds. *Geom. Topol.*, 8:311–334 (electronic), 2004.

[44] Peter Ozsváth and Zoltán Szabó. On the Heegaard Floer homology of branched double-covers. *Adv. Math.*, 194(1):1–33, 2005.

[45] Ina Petkova. The decategorification of bordered Heegaard Floer homology. *J. Symplectic Geom.*, 16(1):227–277, 2018.

[46] Jacob Rasmussen and Sarah Dean Rasmussen. Floer simple manifolds and L-space intervals. *Adv. Math.*, 322:738–805, 2017.

[47] Sarah Dean Rasmussen. L-space intervals for graph manifolds and cables. *Compos. Math.*, 153(5):1008–1049, 2017.

[48] Peter Scott. The geometries of 3-manifolds. *Bull. London Math. Soc.*, 15(5):401–487, 1983.

[49] Gudlaugur Thorbergsson. Closed geodesics on non-compact Riemannian manifolds. *Math. Z.*, 159(3):249–258, 1978.

[50] Liam Watson. Heegaard Floer homology solid tori. AMS special session: Knots, links, and three-manifolds, San Diego, January 2013.

[51] Yang Xiu. Elliptic involution on knot complements. Preprint, arXiv:1511.02298.

[52] Rumen Zarev. Bordered Floer homology for sutured manifolds. Preprint, arXiv:0908.1106.

Department of Mathematics, Princeton University.
E-mail address: jh66@princeton.edu

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge.
E-mail address: J.Rasmussen@dpmms.cam.ac.uk

Department of Mathematics, University of British Columbia.
E-mail address: liam@math.ubc.ca