

Convex-compact subgroups of the Goeritz group

Bena Tshishiku

January 27, 2022

Abstract

Let $G < \text{Mod}_2$ be the Goeritz subgroup of the genus-2 mapping class group. We show that finitely-generated, purely pseudo-Anosov subgroups of G are convex cocompact in Mod_2 , addressing a case of a general question of Farb–Mosher. We also give a simple criterion to determine if a Goeritz mapping class is pseudo-Anosov, which we use to give very explicit convex-cocompact subgroups. In our analysis, a central role is played by the primitive disk complex \mathcal{P} . In particular, we (1) establish a version of the Masur–Minsky distance-formula for \mathcal{P} , (2) classify subsurfaces $X \subset S$ that are infinite-diameter holes of \mathcal{P} , and (3) show that \mathcal{P} is quasi-isometric to a coned-off Cayley graph for G .

Contents

1	Introduction	2
2	Background and auxiliary results	7
3	Classification of holes for $\mathcal{P}(V)$	21
4	Distance formula for $\mathcal{P}(V)$	40
5	Purely pseudo-Anosov implies convex cocompact	48
6	Characterization of Goeritz pseudo-Anosov elements	51
A	Classification of genus-1 fibered knots	53

Key words and phrases. Geometric group theory, convex cocompactness, mapping class groups, curve complexes, 3-manifolds, Heegaard splittings

1 Introduction

Let $S^3 = V \cup_S \widehat{V}$ be the genus-2 Heegaard splitting of S^3 (so V, \widehat{V} are handlebodies and $S = \partial V = \partial \widehat{V}$ is a closed surface of genus 2). The *Goeritz group* $\mathbb{G} < \text{Mod}(S)$ is the group of isotopy classes of homeomorphisms of S that extend to both handlebodies V, \widehat{V} . The algebraic structure of \mathbb{G} is well understood [Goe33, Sch04, Akb08, Cho08]. In particular, \mathbb{G} is generated by the elements $\alpha, \beta, \gamma, \delta$ pictured in Figure 1, and with respect to this generating set, \mathbb{G} has a simple amalgamated structure:

$$\mathbb{G} \cong \left[\begin{array}{ccc} (\mathbb{Z}_2 & \times & \mathbb{Z}_2) & \rtimes & \mathbb{Z} \\ \alpha & & \gamma & & \beta \end{array} \right] *_{\mathbb{Z}_2 \times \mathbb{Z}_2} \left[\begin{array}{ccc} (\mathbb{Z}_3 & \rtimes & \mathbb{Z}_2) & \times & \mathbb{Z}_2 \\ \delta & & \gamma & & \alpha \end{array} \right] \quad (1)$$

In this paper we study geometric aspects of \mathbb{G} as a subgroup of $\text{Mod}(S)$. Our main results follow.

Theorem A (Purely pseudo-Anosov implies convex cocompact). *Finitely-generated, purely pseudo-Anosov subgroups of the genus-2 Goeritz group \mathbb{G} are convex cocompact in the genus-2 mapping class group $\text{Mod}(S)$.*

Farb–Mosher [FM02, Question 1.5] ask if *purely pseudo-Anosov* implies *convex co-compact* for finitely-generated (free) subgroups of mapping class groups. This question is a special case of Gromov’s hyperbolicity question; see [Bes04, Question 1.1] and [KL07]. In particular, by [FM02, Ham05], Theorem A implies that for any purely pseudo-Anosov subgroup G of the Goeritz group \mathbb{G} , the associated extension group \tilde{G}

$$1 \rightarrow \pi_1(S) \rightarrow \tilde{G} \rightarrow G \rightarrow 1$$

is Gromov hyperbolic.

The question of Farb–Mosher remains open in general, but the answer is “yes” in many special cases of geometric interest, including Veech groups, certain hyperbolic 3-manifold subgroups [KLS09, DKL14], and certain right-angled Artin subgroups [MT16, KMT17]. Theorem A is a new direction in this family of results.

Characterizing pseudo-Anosov elements of the Goeritz group. We complement Theorem A with the following Theorem B, which gives a simple criterion to determine if $g \in \mathbb{G}$ is pseudo-Anosov, and in particular shows that the Goeritz group contains many independent pseudo-Anosov elements, and hence many purely pseudo-Anosov subgroups (c.f. [FM02, Thm. 1.4]).

Theorem B (Pseudo-Anosov characterization). *Let $g \in \mathbb{G} < \text{Mod}(S)$ be a mapping class in the genus-2 Goeritz group. Then g is pseudo-Anosov if and only if g is not conjugate into any of the following subgroups, where the elements $\alpha, \beta, \gamma, \delta$ are shown in Figure 1.*

- (i) (primitive-disk stabilizer) $\langle \alpha, \beta, \gamma\delta \rangle \cong (\mathbb{Z}_2 \times \mathbb{Z}) * \mathbb{Z}_2$
- (ii) (reducing-sphere stabilizer) $\langle \alpha, \beta, \gamma \rangle \cong (\mathbb{Z}_2 \times \mathbb{Z}) \rtimes \mathbb{Z}_2$
- (iii) (primitive pants-decomposition stabilizer) $\langle \alpha, \gamma, \delta \rangle \cong \mathbb{Z}_2 \times S_3$
- (iv) (figure-8 knot stabilizer) $\langle \alpha, \beta\delta\beta^{-1}\delta, \gamma\delta \rangle \cong \mathbb{Z}_2 \times (\mathbb{Z} \rtimes \mathbb{Z}_2)$

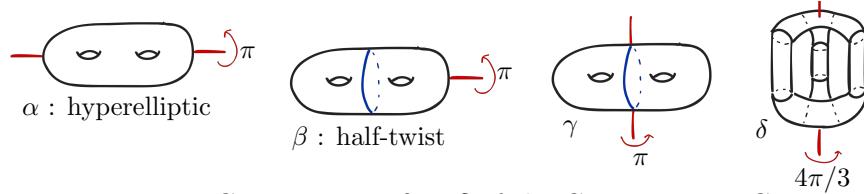


Figure 1: Generators $\alpha, \beta, \gamma, \delta$ of the Goeritz group \mathbb{G} .

In practice, it is easy to use Theorem B and the presentation (1) to decide the Nielsen–Thurston type (finite-order, reducible, or pseudo-Anosov) of a mapping class $g \in \mathbb{G}$ given as a word in these generators.

Remark 1.1. The “figure-8 knot stabilizer” in Theorem B comes from the fibering of the figure-8 knot complement

$$(T^2 \setminus \text{pt}) \rightarrow S^3 \setminus K \rightarrow S^1. \quad (2)$$

The union of two fibers of (2) with K is a genus-2 Heegaard surface in S^3 ; see Figure 2. The figure-8 knot stabilizer is the subgroup of the Goeritz group that fixes K . For example, the monodromy of (2) yields an infinite-order element of the figure-8 knot stabilizer that is conjugate to $\beta\delta\beta^{-1}\delta$.

The trefoil knot is also a genus-1 fibered knot. Its monodromy gives a finite-order element of the Goeritz group, which is accounted for in part (iii) of Theorem B.

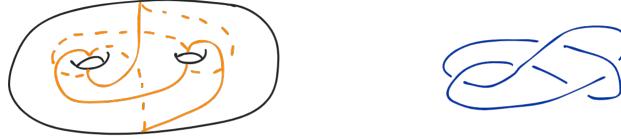


Figure 2: The figure-8 knot embedded on the standard genus-2 Heegaard surface.

Refined Nielsen–Thurston classification. Theorem B also classifies the ways an element of the Goeritz group can be reducible (in terms of what structure is preserved). This provides a refinement of the Nielsen–Thurston classification for reducible elements of the Goeritz group. We illustrate this with the following result.

Corollary 1.2 (Classification of canonical reduction systems for Goeritz elements). *Let $g \in \mathbb{G}$ be an infinite-order, reducible element. Then the canonical reduction system $\text{CRS}(g)$ is either*

- (i) (weakly reducing pair) a pair of curves c and \hat{c} that bound primitive disks in V and \hat{V} respectively,
- (ii) (reducing curve) a curve c that bounds disks in both V and \hat{V} , or
- (iii) (figure-8 curve) a curve c whose embedding in S^3 is the figure-8 knot.

Explicit convex cocompact examples. Using Theorems A and B, one may give explicit examples of convex cocompact subgroups of $G < \text{Mod}(S)$. We illustrate this with the following corollary, which is motivated by Problem 3.6 of [Mos06] that asks for new constructions/examples.

Corollary 1.3 (Explicit convex cocompact subgroups). *Let $\beta, \delta \in \mathbb{G}$ be the elements pictured in Figure 1. For each $n \geq 2$, the subgroup*

$$G_n = \langle \beta^n \delta, \delta \beta^n \rangle < \text{Mod}(S)$$

is purely pseudo-Anosov and hence convex cocompact by Theorem A.

Proof of Corollary 1.3. By the presentation (1), the subgroup $\langle \beta, \delta \rangle \subset \mathbb{G}$ is isomorphic to $\mathbb{Z} * \mathbb{Z}_3$. This implies that nontrivial products of the generators of G_n are cyclically reduced with no cancellation. From this one quickly deduces that no word in G_n is conjugate into any of the subgroups listed in Theorem B. Therefore G_n is purely pseudo-Anosov. \square

Explicit examples are rare in the study of convex cocompact subgroups of mapping class groups. More general constructions of convex cocompact groups, e.g. using ping-pong as in [Mos97, Min11], involve replacing a given collection of pseudo-Anosov elements by large (non-explicit!) powers of these elements. Other explicit examples in genus 2, different from the ones we produce, appear in [MT16, §7] and [Whi00, §6] (the latter examples are convex cocompact by [KLS09]).

Techniques. The main technical results used in the proofs of Theorems A and B are Theorems C and D below.

To show a group $G < \text{Mod}(S)$ is convex cocompact, it suffices to show that the orbit map (with respect to some basepoint) of G acting on the curve complex $\mathcal{C}(S)$ is a quasi-isometric embedding [KL08]. For studying subgroups $G < \mathbb{G}$, it is convenient to choose the basepoint so that the orbit is contained in the primitive disk complex

$\mathcal{P}(V) \subset \mathcal{C}(S)$ associated to the Heegaard splitting $S^3 = V \cup_S \widehat{V}$ (see §2 for the definition). Theorem C relates the distance in $\mathcal{P}(V)$ to the distance in $\mathcal{C}(S)$ and more generally $\mathcal{C}(X)$, where $X \subset S$ is a subsurface whose complement does not support a primitive disk; such subsurfaces are called *holes* for $\mathcal{P}(V)$; see §2.

Theorem C (Distance formula). *For $c \gg 0$ sufficiently large, there is $K > 0$ so that for any two vertices D, E of $\mathcal{P}(V)$,*

$$\frac{1}{K} \sum_X \{d_X(D, E)\}_c - K \leq d_{\mathcal{P}(V)}(D, E) \leq K \sum_X \{d_X(D, E)\}_c + K, \quad (3)$$

where the sums range over subsurfaces $X \subset S$ that are holes for $\mathcal{P}(V)$, $d_X(D, E)$ is defined by subsurface projection, and $\{\cdot\}_c$ is the cutoff function. Definitions appear in §2.

Distance formulas as in (3) originate in the work of Masur–Minsky [MM00]. We arrive at (3) by closely following the work of Masur–Schleimer [MS13], who prove an analogous result for the disk complex $\mathcal{D}(V)$ and provide an axiomatic approach to distance formulas more generally.

To apply Theorem C, it is necessary to classify the holes $X \subset S$ with the property that the image of the subsurface projection $\pi_X : \mathcal{P}(V) \rightarrow \mathcal{C}(X)$ has infinite diameter. For brevity we say that X has *infinite diameter* (with respect to $\mathcal{P}(V)$).

Theorem D (Classification of large holes for $\mathcal{P}(V)$). *Let $X \subset S$ be a hole for $\mathcal{P}(V)$ with diameter ≥ 61 . Then either $X = S$ or X is a genus-1 Seifert surface for the figure-8 knot. In either case there exists $g \in \mathbb{G}$ that preserves X and so that $g|_X$ is pseudo-Anosov. Consequently, every hole of diameter ≥ 61 has infinite diameter.*

Remark 1.4 (Genus-1 fibered knots). The proof of Theorem D uses the classification of genus-1 fibered knots in S^3 . The only such knots are the figure-8 and the (left/right-handed) trefoil [GA70]. The appendix contains a modern proof of this fact.

Remark 1.5 (Relation to the work of Masur–Schleimer). Our proofs of the distance formula and classification of holes for $\mathcal{P}(V)$ (Theorems C and D) build on the work of [MS13], who prove similar results for the disk complex $\mathcal{D}(V)$. The primitive disk complex $\mathcal{P}(V)$ is a subcomplex of $\mathcal{D}(V)$. As such, the techniques of [MS13] are useful for this paper; however, we note that the results of [MS13] cannot be applied directly to prove our main results, since e.g. there are holes for $\mathcal{P}(V)$ that are not holes for $\mathcal{D}(V)$, and there are holes for $\mathcal{D}(V)$ that have infinite diameter with respect to $\mathcal{D}(V)$, but finite diameter with respect to $\mathcal{P}(V)$. See Remark 2.12. Many of the arguments of [MS13] “break” when restricting only to primitive disks, and some of the work in the current paper involves finding ways to “repair” these arguments.

Remark 1.6 (Holes for the reducing sphere complex). Using our analysis one could also classify holes for the complex $\mathcal{R}(V, \widehat{V})$ of reducing spheres for the Heegaard splitting. This problem was posed by Schleimer [Sch20, §7]. One might try to use $\mathcal{R}(V, \widehat{V})$ to prove Theorem A (indeed, this is one motivation for Schleimer’s problem), but it seems more convenient to work with the primitive disks instead of reducing spheres. For example, it’s easier to construct paths between primitive disks using surgery.

Remark 1.7 (The assumption $g = 2$). Our approach to proving each of Theorems A – D fails to extend to genus $g \geq 3$. A fundamental difference between $g = 2$ and $g \geq 3$ is that in genus 2, every surgery path between a pair of primitive disks is a path of *primitive* disks, but this is not true in higher genus; see [CKL20, Thm. 1.1].

About the proof of Theorem A. The proof of Theorem A is given in §5. Here we explain the main ideas using the following diagram.

$$\begin{array}{ccccc}
\mathbb{G} & \xrightarrow{\quad} & \text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_D) & & \\
\uparrow & \nearrow & \downarrow & & \\
G & \xrightarrow{\quad} & \mathcal{P}(V) & \xrightarrow{\quad} & \\
& \searrow & \downarrow & & \\
& & \mathcal{C}(S) & &
\end{array}$$

Assuming $G < \mathbb{G}$ is purely pseudo-Anosov, to show G is convex cocompact in $\text{Mod}(S)$, it suffices to show that the orbit map $G \rightarrow \mathcal{C}(S)$ is a q.i. embedding [KL08]. First we use Theorem C (distance formula) and results of [BBKL20] to reduce to showing that the orbit map $G \rightarrow \mathcal{P}(V)$ is a q.i. embedding (see Prop. 4.3). This step is nontrivial because the inclusion $\mathcal{P}(V) \hookrightarrow \mathcal{C}(S)$ is not a q.i. embedding by Theorems C and D. Next we show, using work of Cho [Cho08], that $\mathcal{P}(V)$ is quasi-isometric to a cone-off of the Cayley graph $\Gamma_{\mathbb{G}}$ with respect to the stabilizer \mathbb{G}_D of a primitive disk (see Prop. 5.1). Finally $G \hookrightarrow \mathbb{G}$ is a quasi-isometric embedding because \mathbb{G} is virtually free, and to prove that the composition $G \hookrightarrow \mathbb{G} \hookrightarrow \text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_D)$ is a quasi-isometric embedding, we use a criterion of Abbott–Manning [AM21] (see Prop. 5.3).

About the proof of Theorem B. Given Theorem D (classification of holes), the proof of Theorem B is a straightforward argument using the canonical reduction system and some results of Oertel [Oer02] about homeomorphisms of handlebodies. The details are in §6.

Acknowledgements. The author thanks J. Manning for helpful conversations about [AM21] and for observing that $\mathcal{P}(V)$ is quasi-isometric to a coned-off Cayley graph. Thanks also to J. Behrstock and J. Gaster for helpful conversations and to C. Leininger for comments on a draft of this paper.

2 Background and auxiliary results

In this section we quickly summarize the necessary terminology and results that we will need in the rest of the paper. At the end of the paper we have included a guide to the notation for easy reference.

2.1 Surfaces and curves

We use $\Sigma_{g,b}$ to denote the compact oriented surface of genus g with b boundary components. A *simple closed curve* (usually referred to simply as a *curve*) c on a surface Σ is an embedded circle. A curve c is *essential* if it does not bound a disk and *non-peripheral* if it is not isotopic to a boundary component. We will often use the same notation for a curve and its isotopy class. For a curve $c \subset \Sigma$, we use $n(c) \subset \Sigma$ to denote a small regular neighborhood of c . For a pair of isotopy classes c_1, c_2 , the *geometric intersection number* is denoted $i(c_1, c_2)$.

2.2 Primitive disks and curve complexes

Fix a genus-2 Heegaard splitting $S^3 = V \cup_S \widehat{V}$. Some of the definitions of this section can be stated more generally, but several of the results are specific to genus 2, so we focus on that case.

Primitive disks. A properly embedded disk $D \subset V$ is called *primitive* if there exists a properly embedded disk $\widehat{D} \subset \widehat{V}$ so that $i(\partial D, \partial \widehat{D}) = 1$.

For example, in Figure 3, the disks $\widehat{E}_1, \widehat{E}_2$ illustrate that $E_1, E_2 \subset V$ are primitive disks. We call E_1, E_2 and $\widehat{E}_1, \widehat{E}_2$ a *dual pair* (these are the curves for the standard genus-2 Heegaard diagram for S^3).

Remark 2.1 (Primitivity testing). A disk $D \subset V$ is primitive if and only if the conjugacy class of ∂D in $\pi_1(\widehat{V}) \cong F_2$ is primitive. More precisely, if we orient ∂D , $\partial \widehat{E}_1$, and $\partial \widehat{E}_2$, then the intersection pattern of ∂D with \widehat{E}_1 and \widehat{E}_2 determines a word w_D in x_1 and x_2 , say. Then disk D is primitive if and only if the word w_D , when cyclically reduced, is part of a free basis for the free group $F_2 = \langle x_1, x_2 \rangle$. See [Zie65] and also [Gor87].

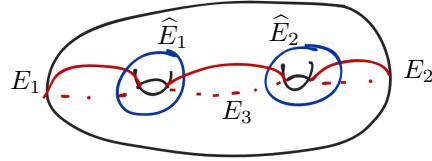


Figure 3: Standard Heegaard diagram for the Heegaard splitting $S^3 = V \cup_S \widehat{V}$. Here E_1, E_2, E_3 are disks in V (the “inside” handlebody), and $\widehat{E}_1, \widehat{E}_2$ are disks in \widehat{V} .

There is an easy algorithm to decide if a word in F_2 is primitive [Pig06]. A simple obstruction to a cyclically reduced word $w \in F_2$ being primitive is if w contains both x_1 and x_1^{-1} (or x_2 and x_2^{-1}). For example, Figure 4 gives an example of a non-separating curve $c \subset S$ that bounds a disk, but does not bound a primitive disk.

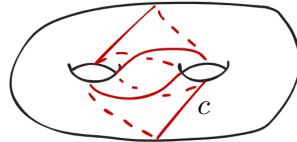


Figure 4: The curve c is nonseparating and bounds a disk in $D \subset V$, but D is not primitive. Indeed, (for a particular choice of orientations) the word w_D is conjugate to $x_1x_2^{-1}x_1x_2x_1^{-1}x_2$, which is not primitive.

Primitive disk complex. For a compact surface X , we use $\mathcal{C}(X)$ to denote the curve complex, which is the simplicial complex with a vertex for each isotopy class of essential, non-peripheral simple closed curve on X , and a k -simplex for each $(k+1)$ -tuple of isotopy classes that can be realized disjointly. We define the distance $d_{\mathcal{C}(X)}(a, b)$ between two vertices of $\mathcal{C}(X)$ as the fewest edges in an edge path in $\mathcal{C}(X)$ connecting them.

The key complex of interest in this paper is the *primitive disk complex* $\mathcal{P}(V)$, which is the full subcomplex of the curve complex $\mathcal{C}(S)$ whose vertices are isotopy classes of simple closed curves on S that bound primitive disks in V . This complex was defined in [Cho08]. Sometimes we abuse notation and refer to D (instead of ∂D) as a vertex of $\mathcal{P}(V)$.

By definition, $\mathcal{P}(V)$ is a subcomplex of the (more familiar) disk complex $\mathcal{D}(V) \subset \mathcal{C}(S)$ whose vertices are curves on S that bound disks in V . The disk complex will be helpful for our study of $\mathcal{P}(V)$ because analogues of Theorems C and D are known for $\mathcal{D}(V)$ [MS13].

To understand $\mathcal{P}(V)$, we will use two other complexes: the reducing sphere complex $\mathcal{R}(V, \hat{V})$ and a new complex that we call the Heegaard marking complex $\mathcal{M}(V, \hat{V})$.

Reducing sphere complex. The *reducing sphere complex* $\mathcal{R}(V, \hat{V})$ is the sub-complex of $\mathcal{C}(S)$ whose vertices are curves that bound disks in both V and \hat{V} (we call these *reducing curves*; the union of the two disks is called a *reducing sphere*). Edges correspond to curves with geometric intersection number 4 (this is the minimal possible intersection since vertices of $\mathcal{R}(V, \hat{V})$ are in particular separating curves on S). See Figure 5 for an example. The complex $\mathcal{R}(V, \hat{V})$ is studied in [Sch04, Akb08].

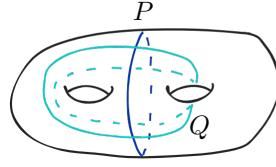


Figure 5: Adjacent vertices of $\mathcal{R}(V, \hat{V})$.

There is an embedding $\mathcal{R}(V, \hat{V}) \hookrightarrow \mathcal{P}(V)$ that appears in [Cho08]. First observe that there is a bijection between vertices of $\mathcal{R}(V, \hat{V})$ and edges of $\mathcal{P}(V)$: given a reducing sphere R , there is a unique pair of disjoint primitive disks $D_1^R, D_2^R \subset V$ that are disjoint from R ; conversely, a pair of disjoint primitive disks D_1, D_2 in V determines a unique reducing curve disjoint from $D_1 \cup D_2$ by [Cho08, Lem. 2.2]. Then there is a map from vertices of $\mathcal{R}(V, \hat{V})$ to $\mathcal{P}(V)$ that sends R to the midpoint of the edge $\{D_1^R, D_2^R\}$. As explained in [Cho08, §6], this map extends to an embedding $\mathcal{R}(V, \hat{V}) \hookrightarrow \mathcal{P}(V)$, where each triangle of $\mathcal{R}(V, \hat{V})$ maps into a unique triangle of $\mathcal{P}(V)$, as pictured in Figure 6.

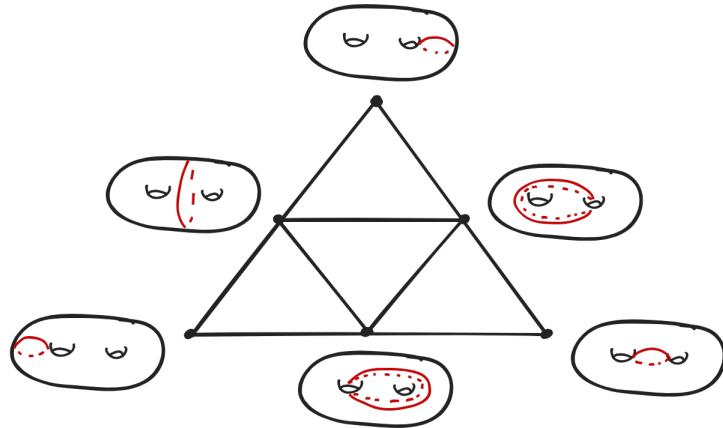


Figure 6: A triangle of $\mathcal{R}(V, \hat{V})$ embedded in a triangle of $\mathcal{P}(V)$.

Remark 2.2. Using the preceding paragraph, we find that $\mathcal{P}(V)$ is obtained from $\mathcal{R}(V, \hat{V})$ by adding, for each primitive disk D , a vertex v_D and the cone $v_D * \mathcal{R}_D$, where $\mathcal{R}_D \subset \mathcal{R}(V, \hat{V})$ is the subgraph spanned by reducing spheres that are disjoint from D .

Heegaard marking complex. We define the *Heegaard marking complex* $\mathcal{M}(V, \hat{V})$ as the graph with

- a vertex for each pair (R, D) , where R is a reducing curve, and $D \subset V$ is a primitive disk that is disjoint from the primitive disks $D', D'' \subset V$ that are disjoint from R .
- an edge between pairs of vertices if they differ by one of the following moves:
 - twist: replace (R, D) by $(R, \beta_R(D))$, where $\beta_R \in \mathbb{G}$ is a half-twist about R (analogous to β in Figure 1).
 - 3-cycle: replace (R, D) by $(\rho(R), \rho(D))$, where $\rho \in \mathbb{G}$ is the order-3 element defined as follows. Let $D', D'' \subset V$ be the two primitive disks that are disjoint from R . Let $\rho \in \mathbb{G}$ be a mapping class that cyclically permutes D, D', D'' and preserves the two components $S \setminus (\partial D \cup \partial D' \cup \partial D'')$ (analogous to δ in Figure 1).

A vertex of $\mathcal{M}(V, \hat{V})$ naturally determines a clean, complete marking in the sense of [MM00, §2.5]. To see this, fix a vertex (R, D) of $\mathcal{M}(V, \hat{V})$. Let $D', D'' \subset V$ and $\hat{D}', \hat{D}'' \subset \hat{V}$ be the primitive disks that are disjoint from R (with the respective pairs D', \hat{D}' and D'', \hat{D}'' dual). Then

$$\mu = \{(R, D), (D', \hat{D}'), (D'', \hat{D}'')\}$$

is a clean, complete marking. We refer to either (R, D) or μ as a *Heegaard marking*.

Recall [MM00, §2.5] that the marking complex $\mathcal{M}(S)$ has vertices for clean, complete markings, and a vertex for markings that differ by a twist or a flip.

Lemma 2.3. *Fix adjacent vertices (R_1, D_1) and (R_2, D_2) of $\mathcal{M}(V, \hat{V})$, and let μ_1, μ_2 be the corresponding vertices of $\mathcal{M}(S)$. Then $d_{\mathcal{M}(S)}(\mu_1, \mu_2) \leq 2$.*

Proof. If (R_1, D_1) and (R_2, D_2) differ by a twist, then μ_1 and μ_2 also differ by a twist, so they are adjacent in $\mathcal{M}(S)$.

Suppose that (R_1, D_1) and (R_2, D_2) differ by a 3-cycle move. Up to homeomorphism, we can assume that the pairs are the ones pictured in Figure 7. Then ρ is the standard generator δ . Figure 8 illustrates a path of length-2 between μ_1 and μ_2 in $\mathcal{M}(S)$. \square

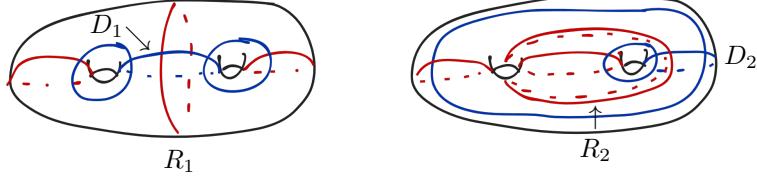


Figure 7: Heegaard markings that differ by a 3-cycle move.

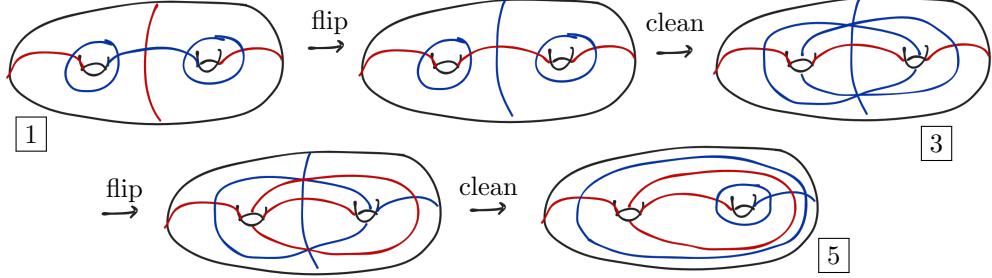


Figure 8: Path $\boxed{1} \rightarrow \boxed{3} \rightarrow \boxed{5}$ of complete, clean markings obtained by doing two flip moves. After performing a flip move, an additional step is needed to replace the resulting marking by a clean marking, c.f. [MM00, §2.5].

Lemma 2.4. \mathbb{G} is quasi-isometric to $\mathcal{M}(V, \widehat{V})$.

Proof. It is easy to see that \mathbb{G} acts transitively on vertices of $\mathcal{M}(V, \widehat{V})$. The stabilizer of the standard pair (P, E_3) (Figure 3 and Figure 5) is the subgroup $\langle \alpha, \gamma \rangle \subset \mathbb{G}$. Since $\langle \alpha, \gamma \rangle \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ is finite, the lemma follows. \square

Goeritz group action. The Goeritz group \mathbb{G} obviously acts on all of the complexes defined above. The following computation will be used in §5.

Proposition 2.5 (Primitive disk stabilizer). *Let $E = E_2$ be the primitive disk pictured in Figure 3. The stabilizer $\mathbb{G}_E < \mathbb{G}$ is the subgroup generated by $\alpha, \beta, \gamma\delta$.*

Proof. The following proof is based on the computation in [Sch04] of a generating set for \mathbb{G} . Let $\mathcal{R}_E \subset \mathcal{R}(V, \widehat{V})$ denote the subcomplex spanned by reducing spheres that are disjoint from E . Let P, Q be the reducing curves pictured in Figure 5.

We prove the proposition by showing that (1) \mathcal{R}_E is connected, (2) \mathbb{G}_E acts transitively on vertices and edges of \mathcal{R}_E , and (3) \mathbb{G}_E contains an edge inversion $P \leftrightarrow Q$. Once we show this, it is a basic result of geometric group theory that \mathbb{G}_E is generated by the stabilizer of P and the edge inversion [FM12, Lem. 4.10]. The stabilizer of P in \mathbb{G} is $\langle \alpha, \beta, \gamma \rangle \cong (\mathbb{Z}_2 \times \mathbb{Z}) \rtimes \mathbb{Z}_2$; see [Sch04, §2]. It is easy to see that the intersection of this group with \mathbb{G}_E is $\langle \alpha, \beta \rangle$. In addition $\gamma\delta$ interchanges P and Q ,

so it only remains to show that \mathcal{R}_E is connected, and that \mathbb{G}_E acts transitively on its vertices and edges.

One can show that \mathcal{R}_E is connected in the same way that Scharlemann shows that $\mathcal{R}(V, \widehat{V})$ is connected. Given $R \in \mathcal{R}_E$, Scharlemann [Sch04, §3] gives a surgery procedure that replaces R by an adjacent curve with fewer intersections with P . This construction does not create intersections with E . (The main input of the surgery operation is an arc of R on one side of P whose slope (defined in [Sch04, §3]) is ∞ and an arc on the other side that has slope 0 and is disjoint from R . By our choice of E , the arc of slope ∞ is not on the same side of P as E . Since E is disjoint from arcs of slope 0, the surgery procedure does not create any new intersections with E . See [Sch04] for details.)

Next we show that \mathbb{G}_E acts transitively on vertices \mathcal{R}_E . Fix a vertex $R \in \mathcal{R}_E$ we show that there is $k \in \mathbb{G}_E$ with $k(P) = R$. By Scharlemann's result, there exists $g \in \mathbb{G}$ with $g(P) = R$. There are two possibilities for $g(E)$ (each of the solid tori components of $V \setminus R$ contains exactly one primitive disk). If $g(E) \neq E$, then $(g \circ \gamma)(E) = E$, and so either $k = g$ or $k = g \circ \gamma$ has the desired properties $k(E) = E$ and $k(P) = R$ (note that $\gamma(P) = P$). The fact that \mathbb{G}_E acts transitively on edges follows by observing that $\beta \in \mathbb{G}_E$ acts transitively on reducing spheres that are adjacent to P in \mathcal{R}_E . \square

To summarize, we have a sequence of forgetful maps

$$\mathcal{M}(V, \widehat{V}) \rightarrow \mathcal{R}(V, \widehat{V}) \rightarrow \mathcal{P}(V),$$

and each is distance non-increasing. Up to quasi-isometry, each map is the inclusion of a space in a coned-off space. Identifying $\mathcal{M}(V, \widehat{V})$ with \mathbb{G} (up to quasi-isometry), the space $\mathcal{R}(V, \widehat{V})$ is quasi-isometric to the cone off of \mathbb{G} by cosets of $\mathbb{G}_P = \langle \alpha, \beta, \gamma \rangle$, and the space $\mathcal{P}(V)$ is quasi-isometric to the cone off of \mathbb{G} by cosets of $\mathbb{G}_E = \langle \alpha, \beta, \gamma\delta \rangle$. Since a finite-index subgroup of \mathbb{G}_P is contained in \mathbb{G}_E , we can also view $\mathcal{P}(V)$ as a cone off of $\mathcal{R}(V, \widehat{V})$, up to quasi-isometry. This is discussed further in §5.

To end this section, we record the following lemma for later use. The proof is elementary.

Lemma 2.6. *Let G be a finitely generated group. Fix complexes \mathcal{X} and \mathcal{Y} with isometric G actions, and assume there is a distance non-increasing G -equivariant map $\pi : \mathcal{X} \rightarrow \mathcal{Y}$. If the orbit map $G \rightarrow \mathcal{Y}$ is a quasi-isometric embedding, then so is the orbit map $G \rightarrow \mathcal{X}$.*

2.3 Surgery of primitive disks

In this section we explain the construction of surgery paths in $\mathcal{P}(V)$.

Surgery. Primitive disks are manipulated by surgery along bigons (boundary compressions). Some care must be taken to ensure that the disks resulting from surgery are primitive. Below we mostly follow the terminology in [MS13, §8].

Fix a primitive disk D . A *surgery bigon* for D is a triple (B, a, b) , where $B \subset V$ is a closed disk whose boundary is decomposed into two arcs a, b with $a = B \cap D$, $b = B \cap S$, and $B \cap \partial D = \partial a = \partial b$. Sometimes we refer to the surgery bigon as B , leaving a and b implicit.

To surger D along (B, a, b) , observe that a decomposes D into closed disks $D = D_1 \cup D_2$ whose intersection is a . The disks obtained from D by surgery along B are $D' = D_1 \cup B$ and $D'' = D_2 \cup B$. After an isotopy, D, D', D'' are pairwise disjoint. Furthermore, D', D'' are not isotopic, since otherwise $\partial D \subset S$ would be separating (hence not primitive).

A surgery bigon (B, a, b) for D is called a *boundary compression* if b is an essential arc in $S \setminus \partial D$. This implies that $\partial D'$ and $\partial D''$ are essential in S .

We call a boundary compression (B, a, b) for D *primitive* if there exists $\widehat{D}_b \subset \widehat{V}$ so that $\partial D, \partial \widehat{D}_b$ are disjoint and $b, \partial \widehat{D}_b$ intersect transversely in a single point. If B is a primitive boundary compression, then D', D'' are both primitive disks; this follows from the definitions. Below we will only ever be interested in primitive boundary compressions.

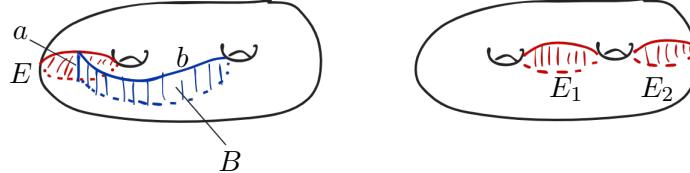


Figure 9: Left: a primitive disk E and a surgery bigon (B, a, b) . Right: the result of surgery is two primitive disks E_1, E_2 .

Example 2.7. The following example gives a basic but important construction of primitive boundary compressions. Fix primitive disks $D, E \subset V$ with $i(\partial D, \partial E) \neq 0$. We assume $\partial D, \partial E$ are in minimal position and D, E intersect transversely. Since V is irreducible, we can isotope D, E so that the 1-manifold $E \cap D$ has no closed components, i.e. each component of $D \cap E$ is a proper arc. An outermost arc a of $D \cap E \subset E$ cuts off a bigon (B, a, b) , which is always a primitive boundary compression for D by [Cho08, Thm. 2.3]. The surgered disks D', D'' satisfy

$$i(\partial D', \partial E) + i(\partial D'', \partial E) < i(\partial D, \partial E),$$

which implies that neither of D', D'' is isotopic to D . Furthermore, neither of D', D'' is isotopic to E because D', D'' are disjoint from D and $i(\partial D, \partial E) \neq 0$.

Remark 2.8. We can repeat the above discussion with a single primitive disk D replaced by a collection $\Delta \subset S$ of $k \geq 1$ disjoint primitive disks. A primitive boundary compression for Δ is a primitive boundary compression B for some component $D \subset \Delta$ such that B is disjoint $\Delta \setminus D$. Surgering D along B results in a collection Δ' of $k+1$ disjoint primitive disks. Furthermore, generalizing Example 2.7, if E is a primitive disk with nonempty minimal intersection with Δ , then an outermost bigon of $E \setminus (\Delta \cap E)$ is a primitive boundary compression for Δ , and $i(\Delta', E) < i(\Delta, E)$.

Subsurfaces and surgery sequences. A subsurface $X \subset S$ is *essential* if every component of ∂X is essential in S . By convention, we will always assume our subsurfaces are connected.

We say a primitive disk D is *supported in X* if $\partial D \subset X$ and ∂D is not parallel to a component of ∂X . More generally, D *cuts X* if ∂D cannot be isotoped to be disjoint from X .

If there is a primitive disk supported in X , we say X is *primitively compressible*. Otherwise, X is *primitively incompressible*. Similarly, if D cuts X and there exists a primitive boundary compression (B, a, b) for D with $b \subset X$, then we say that D is *primitively boundary compressible* into X .

Definition 2.9 (Surgery sequences). Fix a subsurface $X \subset S$ and primitive disk D , and assume $\partial X, \partial D$ are in minimal position and have nonempty intersection. A *primitive compression sequence* is a sequence $\{\Delta_k\}_{k=1}^n$, where (i) $\Delta_1 = \{D\}$, (ii) Δ_{k+1} is a collection of $k+1$ disjoint disks obtained from Δ_k by a primitive boundary compression supported in $S \setminus \partial X$, and (iii) for each k , each $D_k \in \Delta_k$ intersects ∂X nontrivially.

Remark 2.10. If $S \setminus n(\partial X)$ supports a primitive disk E and $\{\Delta_k\}_{k=1}^n$ is a surgery sequence for X , then either (i) E is disjoint from Δ_n or (ii) Δ_n can be primitively boundary compressed into $S \setminus n(\partial X)$; cf. Example 2.7 and Remark 2.8. This motivates the following definition, which agrees with [MS13, Definition 11.2].

Definition 2.11. A primitive compression sequence $\{\Delta_k\}_{k=1}^n$ for X, D is *maximal* if either X supports a primitive disk that is disjoint from Δ_n or there is no primitive boundary compression for Δ_n supported in $S \setminus n(\partial X)$. In the former case, the surgery sequence is said to *end in $S \setminus n(\partial X)$* and in the latter case it is said to *end essentially*.

Whether or not a maximal primitive compression sequence for (X, D) ends essentially or ends in $S \setminus n(\partial X)$ depends on whether or not $S \setminus \partial X$ is primitively compressible. By Remark 2.10, if $S \setminus n(\partial X)$ is primitively compressible, then every maximal surgery sequence ends in $S \setminus n(\partial X)$, and the converse is also true, trivially. Then if $S \setminus \partial X$ is primitively incompressible, and $\{\Delta_k\}_{k=1}^n$ is a maximal primitive

compression sequence, then Δ_n cannot be primitively boundary compressed into $S \setminus n(\partial X)$.

Any surgery sequence can be extended to a maximal sequence. If $\{\Delta_k\}_{k=1}^n$ is not maximal then, there exists a primitive boundary compression for Δ_n supported in $S \setminus n(\partial X)$, which produces a collection Δ_{n+1} . Since the given sequence is not maximal, every disk in Δ_{n+1} intersects ∂X , which extends the sequence. Since each $D_k \in \Delta_k$ intersects ∂X and the total intersection $i(\Delta_k, \partial X)$ is bounded above by $i(\partial D, \partial X)$, every surgery sequence is finite and can be extended to a maximal sequence.

General position. For various arguments it is helpful to put curves on S or disks in V in a “general position”. Precise notions follow.

Recall that simple closed curves $a, b \subset S$ are said to be in *minimal position* if they meet transversely and $S \setminus (a \cup b)$ has no bigon components. Equivalently, a, b realize the geometric intersection number for the corresponding isotopy classes [FM12, §1.2].

Disk $D, E \subset V$ that intersect transversely can be isotoped so that $D \cap E$ is a union of arcs. This is because V is irreducible: if $c \subset D \cap E$ is a circle, then c bounds a disk in both D and E , giving a 2-sphere in V that can be filled by a 3-ball; using these 3-balls, circular components of $D \cap E$ can be removed by a sequence of isotopies. Then we say that $D \cap E$ is *circle free*.

If given simple closed curves $a, b \subset S$ and a subsurface $X \subset S$, we can isotope a, b so that $a, b, \partial X$ are (pairwise) in minimal position. Then $S \setminus (a \cup b \cup \partial X)$ generally will contain triangular components, and we can always isotope further so that these triangle components are all in X or all in $S \setminus X$ (the same observation appears in [MS13, §10]).

Combining the preceding paragraphs, we say D, E, X are in *standard position* if (1) each pair of $\partial D, \partial E, \partial X$ are in minimal position, (2) the triangles of $S \setminus (\partial D \cup \partial E \cup \partial X)$ are either all in X or all in $S \setminus X$, and (3) $D \cap E$ is circle free. Any triple D, E, X can be put in standard position.

2.4 Subsurface projection

In this section we recall the terminology needed to state the distance formula.

Subsurface projections, holes, and diameter. We use subsurface projections to understand the geometry of $\mathcal{P}(V)$. For a subsurface $X \subset S$ and a curve $a \subset S$, the *subsurface projection* $\pi_X(a) \subset \mathcal{C}(X)$ is defined as the union of nontrivial isotopy classes represented by the boundary of $n((a \cap X) \cup \partial X)$. See Figure 10 for an example.

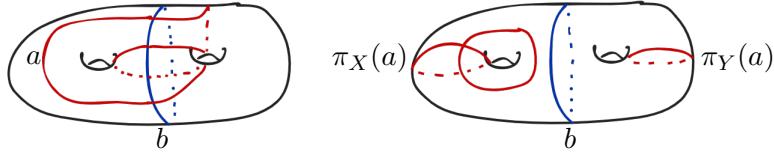


Figure 10: A curve a and its projections to the subsurfaces X and Y on the left and right sides of the curve b .

For curves $a, b \subset S$, the *subsurface-projection distance* is defined as

$$d_X(a, b) = \text{diam}_{\mathcal{C}(X)}(\pi_X(a) \cup \pi_X(b)).$$

For $X \cong \Sigma_{0,2}$ it is useful to (re)define $\mathcal{C}(X)$ to be the arc complex. This requires a different definition of the subsurface projection; see [MM00, §2.4] or [MS13, §4.1].

The function d_X is not a distance function, since it isn't necessarily true that $d_X(a, a) = 0$; on the other hand, d_X is symmetric and satisfies the triangle inequality. If d_X is bounded on $\mathcal{P}(V)$, then its maximum value is called the *diameter* of X (with respect to $\mathcal{P}(V)$). If d_X is unbounded, we say X has *infinite diameter*.

One says X is *hole* for $\mathcal{P}(V)$ if every primitive disk $D \subset V$ cuts X .

Remark 2.12. One can similarly define holes for the disk complex $\mathcal{D}(V)$ and the diameter for $X \subset S$ with respect to $\mathcal{D}(V)$. Since $\mathcal{P}(V) \subset \mathcal{D}(V)$, the following statements follow from the definitions.

- If X is a hole for $\mathcal{D}(V)$, then X is a hole for $\mathcal{P}(V)$.
- If X has infinite diameter with respect to $\mathcal{P}(V)$, then X has infinite diameter with respect to $\mathcal{D}(V)$.

The converses of these statements are not true. For example, if $D \in \mathcal{D}(V)$ is not primitive and ∂D is nonseparating, then $X = S \setminus n(\partial D)$ is a hole for $\mathcal{P}(V)$, but not a hole for $\mathcal{D}(V)$. Regarding the diameter, there exists a genus-1 surface $X \subset S$ that is a Seifert surface for an embedding of the trefoil knot, and X has infinite diameter with respect to $\mathcal{D}(V)$, but finite diameter with respect to $\mathcal{P}(V)$ (see §3.4).

Subsurface projection of a surgery sequence. For the disks in a primitive compression sequence for (X, D) , the subsurface projections to X do not change much. By [MS13, Lem. 11.5], if $\{\Delta_k\}$ is a compression sequence for (X, D) , then there exists $D_n \in \Delta_n$ and arcs $a \subset \partial D \cap X$ and $b \subset \partial D_n \cap X$ so that a and b are

disjoint.¹ As a consequence, one obtains the following result, proved in [MS13, Lem. 11.7].

Lemma 2.13. *Fix a subsurface $X \subset S$, and a primitive disk D that cuts X . Then there exists a primitive disk E such that $d_X(D, E) \leq 6$ and either E supported in X (if X is primitively compressible) or E cannot be primitively compressed into $S \setminus n(\partial X)$ (if X is primitively incompressible).*

2.5 I -bundles

An I -bundle $T \xrightarrow{p} B$ is a fiber bundle with fiber $I = [0, 1]$. In this paper B will always be a compact surface B . The *horizontal boundary* $\partial_h T \subset T$ is the union of endpoints of each fiber, i.e. it is the total space of the induced bundle with fiber ∂I ; the restriction $p : \partial_h T \rightarrow B$ is a 2-fold covering map. The *vertical boundary* $\partial_v T$ is defined as $p^{-1}(\partial B)$. A subset of T is called *vertical* if it is a union of fibers.

If T is orientable and $\partial_h T$ is connected, then the base B is non-orientable. Indeed, if $\partial_h T$ is connected, then there exists a curve $c \subset B$ so that $p^{-1}(c)$ is a Möbius band. In order for T to be orientable, c must be 1-sided. From this it follows that $\partial_h T \rightarrow B$ is the orientation double cover of B .

For example, if $\partial_h T = \Sigma_{0,2n}$, then $B = \mathbb{R}P^2 \setminus (\bigcup_n D^2)$, and if $\partial_h T = \Sigma_{1,2n}$, then $B = (\mathbb{R}P^2 \# \mathbb{R}P^2) \setminus (\bigcup_n D^2)$.

2.6 Fibered knots and Heegaard splittings

Lemma 2.14. *Let $L \subset S^3$ be a fibered link*

$$\mathring{F} \rightarrow S^3 \setminus L \xrightarrow{\pi} S^1.$$

The union of two fibers with L is a Heegaard surface.

Proof. Identify S^1 with \mathbb{R}/\mathbb{Z} . For $\theta \in S^1$, set $\mathring{F}_\theta = \pi^{-1}(\theta)$, and denote $F_\theta = \mathring{F}_\theta \cup L$ (a compact surface with boundary L). We show that

$$\Sigma = F_0 \cup F_{1/2}$$

is a Heegaard surface.

Consider the interval $J = [0, 1/2] \subset S^1$. Then $\pi^{-1}(J) \cong \mathring{F} \times J$. We want to show that $W := (\mathring{F} \times J) \cup K$ is a handlebody.

¹The statement of [MS13, Lem. 11.5] includes the hypothesis that the sequence is maximal, but this is not used in the proof. This is noteworthy because our surgery sequences can be maximal as sequences of primitive disks without being maximal as a sequence of all disks.

First observe that $F \times J$ is a handlebody. Choose a maximal collection of disjoint essential arcs $a_1, \dots, a_m \subset F$ whose complement is a disk. Then $a_1 \times J, \dots, a_m \times J$ is a collection of disjoint disks whose complement in $F \times J$ is a 3-ball.

Next observe that W is homeomorphic to $F \times J \cup C(q)$, where $C(q)$ is the mapping cylinder of the projection $q : \partial F \times J \rightarrow \partial F$. Observe that $C(q)$ is homeomorphic to $\partial F \times C(q')$, where $q' : J \rightarrow \{\ast\}$ (constant map). It is easy to show that $C(q')$ is homeomorphic to $J \times [0, 1]$, and it follows that $W \cong F \times J \cup C(q)$ is homeomorphic to $F \times J$.

The argument works the same for $J = [1/2, 1]$. Therefore, $\Sigma \subset S^3$ is a Heegaard surface. \square

Remark 2.15 (Monodromy). Using Lemma 2.14, we obtain a description for the gluing map of the Heegaard splitting associated to a fibered link. This is pictured schematically in Figure 11; see also Figure 12. Here $\phi \in \text{Mod}(F)$ is a lift of the monodromy of the fibration $\mathring{F} \rightarrow S^3 \setminus L \rightarrow S^1$ with respect to the short exact sequence

$$1 \rightarrow \mathbb{Z}^{|L|} \rightarrow \text{Mod}(F) \rightarrow \text{Mod}(\mathring{F}) \rightarrow 1.$$

Any lift defines a Heegaard splitting, but changing ϕ by an element of the kernel (a boundary twist) may result in a different manifold ($\neq S^3$). This is discussed more in the Appendix, where we give a proof of the classification of genus-1 fibered knots.

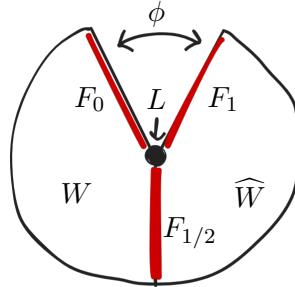


Figure 11: M_ϕ is obtained by gluing two handlebodies $F \times I$ and $\widehat{F} \times I$.

A converse. Lemma 2.14 and Remark 2.15 have an obvious converse. Fix two copies of a compact, oriented surface $F = \widehat{F}$. Let M be a 3-manifold with a decomposition $M = (F \times I) \cup (\widehat{F} \times I)$, where the gluing $h : \partial(F \times I) \rightarrow \partial(\widehat{F} \times I)$ matches the horizontal boundary components $h(F \times \{j\}) = \widehat{F} \times \{j\}$ for $j = 0, 1$. This is a Heegaard splitting and $\partial F \times \{1/2\}$ is a fibered link in M .

The homeomorphism type of M is determined by the gluing map h and more precisely by a single mapping class $[\phi] \in \text{Mod}(F)$. After an isotopy of h , we can

assume that $h|_{(\partial F) \times I}$ is the identity. Define a homeomorphism $\phi : F \rightarrow F$ by the composition

$$\phi : F = F \times 0 \cong F \times 1 \xrightarrow{h} \widehat{F} \times 1 \cong \widehat{F} \times 0 \xrightarrow{h^{-1}} F \times 0 = F.$$

The homeomorphism type of M depends only on the isotopy class of ϕ , and we denote $M = M_\phi$.

Remark 2.16. Since we have a fixed identification $F = \widehat{F}$, we can (and will for convenience) choose h so that $h|_{F \times 0} = \text{Id}$. Then $\phi = h|_{F \times 1}$. See Figure 12.

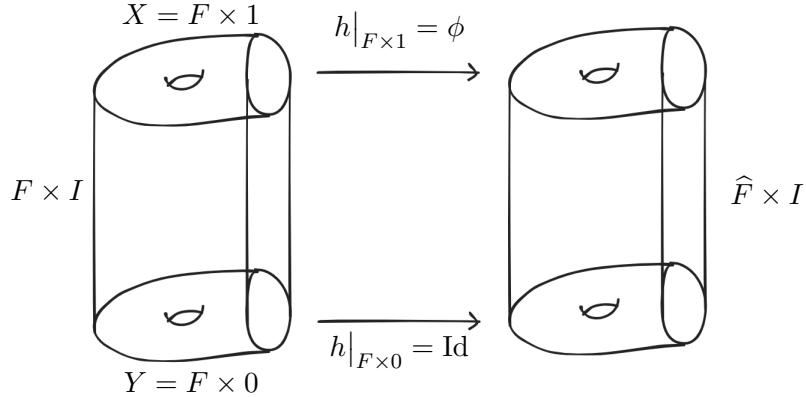


Figure 12: M_ϕ is obtained by gluing two handlebodies $F \times I$ and $\widehat{F} \times I$.

Remark 2.17 (Orientations). To view M as an oriented 3-manifold, fix an orientation on $F = \widehat{F}$, orient $F \times I$ so that the positive normal direction to $F \times \{1/2\}$ points toward $F \times \{1\}$, and orient $\widehat{F} \times I$ in the opposite way, so that the positive normal direction to $\widehat{F} \times \{1/2\}$ points toward $\widehat{F} \times \{0\}$. With this choice, h is orientation reversing, so M inherits its orientation from $F \times I$ and $\widehat{F} \times I$.

Remark 2.18 (Fiber-preserving Goeritz elements). If $M_\phi = (F \times I) \cup (\widehat{F} \times I)$ as above, and $\tilde{\phi} \in \text{Homeo}(F, \partial F)$ represents $\phi \in \text{Mod}(F)$, then the product homeomorphism $\phi \times \text{Id}$ of $F \times I$ and $\widehat{F} \times I$ induces a homeomorphism of M_ϕ that preserves the Heegaard splitting. By construction, the isotopy class of this homeomorphism restricted to $F \times 1$ is ϕ .

Proposition 2.19 (Fiber-preserving homeomorphisms). *Let F be a compact surface, and fix $M = (F \times I) \cup_\phi (\widehat{F} \times I)$ as above. Fix an orientation-preserving homeomorphisms $g : M \rightarrow M$ that preserves this Heegaard splitting. If g preserves $F \times 0 \cup F \times 1$, then g is isotopic, through homeomorphisms preserving the Heegaard splitting, to a homeomorphism whose restriction to each of $F \times I$ and $\widehat{F} \times I$ is fiber preserving.*

Proof. There are two cases: either the components of $F \times 0 \cup F \times 1$ are preserved, or they are interchanged. We explain the case each component is preserved; the other case is similar.

Since a homeomorphism of a handlebody that is the identity on the boundary is isotopic to the identity [Hat76, §1], it suffices to show that there exists a fiber-preserving homeomorphism f so that $f^{-1} \circ g|_{\Sigma}$ is isotopic to the identity, where $\Sigma = \partial(F \times I)$.

Set $g_i = g|_{F \times \{i\}}$ for $i = 0, 1$. Observe that g_0 and g_1 are isotopic homeomorphisms of F , since they induce the same (outer) automorphism of $\pi_1(F)$ (identifying $\pi_1(F \times 1) = \pi_1(F \times I) = \pi_1(F \times 0)$). Note that this isotopy is not necessarily the identity on ∂F . In any case, by extension we can isotope g to a homeomorphism (still denoted g) satisfying $g_0 = g_1$.

Let $f = g_0 \times \text{Id}$ on $F \times I$ and $\widehat{F} \times I$. By construction, $f^{-1} \circ g$ is the identity on $F \times \{0, 1\}$, so $f^{-1} \circ g|_{\Sigma}$ is supported on $(\partial F) \times I$. Then the isotopy class of $f^{-1} \circ g|_{\Sigma}$ is a multitwist.

By [Oer02, Thm. 1.11], a multitwist about $c \subset \Sigma$ extends to a handlebody only if the curves in c bound a union of essential disks and annuli. Since no curve in $c = \partial F \times \{1/2\}$ bounds a disk in $F \times I$, and no two curves of c bound an annulus in $F \times I$, we conclude that $f^{-1} \circ g|_{\Sigma}$ is isotopic to the identity. \square

Remark 2.20. Assume that f is fiber preserving on $F \times I$ and $\widehat{F} \times I$. If f preserves the components of $F \times 0 \cup F \times 1$, then we can write $f = \psi \times \text{Id}$ on $F \times I$ and $f = \widehat{\psi} \times \text{Id}$ on $\widehat{F} \times I$, and the gluing $h : \partial(F \times I) \xrightarrow{\cong} \partial(\widehat{F} \times I)$ forces the relation $\psi = \widehat{\psi}$, and $\psi \phi = \phi \psi$ in $\text{Mod}(F)$. Together these imply that $\psi \phi \psi^{-1} = \phi$.

Suppose instead that f interchanges the components of $F \times 0 \cup F \times 1$. Write $f(x, t) = (\psi(x), 1-t)$ on $F \times I$ and $f(x, t) = (\widehat{\psi}(x), 1-t)$ on $\widehat{F} \times I$. Here the gluing forces the relations $\psi = \widehat{\psi} \phi$ and $\widehat{\psi} = \phi \psi$. Together these imply $\psi \phi \psi^{-1} = \phi^{-1}$.

We also observe that the homeomorphism $(x, t) \leftrightarrow (x, 1-t)$ on $F \times I$ and $\widehat{F} \times I$ defines an orientation-reversing homeomorphism between M_ϕ and M_ϕ^{-1} .

2.7 Heegaard splittings and Mayer–Vietoris

Here we make a simple observation that will be useful in multiple arguments.

Lemma 2.21. *Let M be a 3-manifold with a Heegaard splitting $M = W \cup_{\Sigma} \widehat{W}$. Suppose there exists a multicurve $c \subset \Sigma$ that bounds a surface in W and \widehat{W} . If c can be oriented so that the associated homology class $[c] \in H_1(\Sigma)$ is nonzero, then $H_2(M) \neq 0$.*

Proof. The Mayer–Vietoris sequence associated to the Heegaard splitting contains

the exact sequence

$$H_2(M) \xrightarrow{d} H_1(\Sigma) \xrightarrow{j} H_1(W) \oplus H_1(\widehat{W})$$

By assumption there is a nonzero class $H_1(\Sigma)$ in $\ker(j) = \text{Im}(d)$, so $H_2(M) \neq 0$. \square

3 Classification of holes for $\mathcal{P}(V)$

This section is devoted to the proof of Theorem D. Our analysis is based on work of Masur–Schleimer [MS13, §9-12], who characterize large-diameter holes for the disk complex $\mathcal{D}(V)$. Theorem D does not follow easily from [MS13], but we do use several of the same ideas.

Overview of the proof. Let $X \subset S$ be an essential subsurface, and assume X is a hole for $\mathcal{P}(V)$. It is easy to show that if $X = S$, then X supports a pseudo-Anosov in \mathbb{G} (see Figure 13), so X has infinite diameter. Therefore, we assume that $X \subsetneq S$ is a proper subsurface. We divide into three cases: $X \cong \Sigma_{0,2}$ is an annulus, $X \neq \Sigma_{0,2}$ is primitively compressible, and $X \neq \Sigma_{0,2}$ is primitively incompressible. In the first two cases, we will show that every hole X has diameter ≤ 13 (Thm. 3.1 and Prop. 3.3). The last case is the most complicated and interesting. When $X \neq \Sigma_{0,2}$ is primitively incompressible with diameter ≥ 61 , we construct homeomorphisms

$$V \cong \Sigma_{1,1} \times I \cong \widehat{V}$$

such that X is a component of the horizontal boundary of each I -bundle (Theorem 3.5). In particular, this gives a fibering

$$X \setminus \partial X \rightarrow S^3 \setminus \partial X \rightarrow S^1. \quad (4)$$

Therefore, ∂X is a genus-1 fibered knot. It is well known that this implies that ∂X is either the figure-8 knot or the right/left-handed trefoil knot (this is discussed more in the Appendix). It is easy to show that X has infinite diameter when ∂X is the figure-8 knot (since this knot has hyperbolic monodromy, c.f. Remark 2.18). We show that if ∂X is a trefoil knot, then X has finite diameter (§3.4). Altogether this proves Theorem D.

As part of our analysis, we give a concrete description of the set of primitive disks that are vertical with respect to $V \cong \Sigma_{1,1} \times I$ (in each case), and we give a precise description of the image of the subsurface projection of $\mathcal{P}(V)$ to the horizontal boundary components (§3.5). This is needed for the proof of the distance formula (Theorem C). We also determine the subgroup of \mathbb{G} that preserves the fibering (4) when ∂X is the figure-8 knot (§3.6). This is used for the proof of Theorem B.

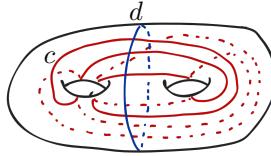


Figure 13: The curves c, d fill S , so by a result of Thurston, the mapping class $T_c T_d^{-1}$ is pseudo-Anosov. Since c, d are reducing sphere curves, $T_c, T_d \in \mathbb{G}$.

3.1 The hole X is an annulus

Theorem 3.1 (Annular holes for $\mathcal{P}(V)$). *Let $X \subset S$ be an essential annulus. If X is a hole, then the diameter of X is at most 11.*

Theorem 3.1 is similar to [MS13, Thm. 10.1], but there is one ingredient from their argument that doesn't work: the first Claim following Theorem 10.1, which asserts that if $X = n(c)$ is an annular hole for $\mathcal{D}(V)$, then $i(\partial D, c) \geq 2$ for each $D \in \mathcal{D}(V)$. Figure 14 shows that this statement is not true if $\mathcal{D}(V)$ is replaced by $\mathcal{P}(V)$. Although we cannot use this claim, we will use the same overall strategy as the proof of [MS13, Thm. 10.1] to prove Theorem 3.1.

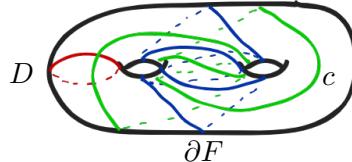


Figure 14: The annulus $X = n(c)$ is a hole for $\mathcal{P}(V)$, and D is a primitive disk with $i(\partial D, c) = 1$. To see that X is a hole, if E is a primitive disk disjoint from c , then (by surgery) there exists primitive disk E' that is disjoint from $D \cup c$. But there is a unique disk $F \subset V$ with ∂F disjoint from $\partial D \cup c$, and this disk is not primitive.

Proof of Theorem 3.1. Fix an essential annulus $X = n(c) \subset S$, and assume that X is a hole for $\mathcal{P}(V)$. Choose primitive $D \subset V$ minimizing $i(\partial D, c)$ among all primitive disks (equivalently, D minimizes the number of components of $\partial D \cap X$). Assume for a contradiction that the diameter of X is at least 12. Then there exists primitive $E \subset V$ so that $d_X(D, E) \geq 6$.

Put D, E, X in standard position (cf. §2.3) with any triangular components of $S \setminus (\partial D \cup \partial E \cup \partial X)$ contained in X . Let $B \subset E$ be an outermost bigon cut by $D \cap E$. Denote $a := B \cap D$ and $b := B \cap \partial E$, and let D', D'' be the disks obtained by surgering D along B ; cf. Example 2.7.

By the Lemma 3.2 below, every component of $\partial D \cap X$ intersects every component of $\partial E \cap X$.

Lemma 3.2. Fix vertices $v, w \in \mathcal{C}(S)$ that cut the annulus $X \subset S$. Choose representatives so that v, w, X are in standard position with triangular components of $S \setminus (v \cup w \cup \partial X)$ contained in X . If $d_X(v, w) \geq 6$, then every component of $v \cap X$ intersects every component of $w \cap X$.

Lemma 3.2 is the second claim after Theorem 10.1 in [MS13]. The statement above is more general than the statement in [MS13], but their proof holds in this generality. We apply the lemma to with $(v, w) = (\partial D, \partial E)$.

The proof proceeds by studying the relationship between the arc b and the arcs $X \cap \partial E$ in ∂E . First note that no component of $X \cap \partial E$ can be strictly contained in b , since then this component would be disjoint from $\partial D \cap X$, contradicting Lemma 3.2. Now we consider cases based on whether the endpoints of b meet X , and how many components of X they meet. These cases are pictured in [MS13, Fig. 10.4] and listed below. Ultimately, we will show that each of these cases is impossible, since otherwise either D and E are not in minimal position or at least one of D', D'' would intersect X in fewer components than D (contradicting our assumption). From this we deduce that there is no pair D, E with $d_X(D, E) \geq 6$, so the diameter of X is at most 11.

The only case that is treated differently from [MS13] is Case 4 below. We include the other cases for completeness.

Case 1: the arc b is disjoint from X . In this case, each component of $D \cap X$ contributes exactly one component to either $D' \cap X$ or $D'' \cap X$ (not both). It follows then that one or both of D', D'' intersect X in fewer components than D , a contradiction. This is illustrated in Figure 15.

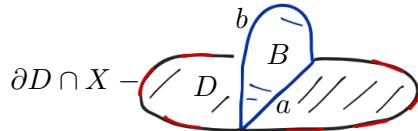


Figure 15: The disk D , the bigon B , and the arcs of $\partial D \cap X$. In the figure, b is disjoint from X , and one observes that $\partial D'$ and $\partial D''$ have fewer intersections with X than ∂D does.

Case 2: the arc b is contained in X . In this case, either the endpoints of b meet (i) the same or (ii) different components of $\partial D \cap X$. In case (i), let d denote the component of $D \cap X$ meeting b . Then the complement of $d \cup b$ in X contains a bigon, which contradicts the assumption that D, E intersect minimally. In case (ii), observe that if ∂D is oriented, then the two intersections of b with ∂D appear on

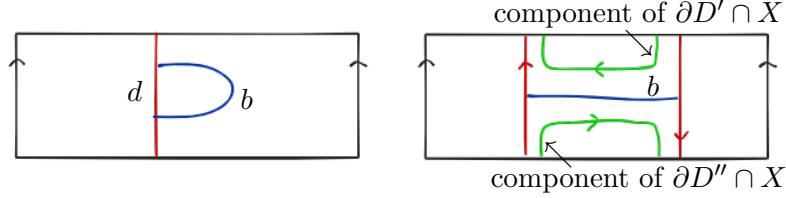


Figure 16: The arc b is contained in X and connects either a component of $\partial D \cap X$ to itself (left) or two distinct components of $\partial D \cap X$ (right).

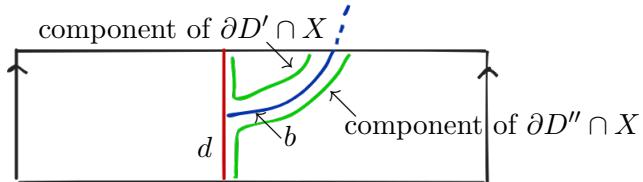


Figure 17: The arc b has one endpoint in X .

the same side of the normal bundle of ∂D in S because b is an arc of a surgery bigon. Then we conclude that there is a component of each of $D' \cap X$ and $D'' \cap X$ whose endpoints lie on the same component of ∂X , so after an isotopy D' and D'' intersects X in fewer components than D , a contradiction. This case is illustrated in Figure 16.

Case 3: exactly one endpoint of b is contained in X . Let d be the component of $D \cap X$ that meets b . The endpoints of d meet different components of X (otherwise ∂D and X do not intersect minimally). Then $b \cap \partial X$ (which is a single point) belongs to the same component as one endpoint of d . This implies that the surgered disk that contains this endpoint (either D' or D'') can be isotoped to intersect X in fewer components than D . See Figure 17.

Case 4: both endpoints of b are contained in X , but $b \not\subset X$. As in Case 2, we divide into two additional subcases: either the endpoints of b meet (i) the same or (ii) different components of $D \cap X$.

In case (i), let d be the component of $D \cap X$ that meets b . Let b_{\pm} and d_{\pm} denote the endpoints of b and d , labeled so that d is a union of intervals $[d_-, b_-] \cup [b_-, b_+] \cup [b_+, d_+]$. We also denote $a_{\pm} \in b$ the point of $b \cap \partial X$ belonging to the same component of $b \cap X$ as b_{\pm} . As previously observed, d_+ and d_- lie on different components of ∂X , and if we orient d , then the intersections of b with d both occur on the same side of d . See Figure 18.

Let D'' be the surgered disk that intersects the interval $[b_-, b_+] \subset \delta$. Consider further

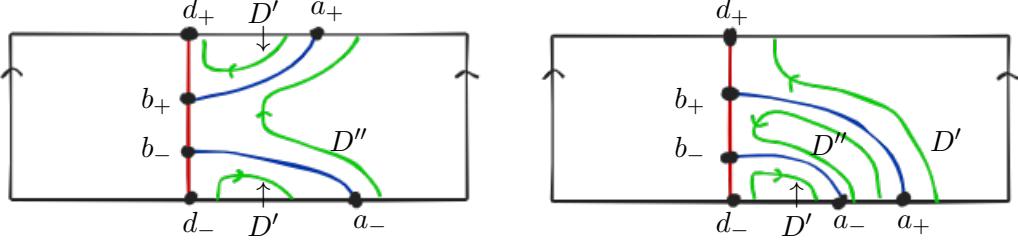


Figure 18: Left: a_+, d_+ lie on the same component of ∂X , and the same holds for a_-, d_- . Right: a_+, d_+ lie on different components of ∂X , and a_-, d_- lie on the same component of ∂X .

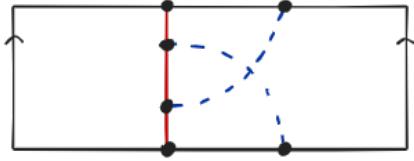


Figure 19: a_+, d_+ lie on different components of ∂X and the same holds for a_-, d_- .

cases, depending on whether a_+ and d_+ belong to the same or different components of ∂X , and similarly for a_- and d_- . If a_+, d_+ lie on the same component of ∂X and the same for a_-, d_- , then D' intersects X in fewer components than D (after an isotopy). If one pair a_+, d_+ or a_-, d_- lie on the same component, and the other pair lies on different components, then D'' intersects X in fewer components than D (after an isotopy). These cases are pictured in Figure 18. Finally, assume that a_+, d_+ lie on different components of ∂X and the same for a_-, d_- . In this situation, the arcs $[b_-, a_-]$ and $[b_+, a_+]$ in the disk $X \setminus d$ have endpoints that link on the boundary. This forces these subarcs of b to intersect, a contradiction. See Figure 19. This concludes case 4(i).

Finally we consider case 4(ii). In this case, D intersects X in exactly 2 components, since otherwise D' or D'' would have smaller intersection with X than D . Let d_\pm denote the two components of $D \cap X$ meeting b . Let d'_\pm be the endpoint of d_\pm that belongs to D' , and define d''_\pm similarly. Define a_\pm, b_\pm as in case 4(i) with $[a_\pm, b_\pm]$ a subinterval of b with $b_\pm \in d_\pm$ and $a_\pm \in \partial X$. Since the endpoints d'_+, d''_+ of d_+ belong to different components of ∂X , exactly one endpoint lies in the same component as a_+ . If a_+ (resp. a_-) lies on the same component of ∂X as d'_+ (resp. d'_-), then D' will be disjoint from X (after isotopy); we conclude similarly when a_+, a_- lie on the same component of ∂X as d''_+, d''_- . If a_+ (resp. a_-) lies on the same component of ∂X as d'_+ (resp. d''_-), then $|D' \cap X| = |D'' \cap X| = 1 < |D \cap X| = 2$, contradicting the fact that D is minimal. The remaining case $(a_+, a_-) \leftrightarrow (d''_+, d'_-)$ is similar. These different cases are pictured in Figure 20.

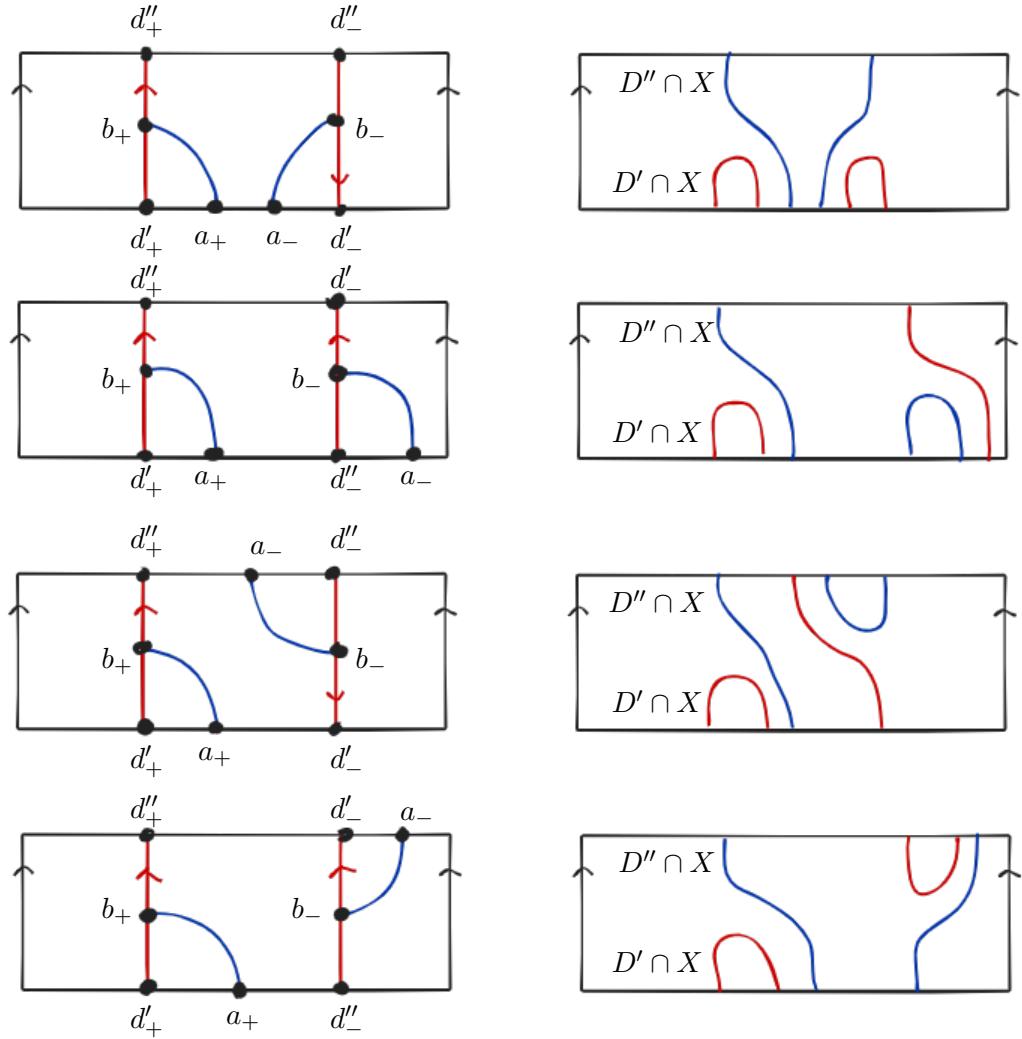


Figure 20: Left: The entire intersection of ∂D and b with X in the various cases, depending on whether a_{\pm} lies on the same component as d'_{\pm} or d''_{\pm} . Right: The entire intersection of $\partial D'$ and $\partial D''$ with X , which illustrates that (after an isotopy) D', D'' intersect X in fewer components than D does.

This completes the proof the proof of Theorem 3.1. \square

3.2 The hole X is primitively compressible

Recall from §2 that $X \subset S$ is primitively compressible if there exists a primitive disk $E \subset V$ such that $\partial E \subset X$ is nonperipheral.

Proposition 3.3 (Primitively compressible holes). *Let $X \subset S$ be a non-annular subsurface that is primitively compressible. If X is a hole for $\mathcal{P}(V)$, then either $X = S$ or X has diameter ≤ 13 .*

Proof of Proposition 3.3. Fix X as in the statement, and suppose that there exist primitive disks D_1, E_1 with $d_X(D_1, E_1) \geq 14$. We will show $X = S$, proceeding by contradiction. Since $X \neq \Sigma_{0,2}$ (by assumption) and $X \neq \Sigma_{0,3}$ (because $\mathcal{C}(\Sigma_{0,3})$ is empty), X is homeomorphic to one of $\Sigma_{0,4}, \Sigma_{1,1}$, and $\Sigma_{1,2}$.

By Lemma 2.13, there exist disks D, E supported in X with $d_X(D, E), d_X(E, D) \leq 6$. Then $d_X(D, E) \geq 2$ by the triangle inequality. Since $\partial D, \partial E$ are vertices of $\mathcal{C}(X)$, this implies that D and E have nonzero intersection number. After an isotopy, we may assume that $D \cap E$ is circle free (c.f. §2.3).

Consider an outermost bigon (B, a, b) cut off from E by $D \cap E$, and the surgered disks D', D'' . The disks D, D', D'' are disjoint, and no two are isotopic (Example 2.7). Furthermore, $\partial D, \partial D', \partial D''$ bound a subsurface $Y \cong \Sigma_{0,3}$ that is contained in X . Since X is homeomorphic to one of $\Sigma_{0,4}, \Sigma_{1,1}$, or $\Sigma_{1,2}$, one of the boundary component of Y is parallel to a boundary component of X . This means the corresponding disk (D' or D'') can be isotoped to be disjoint from X , contradicting the assumption that X is a hole. \square

Remark 3.4. By the proof of Proposition 3.3, if $X \subset S$ is a hole and there exists primitive disks D, E supported in X such that $i(\partial D, \partial E) \neq 0$, then there exist a triple (D, D', D'') of disjoint disks supported in X , and this implies that $X = S$. We will use this observation later.

3.3 The hole X is primitively incompressible

In this section we prove that if X is primitively incompressible with sufficiently large diameter, then X is a genus-1 Seifert surface for a fibered knot $K \subset S \subset S^3$. This is the content Theorem 3.5, although we formulate the theorem slightly differently.

Theorem 3.5 (Primitively incompressible holes). *If $X \subset S$ is a primitively incompressible hole for $\mathcal{P}(V)$ with diameter ≥ 61 , then V and \widehat{V} are homeomorphic to the I -bundle $\Sigma_{1,1} \times I$ with X a component of the horizontal boundary of each.*

Theorem 3.5 is similar to [MS13, Thm. 12.1], which we state below in a special case.

Theorem 3.6 (Masur–Schleimer). *Suppose X is an incompressible hole for $\mathcal{D}(V)$ with diameter ≥ 57 . Then there is an I -bundle T with an embedding $T \hookrightarrow V$ so that $\partial_h T \subset S$ and with X a component of $\partial_h T$.*

We cannot use Theorem 3.6 directly to prove Theorem 3.5 since a primitively incompressible hole is not necessarily incompressible. However, we will use some arguments from [MS13, Thm. 12.1] in our proof.

Proof of Theorem 3.5. Fix X as in the statement of the Theorem. We divide the proof into several steps:

1. Assuming that X has diameter ≥ 57 with respect to $\mathcal{P}(V)$, we build an embedded I -bundle $T \hookrightarrow V$ that contains X as a component of its horizontal boundary. For this we follow the argument of [MS13, §12] as far as possible. We show that either $T \hookrightarrow V$ is (isotopic to) a homeomorphism or every component of ∂X is nullhomologous in V .
2. We show that if X has diameter ≥ 61 , then X is also a primitively incompressible hole for $\mathcal{P}(\widehat{V})$ with diameter ≥ 57 . Our proof of this relies on details of the proof of Step 1.
3. Combining Steps 1 and 2, we have embedded I -bundles $T \hookrightarrow V$ and $\widehat{T} \hookrightarrow \widehat{V}$, each with X as a component of the horizontal boundary. We use the topology of $S^3 = V \cup \widehat{V}$ to rule out all cases where either $T \hookrightarrow V$ or $\widehat{T} \hookrightarrow \widehat{V}$ is not a homeomorphism, and from this we conclude.

Step 1. In this step we construct an embedded I -bundle $T \hookrightarrow V$. The assumption needed in the construction is that X is a primitively incompressible hole for $\mathcal{P}(V)$ with diameter ≥ 57 . Part of the construction follows [MS13, §12] and we refer the reader there for more details. We mostly use their notation. We also remark that a similar construction appeared in earlier work of Oertel; cf. [Oer02, Lem. 2.13].

Fix primitive disks $D_0, E_0 \subset V$ so that $d_X(D_0, E_0) \geq 57$. By Lemma 2.13, we can find primitive disks D, E that cannot be primitively boundary compressed into $S \setminus n(\partial X)$ and with $d_X(D, E) \geq 45$.

We put D, E, X in standard position with triangular components of $S \setminus (\partial D \cup \partial E \cup \partial X)$ contained in $S \setminus X$ (c.f. §2.3).

Give D the structure of a polygon with vertices $\partial D \cap \partial X$, and similarly for E . The edges of these polygons alternate between X and $Y := \overline{S \setminus X}$. Set $\Gamma = D \cap E$. The assumption that D, E cannot be primitively boundary compressed into $S \setminus n(\partial X)$ implies that every arc in Γ is a *diagonal*, i.e. it connects a pair of distinct sides.

(This argument is the same as [MS13, first claim of §12.5], although our assumption is slightly different.)

A side with not many different diagonals. A counting argument shows that there are sides $a \subset \partial D \cap X$ and $b \subset \partial E \cap X$ meeting at most 8 different types of diagonals (two diagonals have the same type if they are parallel, i.e. their endpoints share the same sides). See [MS13, Lem. 12.4].

Arcs and rectangles. We decompose a into at most 8 subarcs $\{a_i\}$, each containing one parallel collection of diagonals, and we choose rectangles $R_i \subset D$ with one side of R_i equal to a_i and so that R_i contains all the diagonals in Γ meeting a_i . We write $a'_i \subset \partial D$ for the side of R_i that's parallel to a_i . We do the same for b , getting subarcs $\{b_j\}$, rectangles $Q_j \subset E$, and parallel subarcs $b'_j \subset \partial E$.

Large arcs. If $|a_i \cap b_j| \geq 3$ for some i, j , then a_i and b_j are called *large*. Similarly, say a'_i and b'_j are large if $|a'_i \cap b'_j| \geq 3$. Note that $|a_i \cap b_j| = |a'_i \cap b'_j|$, so if a_i and b_j are large, then a'_i and b'_j are large (and vice versa).

Graphs $\Theta, \Theta' \subset S$. Define $\Theta \subset X$ the union of all the large a_i and b_j , and define Θ' the union of the large a'_i and b'_j . The set Θ (and Θ' similarly) has the structure of a graph where every vertex has degree either 1 (coming from endpoints of a_i and b_j) or 4 (coming from intersections of a_i and b_j). In addition, Θ and Θ' are isomorphic as graphs in an obvious way.

Define Z, Z' as a small regular neighborhood of Θ, Θ' respectively. We have the following properties, which are proved in [MS13]. We give some brief explanation of the proofs to help assure the reader that these proofs hold in our setting.

- [MS13, Claim 12.7]: The graph Θ is nonempty.

Proof sketch. If $\Theta = \emptyset$, then all the a_i, b_j are small. From this one deduces that $|a \cap b| \leq 128$. Since a, b are components of $\partial D \cap X$ and $\partial E \cap X$, this implies that $d_X(D, E)$ can't be large. In fact, $d_X(D, E) \leq 24$, contradicting our assumption.

- [MS13, Claim 12.8]: No component of either Θ or Θ' is contained in a disk in S or an annulus in S that's peripheral in X .

Proof sketch. If some component of Θ is contained in a disk or peripheral annulus, then one argues that either some a_i, b_j cut a bigon from S , or $a_i, b_j, \partial X$ cut a triangle contained in X . Both of these contradict our assumption that D, E, X are in standard position. The argument for Θ' is the same.

- [MS13, Claim 12.9]: Let Z_1 be a component of Z , and let d be a component of ∂Z_1 . Then d is either inessential or peripheral in X .

Proof sketch. Supposing that d is both essential and non-peripheral, one can deduce that $|a \cap d| < 256$ (counting separately intersections with d coming from large and small a_i). Similarly, $|b \cap d| < 256$, and this implies that $d_X(D, E) < 45$, a contradiction.

- Θ (and hence also Θ') is connected, and Θ fills X .

Proof sketch. This follows from the previous items: any component $Z_1 \subset Z$ fills X by [MS13, Claim 12.9]. Then any other component must be contained in a disk or peripheral annulus in X , which contradicts [MS13, Claim 12.8]. Thus Z is connected and fills X , and this implies the same for Θ .

Let $\mathcal{R} = \{R_i\}$ and $\mathcal{Q} = \{Q_j\}$ be the large rectangles. Observe that $\mathcal{R} \cup \mathcal{Q}$ is an I -bundle whose horizontal boundary is $\Theta \cup \Theta'$. We can thicken $\mathcal{R} \cup \mathcal{Q}$ to an I -bundle T_0 with horizontal boundary $Z \cup Z'$.

Case 1: Θ and Θ' intersect nontrivially. Then $Z \cup Z'$ is a connected subsurface of X . Since complementary components of Z in X are disks and peripheral annuli, the same is also true for complementary components of $Z \cup Z'$ in X .

In this case, the horizontal boundary $\partial_h T_0$ is connected. Then since T_0 is orientable (because V is), the base space of the I -bundle is a nonorientable surface. See §2.5.

The vertical boundary $\partial_v T_0$ is a union of annuli, each connecting a pair of components of $\partial(Z \cup Z')$. If A is such an annulus, and both components of ∂A bound disks in X , then the union of A and these disks is an embedded 2-sphere, which can be filled by a 3-ball, which we parameterize as $D^2 \times I$ to preserve the I -bundle structure. In this way we obtain an I -bundle T with the following properties:

- (1) For every component A of $\partial_v T$, at least one component of ∂A is peripheral in X . In particular the number of components of $\partial_v T$ is at most the number of components of ∂X .
- (2) Every component of ∂X is parallel to a boundary component of some annulus $A \subset \partial_v T$ (because $Z \cup Z'$ fills X , so every component of ∂X is parallel to a component of $\partial(Z \cup Z')$). In particular the number of components of $\partial_v T$ is at least half the number of components of ∂X .

Next we use T to show that every component of ∂X is null-homologous in V . We separate into cases, depending on whether X is homeomorphic to $\Sigma_{1,1}$, $\Sigma_{1,2}$, or $\Sigma_{0,4}$ (we do not need to consider $\Sigma_{0,2}$, $\Sigma_{0,3}$, or Σ_2).

Case 1(a): $X \cong \Sigma_{1,1}$. By properties (1) and (2), $\partial_v T$ has a single component A . One component of ∂A bounds a disk in X , and the other is parallel to ∂X . Thus ∂X bounds a disk in V .

Case 1(b): $X \cong \Sigma_{1,2}$. Then $\partial_v T$ has either one or two components.

If $\partial_v T$ has one component A , then the two components of ∂A are parallel to the two components of ∂X . If A is compressible, then each component of ∂X bounds a disk in V . Otherwise, if A is incompressible, then A can be isotoped into S ; the proof of this is contained in [MS13, Claim 12.16]. From this we conclude that ∂X bounds an embedded $(\mathbb{R}P^2 \# \mathbb{R}P^2) \setminus D^2$ in V because in this case T is a bundle over $(\mathbb{R}P^2 \# \mathbb{R}P^2) \setminus D^2$, and $(\mathbb{R}P^2 \# \mathbb{R}P^2) \setminus D^2$ embeds as a section in the I -bundle, and the boundary of this section is isotopic to ∂X .

If $\partial_v T$ has two components A_1, A_2 , then each A_i has one boundary component parallel to a component of ∂X and another boundary component that is inessential in X . Then each component of ∂X bounds a disk in V (if one component of ∂X bounds a disk then the other does automatically since the boundary components of X are necessarily parallel in S).

Note that in every case, the components of ∂X are nonseparating curves that are null-homologous in V .

Case 1(c): $X \cong \Sigma_{0,4}$. In this case $\partial_v T$ can have two, three, or four components.

If $\partial_v T$ has four components, then every component of ∂X bounds a disk in V .

If $\partial_v T$ has three components, then two components of ∂X bound disks in V . The other two components of ∂X are joined by a component A of $\partial_v T$. If A is compressible, then every component of ∂X bounds a disk. If A is incompressible, then A is isotopic into S , again by [MS13, proof of Claim 12.16]. In this case, T is an I -bundle over $\mathbb{R}P^2 \setminus (D^2 \cup D^2 \cup D^2)$ (because $\partial_h T \cong \Sigma_{0,6}$, c.f. §2.5), and there is an embedding of $\mathbb{R}P^2 \setminus D^2$ (the Möbius band) with boundary isotopic to the components of ∂X parallel to ∂A .

Finally suppose that $\partial_v T$ has two components. In this case $\partial_h T \cong \Sigma_{0,4}$, so T is an I -bundle over $\mathbb{R}P^2 \setminus (D^2 \cup D^2)$.

First suppose that X is incompressible (i.e. there is no disk $D \subset V$ supported in X , primitive or not) and that X is a hole for the disk complex $\mathcal{D}(V)$. Then since the curve complex of $B := \mathbb{R}P^2 \setminus (D^2 \cup D^2)$ has diameter ≤ 4 [Sch82, §2] and the natural map $\mathcal{C}(B) \rightarrow \mathcal{C}(X)$ is distance non-increasing [MS13, §6], this contradicts the fact that the diameter of X with respect to $\mathcal{P}(V)$ is ≥ 57 (any primitive disk D can be surgered to a (not necessarily primitive) vertical disk D' with $d_X(D, D') \leq 6$ [MS13, Lems. 8.12 and 11.7]). See also the argument of [MS13, §12.18].

If X is incompressible, but is not a hole for $\mathcal{D}(V)$, then two components of ∂X bound disks, and this implies that at least one component of $\partial_v T$ is compressible. Suppose the other component A is incompressible (otherwise every component of ∂X bounds a disk), then similar to a previous case, we find that there are two components of ∂X that each bound an embedded $\mathbb{RP}^2 \setminus D^2$ in V .

Now suppose that X is compressible, i.e. there exists a (not necessarily primitive) disk $D \subset V$ with $\partial D \subset X$. If each component of $\partial_v T$ is incompressible, then D can be isotoped to be contained in T , which contradicts the fact that the horizontal boundary component of an I -bundle is incompressible [MS13, Ex. 5.5]. Therefore, at least one component of $\partial_v T$ is compressible, and we conclude as in the previous paragraph.

Overall, we find that in every case, every component of ∂X is null-homologous in V . Note that the only case when ∂X has a component that is separating in S is when $X \cong \Sigma_{1,1}$.

Case 2: Θ and Θ' are disjoint. In this case $T_0 \cong Z \times I$ with horizontal boundary $\partial_h T_0 = Z \sqcup Z'$ and $Z \subset X$ and $Z' \subset Y$.

Claim. $\xi(Y) \geq \xi(X)$, where ξ denotes the complexity $\xi(\Sigma_{g,b}) = 3g - 3 + b$.

Since $X \subset S$ is homeomorphic to one of $\Sigma_{1,1}, \Sigma_{1,2}, \Sigma_{0,4} \subset \Sigma_2$, the claim implies that X and Y are both homeomorphic to $\Sigma_{1,1}$.

Proof of Claim. Since $S = X \cup Y$ (disjoint subsurfaces glued along boundary), X and Y have the same number of boundary components. Thus to prove the claim, it suffices to show that $g(Y) \geq g(X)$. Since Z fills X , $g(X) = g(Z)$. Since $Z \cong Z'$ embeds in Y , $g(Y) \geq g(Z)$. These combine to give $g(Y) \geq g(X)$. This proves the claim.

As a consequence, we also conclude that Z' fills Y , since if $b \geq 1$, then any embedded subsurface $\Sigma_{g,b} \hookrightarrow \Sigma_{g,1}$ fills.

Every component $\partial_v T_0$ is an annulus A connecting a component of ∂Z to a component of $\partial Z'$. If both components of ∂A are inessential, then we obtain a 2-sphere that can be filled with a 3-ball ($\cong D^2 \times I$). In this way, we enlarge T_0 to an I -bundle T so that every component of $\partial_v T$ is an annulus A so that at least one component of ∂A is isotopic to $\partial X \cong \partial Y$.

Consequently, the vertical boundary $\partial_v T$ has either one or two components.

If $\partial_v T$ has two components, then it follows that ∂X bounds a disk in V .

Suppose $\partial_v T$ has one component A . Then $T \cong \Sigma_{1,1} \times I$. The two components of ∂A are parallel to ∂X and ∂Y . If A is compressible, then ∂X bounds a disk in V (and same for ∂Y). If A is incompressible, then A is isotopic into S by [MS13, proof of

Claim 12.16]. This implies that $T \hookrightarrow V$ is isotopic to a homeomorphism.

Step 2. In this step we show that if X has diameter ≥ 61 , then we can also construct an embedded I -bundle \widehat{T} in \widehat{V} .

Lemma 3.7. *Assume that X is a primitively incompressible hole for $\mathcal{P}(V)$ with diameter ≥ 57 . Then X is also a hole for $\mathcal{P}(\widehat{V})$.*

Proof. Suppose for a contradiction that X is not a hole for $\mathcal{P}(\widehat{V})$. This means that there is a primitive disk $\widehat{D} \subset \widehat{V}$ that is supported in $S \setminus X$.

If $X \cong \Sigma_{1,2}$ or $X \cong \Sigma_{0,4}$, then by Step 1, every component of ∂X is nonseparating and nullhomologous in V . This implies that no component of ∂X is isotopic to $\partial \widehat{D}$, since otherwise this would imply that $H_2(S^3) \neq 0$ by Lemma 2.21.

Assume now that X is homeomorphic to $\Sigma_{1,1}$. By Step 1, ∂X bounds a disk in V .

Claim. ∂X also bounds a disk in \widehat{V} .

The claim implies that ∂X is a reducing sphere, which implies that there is a (unique) primitive disk $D \subset V$ with $\partial D \subset X$. This contradicts the fact that X is primitively incompressible. Thus to prove the lemma it suffices to prove the claim.

Proof of Claim. We want to show that ∂X is homotopically trivial in \widehat{V} . Let $\widehat{E} \subset \widehat{V}$ be a primitive disk that is disjoint from \widehat{D} . An element of $\pi_1(\widehat{V})$ is determined by its intersection with \widehat{D} and \widehat{E} (Remark 2.1). Denoting $\delta, \epsilon \in \pi_1(\widehat{V})$ the free generators dual to \widehat{D} and \widehat{E} , then the word $w \in \langle \delta, \epsilon \rangle$ determined by ∂X is a power of ϵ since ∂X is disjoint from \widehat{D} . In addition ∂X is in the commutator subgroup since ∂X bounds $X \cong \Sigma_{1,1}$. This implies w is trivial. \square

Next we show that the diameter of X with respect to $\mathcal{P}(\widehat{V})$ is ≥ 57 . By assumption, there exist primitive disks $D, E \subset V$ with $d_X(D, E) \geq 61$. Let $\widehat{D}, \widehat{E} \subset \widehat{V}$ be the (unique) dual pair of primitive disks. In particular, $D \cap \widehat{E} = \emptyset = \widehat{D} \cap E$, which implies $d_X(D, \widehat{E}), d_X(\widehat{D}, E) \leq 2$ [MM00, Lem. 2.3]. Now the triangle inequality gives $d_X(\widehat{D}, \widehat{E}) \geq 57$ and hence the diameter bound.

Finally, X is primitively incompressible in \widehat{V} because primitively compressible holes for $\mathcal{P}(\widehat{V})$ have diameter ≤ 13 by Proposition 3.3.

Step 3. By Steps 1 and 2, there are embedded I bundles $T \hookrightarrow V$ and $\widehat{T} \hookrightarrow \widehat{V}$. Both T and \widehat{T} have X as a component of their horizontal boundary. We will show that most of the cases enumerated in Step 1 are impossible, for a variety of reasons.

First suppose that X is homeomorphic to $\Sigma_{1,2}$ or $\Sigma_{0,4}$. By Step 1, every component

of ∂X is null-homologous in both V and \widehat{V} . Since the components of ∂X are non-separating in S , this implies that $H_2(S^3) \neq 0$ (Lemma 2.21), which is a contradiction.

Next suppose that $X \cong \Sigma_{1,1}$. Either ∂X bounds a disk in both, exactly one, or neither of V and \widehat{V} . We consider these cases separately. Recall from Step 1 that if ∂X does not bound a disk in V then the inclusion $T \hookrightarrow V$ is (isotopic to) a homeomorphism (and similarly for \widehat{V}).

If ∂X bounds a disk in both, then together these disks form a reducing sphere for the Heegaard splitting. In particular, on either side of this reducing sphere we can find primitive disks in V , but this contradicts the fact that X is a hole for $\mathcal{P}(V)$.

Suppose that ∂X bounds a disk in $\widehat{D} \subset \widehat{V}$ and that $T \hookrightarrow V$ is a homeomorphism. The disk \widehat{D} separates \widehat{V} into two genus-1 handlebodies, one containing X , and each containing a unique disk $\widehat{D}', \widehat{D}''$. These disks are primitive, since we can easily find a vertical disk in $T \cong V$ that intersects \widehat{D}' or \widehat{D}'' once. But this implies that X is not a hole for $\mathcal{P}(\widehat{V})$, contradicting Step 2.

The only remaining possibility is that both $T \hookrightarrow V$ and $\widehat{T} \hookrightarrow \widehat{V}$ are homeomorphisms, which is the desired conclusion of the Theorem. \square

3.4 Subsurface projection for the trefoil knot

In this section we prove the last part of Theorem D, showing if X is a large-diameter hole for $\mathcal{P}(V)$, then there exists $g \in \mathbb{G}$ that preserves X and so that $g|_X$ is pseudo-Anosov. Given the discussion in the introduction to §3, it remains to show that if X is a genus-1 Seifert surface for the trefoil knot, then X has bounded diameter.

In this section we write $S^3 = V \cup \widehat{V}$ as $(F \times I) \cup_\phi (\widehat{F} \times I)$ with $F = \widehat{F} = \Sigma_{1,1}$ and $\phi = T_a T_b$. We also denote $\Sigma = \partial(F \times I)$. For more on this notation, see §2.6.

Theorem 3.8. *Fix $F = \widehat{F} = \Sigma_{1,1}$ and fix $\phi = T_a T_b \in \text{Mod}(F)$. Write $S^3 = (F \times I) \cup_\phi (\widehat{F} \times I)$, and identify this with the standard Heegaard splitting $S^3 = V \cup \widehat{V}$. Then the subsurfaces $X = F \times \{1\}$ and $Y = F \times \{0\}$ have finite diameter with respect to $\mathcal{P}(V)$.*

Proof of Theorem 3.8. The proof for X and Y is the same. For concreteness we work with X . Let $\mathcal{V} \subset \mathcal{P}(V)$ denote the set of vertical disks $D \subset F \times I \cong V$. First we show that \mathcal{V} is finite. Then we use this to show X has finite diameter with respect to $\mathcal{P}(V)$.

Finitely many vertical primitive disks. First we specify the coordinates we will use to compute. Consider the curves and arcs in Figure 21; as pictured, we cut $F = \Sigma_{1,1}$ along the arcs a', b' to get an octagon, which we use to draw pictures.

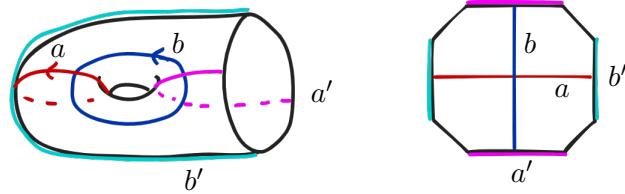


Figure 21: Coordinates for the computation.

Figure 22 shows the intersection of two vertical primitive disks in $\widehat{F} \times I$ with $F \times 0$ and $F \times 1$.

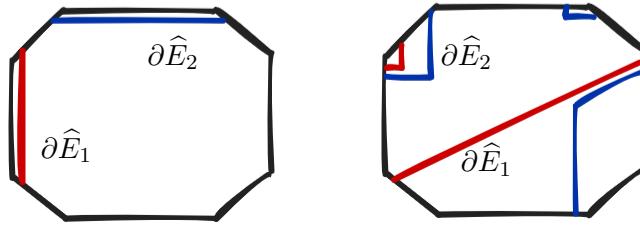


Figure 22: Dual primitive disks.

Given an arc $\lambda \subset \Sigma_{1,1}$, we define the slope as $\frac{\lambda \cdot a}{\lambda \cdot b} \in \mathbb{Q} \cup \{\infty\}$, where the dot product denotes the algebraic intersection number. This depends on our fixed orientation of a, b , but doesn't depend on how we orient λ . For any slope $\frac{p}{q} \in \mathbb{Q} \cup \{\infty\}$, there is a unique arc with that slope; see Figure 23.

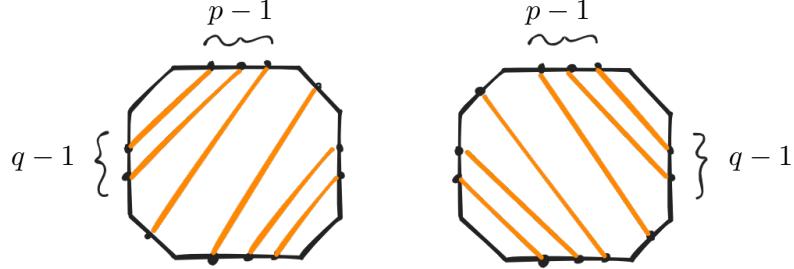


Figure 23: Arcs on $\Sigma_{1,1}$ of slope $4/3$ (left) and $-4/3$ (right).

It is easy to check that the vertical disks corresponding to the arcs of slope

$$\pm 1, 0, \infty, 1/2, 2$$

are all primitive. These fall into two orbits of $\phi = T_a T_b$, which has order 3 in $\text{PSL}_2(\mathbb{Z})$. We show that no other vertical disk is primitive. Let $D = D_{p/q}$ be the

vertical disk corresponding to an arc of slope $\frac{p}{q}$. If $\frac{p}{q} \notin \{0, 1, \infty\}$, then up to the action of ϕ , we can assume that $0 < \frac{p}{q} < 1$. The word w in r, b corresponding to ∂D contains $(q - 1)$ occurrences of r and b , $(q - 1 - p)$ occurrences of r^{-1} , and $(p - 1)$ occurrences of b^{-1} . See Figure 24. Here it's important to remember that the orientations on $F \times 0$ and $F \times 1$ are induced from $F \times I$, and the obvious homeomorphism $F \times 0 \cong F \times 1$ is orientation-reversing.

It's easy to check that w is cyclically reduced (the subword corresponding to $F \times 0$ starts and ends with r , and the subword corresponding to $F \times 1$ starts and ends with b , so there is no cancellation). A cyclically-reduced primitive word doesn't contain both r and r^{-1} , and also doesn't contain both b and b^{-1} . Then at least one of $q - 1$, $p - 1$, or $q - 1 - p$ is equal to 0. The case $q - 1 = 0$ is impossible since $0 < \frac{p}{q} < 1$. If $p - 1 = 0$, then $w = r^q(br^{-1})^{q-1}b$, which implies that $q = 1$, which again is impossible. Finally, if $q - 1 - p = 0$, then $w = r(b^{-1}r)^{p-1}b^p$, which implies that $p = 1$, so $\frac{p}{q} = \frac{1}{2}$.

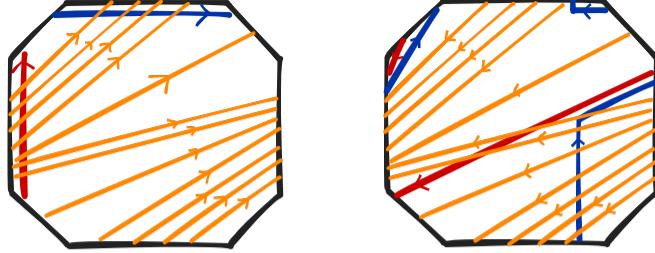


Figure 24: Arc slope $0 < \frac{p}{q} < 1$ pictured on $F \times 0$ and $F \times 1$.

This proves that the only vertical primitive disks are the ones corresponding to arcs of slope $\frac{p}{q} \in \{\pm 1, 0, \infty, 1/2, 2\}$.

X has finite diameter. Fix a primitive disk $D_0 \subset V$. By Lemma 2.13, there exists primitive $D \subset V$ such that D does not admit a primitive boundary compression into $S \setminus n(\partial X)$ and $d_X(D_0, D) \leq 6$. To prove that X has finite diameter, it suffices to show that D is vertical, since we showed above that there are only finitely many vertical disks. The argument below is similar to [MS13, Lem. 8.12] but is more subtle.

Let $D_0, D_1, D_\infty \subset F \times I$ be the vertical primitive disks corresponding to slopes $0, 1, \infty$. These disks split $F \times I$ into two hexagonal prisms. Let $D \subset F \times I$ be a primitive disk, and assume that D cannot be primitively compressed into $S \setminus n(\partial X)$. We show that D can be isotoped to be vertical.

First we can isotope D to be in minimal position with ∂X , and by a further

isotopy we can assume that D is vertical on $\partial F \times I$. Next we can isotope D to be vertical on $D_0 \cup D_1 \cup D_\infty$, since otherwise there would exist a primitive compression of D into $S \setminus n(\partial X)$. Compare with [MS13, Proof of Lem. 8.12].

It remains to show we can isotope D to be vertical in the two hexagonal prisms. Fix one prism $H \times I$, and let D' be the intersection of D with $H \times I$. The surface D' can be built from $D' \cap H \times [1 - \epsilon, 1]$ by a sequence of handle attachments. The first handle attachment is dual to boundary compression (B, a, b) of D' into $X = F \times 1$; here $b \subset X$ is an arc connecting two components δ_1, δ_2 of $D' \cap X$, and a is the core of the handle being attached. One can deduce that the bigon B is primitive (defined in §2.3) by showing that B is an outermost bigon cut from a vertical primitive disk D'' by D' . To see this, observe that either the arc of slope $1/2$ or -1 on X can be isotoped to intersect both δ_1 and δ_2 exactly once. See Figure 25 for an example. It follows that D' cuts a bigon isotopic to B from either $D'' = D_{1/2}$ or $D'' = D_{-1}$. This contradicts the fact that D is primitively incompressible, and from this we deduce that we can isotope D to be vertical in each hexagonal prism, as desired. \square

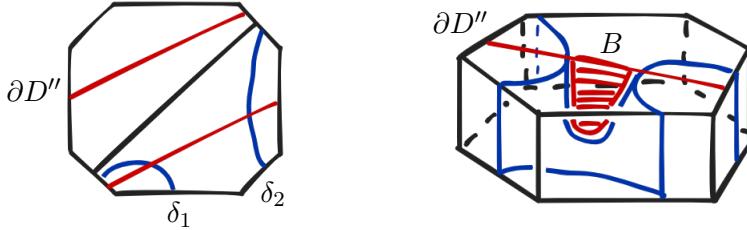


Figure 25: Left: $F \times 1$. Right: hexagonal prism $H \times [1 - \epsilon, 1]$ and a bigon B cut from the vertical disk D'' by D' .

This completes the proof of Theorem D.

3.5 Subsurface projection for the figure-8 knot

In this section, similar to §3.4, we write $S^3 = V \cup \widehat{V}$ as $(F \times I) \cup_\phi (\widehat{F} \times I)$ with $F = \widehat{F} = \Sigma_{1,1}$, but now with $\phi = T_a T_b^{-1}$.

Here we compute the set $\mathcal{V} \subset \mathcal{P}(V)$ of primitive disks that are vertical in $F \times I \cong V$, and prove that the subcomplex $\mathcal{P}(V; X) \subset \mathcal{P}(V)$ spanned by \mathcal{V} is quasi-isometric to a line. This computation is necessary for proving Theorem C (distance formula).

We begin by identifying \mathcal{V} . As in the previous section, let $D_{p/q}$ be the vertical disk whose intersection with X has slope p/q . Figure 26 shows the intersection of two (vertical) primitive disks in $\widehat{F} \times I$ with $F \times 0$ and $F \times 1$. Using these, it is easy to

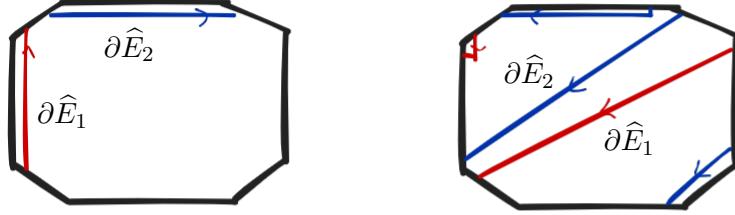


Figure 26: Dual primitive disks.

see that D_0 and D_∞ are primitive. Then the same is true for the orbits of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ under $\phi = T_a T_b^{-1} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. See Figure 27.

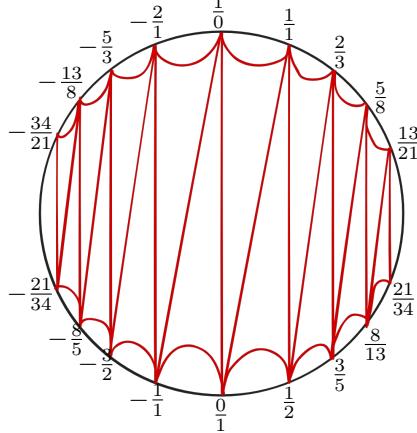


Figure 27: The complex $P(V; X)$ embeds in the Farey graph.

We show that any vertical primitive disk is in one of these two orbits. Fix $D_{p/q}$ such that $\frac{p}{q}$ is not in one of the above two orbits. Up to the action of ϕ , we can assume that $\frac{p}{q}$ belongs to one of the two intervals $(-1, 0)$ or $(1, \infty)$ in \mathbb{Q} .

First suppose that $\frac{p}{q} \in (1, \infty)$. The word $w \in \langle r, b \rangle$ corresponding to $D_{p/q}$ has $(2p - q - 1)$ occurrences of b and $(p - 1)$ occurrences of b^{-1} . See Figure 28. It's easy to check that w is cyclically reduced (the subword corresponding to $F \times 0$ starts and ends with b^{-1} , and the subword corresponding to $F \times 1$ starts and ends with r , so there is no cancellation). Then either $2p - q - 1 = 0$ or $p - 1 = 0$. Both of these are incompatible with the assumption that $\frac{p}{q} \in (1, \infty)$. This shows $D_{p/q}$ is not primitive when $\frac{p}{q} \in (1, \infty)$.

Next suppose that $\frac{p}{q} \in (-1, 0)$. The word $w \in \langle r, b \rangle$ corresponding to $D_{p/q}$ has $(|q| - 1)$ occurrences of r and $(|p| + |q| - 1)$ occurrences of r^{-1} . See Figure 28. The word w is cyclically reduced because the subword corresponding to $F \times 0$ starts and

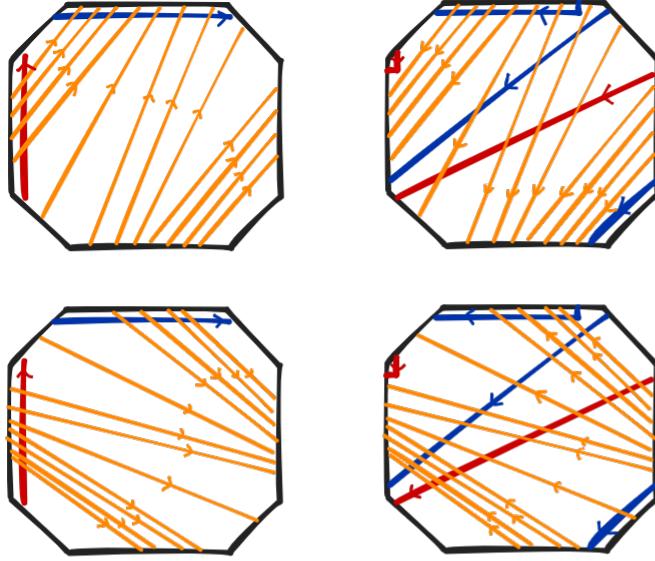


Figure 28: An arc with slope $1 < \frac{p}{q} < \infty$ (top row) or $-1 < \frac{p}{q} < 0$ (bottom row) pictured on $F \times 0$ and $F \times 1$.

ends with r , the subword corresponding to $F \times 1$ starts and ends with b^{-1} , which implies there is no cancellation. Either of the condition $|q|-1=0$ or $|p|+|q|-1=0$ is incompatible with the assumption that $\frac{p}{q} \in (-1, 0)$. This shows that $D_{p/q}$ is not primitive.

Now we claim that for every vertex $D \in \mathcal{P}(V)$ there is a vertical disk D' so that $d_X(D, D') \leq 6$. This follows by the exact same argument as in §3.4; the special fact we needed for that argument is that the vertical disks with slopes $\pm 1, 0, \infty, 1/2$ are primitive; this also holds in the current case.

It follows that the projection of $\mathcal{P}(V)$ to $\mathcal{C}(X)$ is quasi-isometric to the subcomplex pictured in Figure 27, which is quasi-isometric to a line.

Remark 3.9. The curve and arc complexes of $X \cong \Sigma_{1,1}$ are isometric via the map $\mathcal{A}(X) \rightarrow \mathcal{C}(X)$ used to define the subsurface projection (§2.4). Consequently the subsurface projection of $\mathcal{P}(V; X)$ to $\mathcal{C}(X)$ is an embedding, which is in fact a quasi-isometric embedding. In particular, the distance in $\mathcal{P}(V; X)$ between vertices $D, E \in \mathcal{P}(V; X)$ is (coarsely) bounded above by $d_X(D, E)$.

Remark 3.10. In a similar way, we arrive at the same description for $\mathcal{P}(V; Y)$, where $Y = F \times 0$ is the other horizontal boundary component. Since every primitive disk can be surgered to a vertical disk, while only changing its subsurface projection by a uniformly bounded amount, it follows that the infinite-diameter holes X and Y are *paired* in the sense of [MS13, Defn. 5.6]. This will be used in the discussion

of the distance formula.

3.6 Goeritz symmetries preserving the figure-8 knot

The following result is needed for Theorem B.

Proposition 3.11 (figure-8 knot stabilizer). *Let K be the embedding of the figure-8 knot on $S \subset S^3$ in Figure 2. The subgroup of \mathbb{G} that fixes K is generated by α , $\beta\delta\beta^{-1}\delta$, and $\gamma\delta$.*

Proof. First we apply Proposition 2.19 and Remark 2.20. If $g \in \mathbb{G}$ preserves K and preserves each component of $S \setminus K$, then g is represented by a product homeomorphism $g = \psi \times \text{Id}$ on $F \times I$ and $\widehat{F} \times I$, where ψ commutes with $\phi = T_a T_b^{-1}$. The centralizer of

$$A := \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \leftrightarrow \phi \quad \text{in} \quad \text{SL}_2(\mathbb{Z}) \cong \text{Mod}(F)$$

is generated by A and $-I$.

One can show that $\beta\delta\beta^{-1}\delta$ fixes K , preserves each of $F \times 1$ and $F \times 0$, and acts on the homology of both by the matrix A (we found this element by trial and error). The hyperelliptic α (of course) also fixes K and acts on homology by $-I$. Therefore the subgroup of \mathbb{G} that both fixes K and preserves each component of $S \setminus K$ is generated by α and $\beta\delta\beta^{-1}\delta$. This subgroup has index (at most) 2 in the stabilizer of K . Finally, observe that $\gamma\delta$ conjugates $\beta\delta\beta^{-1}\delta$ to its inverse. This element preserves K and swaps the two components of $S \setminus K$ (c.f. Remark 2.20). Therefore, the stabilizer of K is generated by $\alpha, \beta\delta\beta^{-1}\delta$, and $\gamma\delta$. \square

4 Distance formula for $\mathcal{P}(V)$

Our proof of the distance formula (Theorem C) follows closely the approach of [MS13]. There are two parts: an upper and a lower bound on $d_{\mathcal{P}(V)}(D, E)$. The lower bound follows directly from [MM00, Thms. 6.10, 6.12], as explained in [MS13, Thm. 5.14]. To prove the upper bound requires additional work, and for this we use the axiomatic approach of [MS13, §13].

The general axiom-based proof of the upper bound [MS13, Thm. 13.1] is an inductive argument, which, in our case, necessitates the definition of relative versions of $\mathcal{P}(V)$. If $X \subset S$ is a hole for $\mathcal{P}(V)$, then we define $\mathcal{P}(V; X) \subset \mathcal{P}(V)$ as the subcomplex spanned by primitive disks supported in X . Note that we do not necessarily assume that X has infinite diameter. If X is primitively incompressible, then $\mathcal{P}(V; X)$ is empty, which is undesirable, so when X also has infinite diameter, we define $\mathcal{P}(V; X)$ as the complex spanned by vertical primitive disks in the I -bundle constructed in

the proof of Theorem D. By the nature of the induction, we will not need to deal with the case X is a primitively incompressible hole with bounded diameter, so we keep $\mathcal{P}(V; X)$ empty in this case.

For the induction, we need to prove the following more general version of the distance formula upper bound.

Theorem 4.1. *Let $X \subset S$ be a hole for $\mathcal{P}(V)$. Given $c \geq 0$, there is a constant A depending only on c and the topological type of X , such that for any two vertices $D, E \in \mathcal{P}(V; X)$, we have*

$$d_{\mathcal{P}(V; X)}(D, E) \leq A \sum \{d_Y(D, E)\}_c + A,$$

where the sum is over all holes $Y \subset X$ for $\mathcal{P}(V)$.

Given our previous work, the most interesting case of Theorem 4.1 is when $X = S$, as we now explain. If $X \subsetneq S$ is an infinite-diameter hole, then the distance formula holds by our classification of such X in Theorem D and by the computation of $\mathcal{P}(V; X)$ given in §3.5; see Remark 3.9. In the remaining case, when X is a primitively compressible hole, Theorem 4.1 holds trivially by the following proposition.

Proposition 4.2. *Let $X \subsetneq S$ be a primitively compressible hole for $\mathcal{P}(V)$. Then $\mathcal{P}(V; X)$ is a connected, graph with at most two vertices.*

Proof. As noted in Remark 3.4, if a hole $X \subset S$ supports two primitive disks with nonzero intersection number, then $X = S$. Therefore, our assumption $X \neq S$ implies that $\mathcal{P}(V; X)$ is either a single vertex or is two vertices connected by an edge. \square

It remains to prove Theorem 4.1 for $X = S$. It is here that we use the axiomatic approach of [MS13].

4.1 Axiomatic approach to the distance formula for $\mathcal{P}(V)$

This section explains a proof of Theorem 4.1 in the main case of interest $X = S$.

Our approach is to use [MS13, Thm. 13.1], which states that there are seven axioms that together imply the desired upper bound. Instead of writing the full statement of the axioms with all the necessary definitions (this takes multiple pages in [MS13]), we give precise references to [MS13] and follow their notation as closely as possible. Our goal is to explain why the axioms hold for $\mathcal{P}(V)$, focusing on the parts that are the least straight-forward. Ultimately, the argument is very similar to the proof in [MS13, §18-19] that the axioms are satisfied for the disk complex $\mathcal{D}(V)$, using our analogous results. In fact, there are even some simplifications because we only work

in genus 2 and because the classification of large holes (Theorem D) is simpler than the corresponding result for $\mathcal{D}(V)$, c.f. [MS13, Thm. 1.1].

Below we list the axioms and explain why they hold for $\mathcal{P}(V)$. See [MS13, §13] for statements of the axioms, and see [MS13, §18-19] for additional information about why the axioms hold for $\mathcal{D}(V)$.

Large Holes Axiom. This is stated in [MS13, Axiom 13.3]. Here we may take “large” to mean diameter ≥ 61 , which implies the hole has infinite diameter (Theorem D). This axiom states that any two large holes either overlap [MS13, Defn. 2.8] or are paired [MS13, Defn. 5.6]. This holds trivially for $\mathcal{P}(V)$ because all proper holes are homeomorphic to $\Sigma_{1,1}$, so the only way two large holes can be disjoint is if they are paired, cf. Remark 3.10.

Marking Path, Accessibility Intervals, Combinatorial Sequence Axioms. These three axioms concern properties of the following data associated to a pair of primitive disks $D, E \subset V$. These axioms hold for the exact same reason as they hold for $\mathcal{D}(V)$. Start with a surgery path $D = D_0, \dots, D_N = E$ as defined in §2.3.

- The *marking path* is defined as follows. The sequence $\{D_n\}$ induces a nested sequence of train tracks $\tau_0 \succ \dots \succ \tau_N$ by the construction explained in [MM04, §4]. A train track τ has an associated vertex cycle $\text{Vert}(\tau) \subset \mathcal{C}(S)$ (see e.g. [MM04, §3]), which is a marking in the sense of [MS13, §2.9]. The marking path is defined as $\mu_n = \text{Vert}(\tau_n)$.
- Define the *combinatorial sequence* as follows. By [MM04, Thm. 1.3], the subset $\bigcup_{n=0}^N \text{Vert}(\tau_n) \subset \mathcal{C}(S)$ is an unparameterized quasi-geodesic, which means that there is a subsequence $\tau_0 \succ \dots \succ \tau_K$ such that for any choice of $v_k \in \text{Vert}(\tau_k)$, the sequence v_0, \dots, v_K is a quasi-geodesic in $\mathcal{C}(S)$. The combinatorial sequence, is the associated sequence $D = D_0, \dots, D_K = E$ of primitive disks.
- By construction of the combinatorial sequence, there is an increasing *re-indexing function* $r : [0, K] \rightarrow [0, N]$.
- Given the sequence $\{\tau_n\}_{n=0}^N$, for each subsurface $X \subset S$, an *accessibility interval* $J_X \subset [0, N]$ is defined in [MMS12, §5.2].

The Marking Path Axiom [MS13, Axiom 13.4] states that the μ_n have nested supports and that the projection $\pi_X(\mu_n)$ to any subsurface $X \subset S$ is an unparameterized quasi-geodesic in $\mathcal{C}(X)$ with constant depending only on $\mathcal{P}(V)$. The former property holds by construction, and the latter property holds as a consequence of a general result proved by Masur–Mosher–Schleimer [MMS12, Thm. 5.5].

The Accessibility Intervals Axiom (see [MS13, Axiom 13.5]) is verified in [MMS12, Thm. 5.3], just as for $\mathcal{D}(V)$; see [MS13, §18].

The Combinatorial Sequence Axiom (see [MS13, Axiom 13.6]) follows from the construction of $\{D_k\}_{k=0}^K$ and $\{\mu_n\}_{n=0}^N$, just as for $\mathcal{D}(V)$. Details are in [MM04, §3-4]; see also [MS13, Thm. 19.3].

Replacement Axiom. This is stated in [MS13, Axiom 13.7]. The proof of this axiom for $\mathcal{P}(V)$ follows quickly from the following claim (cf. [MS13, §19.5]).

Claim. Let $Y \subsetneq S$ be a large incompressible hole for $\mathcal{P}(V)$ with corresponding I -bundle T , and fix a primitive disk D such that $i(\partial D, \partial Y) < K$ for some constant K , then there exists a primitive disk D' that is vertical in T and such that $d_{\mathcal{P}(V)}(D, D') \leq \log_2(K)$.

Proof of Claim. The proof is similar to [MS13, §19.5]. Consider a maximal sequence of primitive compressions of D into $S \setminus \partial Y$. Each compression replaces D with two disks, and we choose the one with at most half as many intersections with ∂Y . This process ends after $\leq \log_2(K)$ steps with a disk D' with no primitive compression into $S \setminus \partial Y$, and by the argument in §3.5, D' can be isotoped to be vertical. This proves the claim.

Straight Intervals Axiom. See [MS13, Axiom 13.13]. This axiom holds for $\mathcal{P}(V)$ by the exact same argument as [MS13, §19.7].

Shortcut Intervals Axiom. See [MS13, Axiom 13.14]. The proof of this axiom ultimately boils down to the following claim (cf. [MS13, §19.8]).

Claim. Let Z be a non-hole for $\mathcal{P}(V)$, and fix disks D, E such that $i(\partial D, \partial Z) < K$ and $i(\partial E, \partial Z) < K$. Then $d_{\mathcal{P}(V)}(D, E) < K'$ for some constant K' depending only on K .

Proof of Claim. Let $Y \subset S$ be the (closure of the) complement of Z . Note that Y is primitively compressible because Z is a non-hole for $\mathcal{P}(V)$ and $Y \cup Z = S$. As in the argument for the Replacement Axiom, we choose maximal compression sequences $D \rightsquigarrow D'$ and $E \rightsquigarrow E'$ of D and E into $S \setminus \partial Z$ with each sequence of length $\leq \log_2(K)$. It suffices to give a uniform bound on $d_{\mathcal{P}(V)}(D', E')$. If D' and E' are disjoint, then we are done, so we assume D', E' intersect nontrivially. Then D', E' lie in the same component of $S \setminus \partial Z$. If this component is Z , then $d_{\mathcal{P}(V)}(D', E') \leq 2$ since Y is primitively compressible. Next suppose D', E' are contained in a component Y' of Y . Observe that Y' is not a hole for $\mathcal{P}(V)$ since if it were, this would imply $Y = S$ by Remark 3.4. Then Y is not a hole, so again we conclude that $d_{\mathcal{P}(V)}(D', E') \leq 2$.

4.2 Consequences of the distance formula

In this section we explain several consequences of the distance formula that we use in Section 5 to prove Theorem A. The results in this section appear in a similar form in [BBKL20], although we need to formulate them differently, replacing the assumption that G is quasi-isometrically embedded in $\text{Mod}(S)$ with the assumption that $G < \mathbb{G}$ and the orbit map $G \rightarrow \mathcal{P}(V)$ is quasi-isometrically embedded. In fact, the results we prove work more generally:

In this section (and this section only), we let S denote any closed oriented surface of genus ≥ 2 , and we fix a subgroup $\mathbb{G} < \text{Mod}(S)$ that acts on a subcomplex $\mathcal{P} \subset \mathcal{C}(S)$. We assume that \mathcal{P} has a distance formula. Similar to [BBKL20, §2], we express the distance formula in the following form, which is equivalent to (3): given $\beta > 0$ (sufficiently large), there is $\lambda > 0$ so that

$$\frac{1}{\lambda} d_{\mathcal{P}}(u, u') \leq \sum_{\substack{X \subset S \\ \text{hole for } \mathcal{P}}} \{d_X(u, u')\}_{\beta} \leq \lambda d_{\mathcal{P}}(u, u') \quad (5)$$

for any vertices u, u' of \mathcal{P} satisfying either $\sum_X \{d_X(u, u')\}_{\beta} \neq 0$ or $d_{\mathcal{P}}(u, u') \geq \lambda$.

We also assume that there is complex \mathcal{M} whose vertices are markings (i.e. vertices of $\mathcal{M}(S)$; see §2.2) such that (1) \mathcal{M} is quasi-isometric to \mathbb{G} , and (2) there is a uniform bound on the distance in $\mathcal{M}(S)$ between adjacent markings in \mathcal{M} .

All of these assumptions hold for \mathbb{G} the genus-2 Goeritz group, $\mathcal{P} = \mathcal{P}(V)$ the primitive disk complex, and $\mathcal{M} = \mathcal{M}(V, \hat{V})$ the Heegaard marking complex. See Theorem C and Lemmas 2.3 and 2.4.

The main result needed for the proof of Theorem A is Proposition 4.3. This result and its proof are similar to the main theorem of [BBKL20].

Proposition 4.3. *Fix a finitely-generated subgroup $G < \mathbb{G}$. If the orbit map $G \rightarrow \mathcal{P}$ is a q.i. embedding, but the orbit map $G \rightarrow \mathcal{C}(S)$ is not a q.i. embedding, then G contains a reducible element.*

To prove Proposition 4.3, we use two additional results. The first, Proposition 4.4, is proved in [BBKL20, Prop. 3.1]. The second, Proposition 4.5, is similar to [BBKL20, Prop. 4.1] but with a slightly different assumption. We sketch the proof below.

For a subsurface $Z \subset S$, we write $\mathcal{M}(Z)$ for the marking complex; see [BBKL20, §2]. Here ‘‘marking’’ refers to the clean, complete markings defined in [MM00, §2.5]; this differs from the more flexible notion of markings [MS13, §2.9] that is used in §4.1.

Proposition 4.4 (Reducibility criterion). *Let $G < \text{Mod}(S)$ be finitely generated, and let $|g|$ denote the word length of $g \in G$ with respect to a finite generating set.*

Fix a marking μ . For any $c > 0$ there exists $R = R(c) > 0$ so that if $|g| > R$ and there exists a proper subsurface $Z \subset S$ with $d_{\mathcal{M}(Z)}(\mu, g\mu) \geq c|g|$, then G contains a reducible element.

Proposition 4.5 (Linearly summing projections). *Fix $G < \mathbb{G}$ with finite generating set Ω , and fix a marking $\mu \in \mathcal{M}$. Assume that the orbit map $G \rightarrow \mathcal{P}$ is a quasi-isometric embedding. Then there exists $K, C > 0$ with the following property. For $g \in G$, if $|g| > C$ and $d_S(\mu, g\mu) < \frac{|g|}{K}$, then we can find subsurfaces $Z_1, \dots, Z_k \subset S$ and write $g = g_1 \cdots g_k$ geodesically in G with $k \leq d_S(\mu, g\mu)$ so that*

$$|g| \leq K \sum_{j=1}^k d_{\mathcal{M}(Z_j)}(\mu, g_j\mu).$$

Furthermore, by increasing K we can also arrange that $d_{\mathcal{M}(Z_j)}(\mu, g_j\mu) \leq K|g_j|$ for each j .

Proof of Proposition 4.3. Fix a marking $\mu \in \mathcal{M}$. Let K, C be the constants from Proposition 4.5, and fix $M > K$ (to be chosen more precisely later). Since $G \rightarrow \mathcal{C}(S)$ is not a q.i. embedding, we can find group elements $g \in G$ with $|g|$ arbitrarily large and $d_S(\mu, g\mu) \leq \frac{|g|}{M}$. Then we can apply Proposition 4.5: write $g = g_1 \cdots g_k$ with $k \leq d_S(\mu, g\mu)$ and take $Z_1, \dots, Z_k \subset S$ so that $|g| \leq K \sum_{j=1}^k d_{\mathcal{M}(Z_j)}(\mu, g_j\mu)$.

For $t > 0$, let $R(\frac{1}{t})$ be the constant from Proposition 4.4, and consider the partition of $\{1, \dots, k\}$ into

$$J_{\leq}(t) = \{j : d_{\mathcal{M}(Z_j)}(\mu, g_j\mu) \leq \frac{|g_j|}{t}\} \quad \text{and} \quad J_{>}(t) = \{j : d_{\mathcal{M}(Z_j)}(\mu, g_j\mu) > \frac{|g_j|}{t}\}.$$

If $|g_j| \leq R(\frac{1}{t})$ for each $j \in J_{>}(t)$, then using the conclusions of Proposition 4.5 and the assumption $d_S(\mu, g\mu) \leq \frac{|g|}{M}$, we obtain the following estimate (cf. [BBKL20, §5]):

$$\begin{aligned} |g| &\leq K \sum_{j=1}^k d_{\mathcal{M}(Z_j)}(\mu, g_j\mu) \\ &= K \left[\sum_{j \in J_{\leq}(t)} d_{\mathcal{M}(Z_j)}(\mu, g_j\mu) + \sum_{j \in J_{>}(t)} d_{\mathcal{M}(Z_j)}(\mu, g_j\mu) \right] \\ &\leq K \left[\frac{|g|}{t} + \sum_{j \in J_{>}(t)} K|g_j| \right] \\ &\leq K \left[\frac{|g|}{t} + k \cdot K \cdot R(\frac{1}{t}) \right] \\ &\leq K \left[\frac{|g|}{t} + d_S(\mu, g\mu) \cdot K \cdot R(\frac{1}{t}) \right] \\ &\leq K \left[\frac{|g|}{t} + \frac{|g|}{M} \cdot K \cdot R(\frac{1}{t}) \right] = \left(\frac{K}{t} + \frac{K \cdot R(\frac{1}{t})}{M} \right) \cdot |g| \end{aligned}$$

Now choose $t > 2K$ and any $M > 2K \cdot R(\frac{1}{t})$, so that $\left(\frac{K}{t} + \frac{K \cdot R(\frac{1}{t})}{M} \right) < 1$. By the computation above, there must be some $j \in J_{>}(t)$ with $|g_j| > R(\frac{1}{t})$. Then we may apply Proposition 4.4 to conclude that G contains a reducible element. \square

It remains to explain the proof of Proposition 4.5. Before this, we recall the distance formula for the marking complex due to [MM00]; see also [BBKL20, Thm. 2.7] for the form that we state it. Let $Z \subset S$ be a subsurface. Given $\beta > 0$ (sufficiently large), there is $\kappa > 0$ so that

$$\frac{1}{\kappa} d_{\mathcal{M}(Z)}(\mu, \mu') \leq \sum_{Y \subset Z} \{d_Y(\mu, \mu')\}_\beta \leq \kappa d_{\mathcal{M}(Z)}(\mu, \mu') \quad (6)$$

for every $\mu, \mu' \in \mathcal{M}(S)$ so that either the middle term is nonzero or $d_{\mathcal{M}(Z)}(\mu, \mu') \geq \kappa$.

Proof of Proposition 4.5. We explain the last statement first (this is one difference between our argument and the argument in [BBKL20, §4]). Since the orbit map $G \rightarrow \mathcal{P}$ is a quasi-isometric embedding, so too is the orbit map $G \rightarrow \mathcal{M}$ (see Lemma 2.6). Then by Lemma 2.3, $d_{\mathcal{M}(S)}(\mu, g_j\mu)$ is coarsely bounded above by $|g_j|$. Combining this with the fact that subsurface projection of markings is coarsely Lipschitz (see [BBKL20, Prop. 2.3]), one deduces the last sentence of the proposition.

The rest of the proof is very similar to [BBKL20, §4]. The only difference is that we substitute, in the appropriate place, the distance formula for \mathcal{P} (5) for the Masur–Minsky distance formula (6). (Note, however, that we still use the Masur–Minsky distance formula in a different part of the argument.) Since precise details are contained in [BBKL20, §4], we will only sketch the argument. In particular, we will be imprecise with some of the constants (e.g. we will not carefully specify C).

Fix $\beta \gg 0$ sufficiently large (precisely how large will be evident at the end of the argument). Since $G \rightarrow \mathcal{P}$ is a q.i. embedding, by the distance formula (5), there is a constant K_1 so that if $|g| \gg 0$, then

$$|g| \leq K_1 \sum_{\substack{X \subset S \\ \text{hole for } \mathcal{P}}} \{d_X(\mu, g\mu)\}_\beta$$

BBF factors and subsurface order. By [BBF15], it is possible to decompose the set of essential subsurfaces of S into finitely many equivalence classes (called BBF factors), one of which is $\{S\}$, so that surfaces in the same BBF factor overlap (i.e. are neither disjoint nor nested). Combining this fact with the preceding inequality, there is a constant $K_2 \geq K_1$ and a BBF factor \mathbb{Y} so that

$$|g| \leq K_2 \sum_{\substack{Y \in \mathbb{Y} \\ \text{hole}}} \{d_Y(\mu, g\mu)\}_\beta. \quad (7)$$

Ultimately, we will choose our constant K to be larger than K_2 , for then the assumption $d_S(\mu, g\mu) < \frac{|g|}{K}$ implies that $\mathbb{Y} \neq \{S\}$ in (7). Assuming this, let $Y_1, \dots, Y_n \in \mathbb{Y}$ be the holes with $d_Y(\mu, g\mu) \geq \beta$. If β is large enough, it's possible to put the Y_i in increasing order with respect to the subsurface order, which is defined using the Behrstock inequality, c.f. [BBKL20, Prop. 2.6].

Maximal subsurfaces. Define $0 = i_0 < i_1 < \dots < i_k = n$ inductively so that i_j is the largest index so that

$$Y_{i_{j-1}+1} \cup \dots \cup Y_{i_j}$$

is a proper subsurface. Denote this subsurface by Z'_j . If β is sufficiently large, then the number k of these maximal subsurfaces is bounded above by $d_S(\mu, g\mu)$ by [BBKL20, Lem. 4.4].

We use this to decompose the sum in (7)

$$\sum_{i=1}^n d_{Y_i}(\mu, g\mu) = \sum_{j=1}^k \sum_{\ell=i_{j-1}+1}^{i_j} d_{Y_\ell}(\mu, g\mu)$$

Prefixes of g . Next we bound the preceding sum. Write $g = s_1 \dots s_m$ in the generating set Ω . Let $b > 0$ be an upper bound on $d_{\mathcal{M}(Z)}(\mu, s\mu)$ and $d_Z(\mu, s\mu)$ for $s \in \Omega$ and any subsurface $Z \subset S$. For $1 \leq j < k$ define g'_j as the largest prefix of $g = s_1 \dots s_m$ so that $d_{Y_{i_j}}(g'_j \mu, g\mu) \geq 2b$. Set also $g_0 = \text{Id}$ and $g_k = g$. By [BBKL20, Lem. 4.3], g'_{j-1} is a prefix of g'_j , and

$$d_{Y_\ell}(\mu, g\mu) - 5b \leq d_{Y_\ell}(g'_{j-1}\mu, g'_j\mu) \tag{8}$$

for $i_{j-1} < \ell \leq i_j$. (Note: here our notation differs from [BBKL20]; it would agree if we wrote g'_{i_j} instead of g'_j .) Since $d_{Y_\ell}(\mu, g\mu) \geq \beta$ by construction, if we choose β sufficiently large, we can assume that $d_{Y_\ell}(\mu, g\mu) - 5b \geq \frac{1}{2}d_{Y_\ell}(\mu, g\mu)$. Combining this with (8) gives

$$d_{Y_\ell}(\mu, g\mu) \leq 2 d_{Y_\ell}(g'_{j-1}\mu, g'_j\mu).$$

Compare with [BBKL20, Eqn. (5)]. Denoting $g_j = (g'_{j-1})^{-1}g'_j$ and $Z_j = (g'_{j-1})^{-1}(Z'_j)$, we have

$$\begin{aligned} \sum_{\ell=i_{j-1}+1}^{i_j} d_{Y_\ell}(\mu, g\mu) &\leq 2 \sum_{\ell} d_{Y_\ell}(g'_{j-1}\mu, g'_j\mu) \\ &= 2 \sum_{\ell} d_{(g'_{j-1})^{-1}(Y_\ell)}(\mu, g_j\mu) \\ &\leq 2 \sum_{Y \subset Z_j} \{d_Y(\mu, g_j\mu)\}_\beta \\ &\leq 2\kappa d_{\mathcal{M}(Z_j)}(\mu, g_j\mu) \end{aligned}$$

The last inequality uses the distance formula for the marking complex (6). (We know that the distance formula applies because the sum on the left-hand side above is nonzero.)

Conclusion. Now summing over j , we obtain

$$|g| \leq K_2 \sum_{i=1}^n d_{Y_i}(\mu, g\mu) \leq 2\kappa K_2 \sum_{j=1}^k d_{\mathcal{M}(Z_j)}(\mu, g_j\mu)$$

which is the desired inequality. Therefore, if $K = 2\kappa K_2$ and $|g|$ is sufficiently large with $d_S(\mu, g\mu) \leq \frac{|g|}{K}$, then we obtain $Z_1, \dots, Z_k \subset S$ and $g = g_1 \cdots g_k$ with the desired properties. \square

5 Purely pseudo-Anosov implies convex cocompact

This section contains the proof of Theorem A.

Fix $G < \mathbb{G}$ finitely-generated, purely pseudo-Anosov. To show that G is convex cocompact in $\text{Mod}(S)$, it suffices to show that the orbit map $G \rightarrow \mathcal{C}(S)$ is a q.i. embedding; furthermore, by Proposition 4.3, it suffices to show the orbit map $G \rightarrow \mathcal{P}(V)$ is a q.i. embedding. We proceed in the following steps.

- Step 1. We show that $\mathcal{P}(V)$ is quasi-isometric to a certain coned-off Cayley graph $\text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_E)$.
- Step 2. We show that $G \rightarrow \text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_E)$ is a q.i. embedding, using in particular a result of [AM21] and the fact that \mathbb{G} is virtually free.

5.1 Step 1: quasi-isometry type of $\mathcal{P}(V)$

Let $\Gamma_{\mathbb{G}}$ be the Cayley graph of \mathbb{G} with respect to the standard generating set (Figure 1). As usual, we view $\Gamma_{\mathbb{G}}$ as a metric space by giving each edge length 1. Given a subgroup $H < \mathbb{G}$, we denote $\text{Cone}(\Gamma_{\mathbb{G}}, H)$ be the space obtained from $\Gamma_{\mathbb{G}}$ by coning off each translate of $H \subset \Gamma_{\mathbb{G}}$. Specifically, $\text{Cone}(\Gamma_{\mathbb{G}}, H)$ is the graph obtained from $\Gamma_{\mathbb{G}}$ by adding an additional vertex $*_{gH}$ for each coset $gH \in \mathbb{G}/H$ and edges from $*_{gH}$ to the vertices corresponding to gH .

More generally, given a \mathbb{G} -space X with basepoint x_0 and a subgroup $H < \mathbb{G}$, we denote $\text{Cone}(X, H)$ the space obtained by coning off $gH(x_0)$ for each $gH \in \mathbb{G}/H$.

Proposition 5.1 ($\mathcal{P}(V)$ as a coned-off Cayley graph). *Fix a primitive disk $E \subset V$, and let \mathbb{G}_E be its stabilizer in \mathbb{G} . Then the complexes $\text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_E)$ and $\mathcal{P}(V)$ are quasi-isometric.*

Proof. Recall from §2.2 that $\mathcal{R}(V, \hat{V})$ denotes the reducing sphere complex. Choose as basepoint the reducing sphere P pictured in Figure 5, and let $E = E_2$ be the primitive disk pictured in Figure 3.

Claim 1. The complex $\text{Cone}(\mathcal{R}(V, \hat{V}), \mathbb{G}_E)$ is isomorphic to a simplicial subdivision of $\mathcal{P}(V)$.

Proof of Claim 1. This follows from the work of [Cho08] that was discussed in §2.2: By Remark 2.2, $\mathcal{P}(V)$ is obtained from $\mathcal{R}(V, \widehat{V})$ by coning off, for each primitive disk D , the subcomplex $\mathcal{R}_D \subset \mathcal{R}(V, \widehat{V})$ spanned by reducing spheres that are disjoint from D . By the proof of Proposition 2.5, \mathbb{G}_E acts transitively on the vertices of \mathcal{R}_E , so coning \mathcal{R}_E is the same as coning off the orbit of P under \mathbb{G}_E . Since \mathbb{G} acts transitively primitive disks, we conclude that $\mathcal{P}(V)$ is obtained from $\mathcal{R}(V, \widehat{V})$ by coning the $g\mathbb{G}_E(P)$ for each coset $g\mathbb{G}_E$. This proves Claim 1.

Replacing each triangle of $\mathcal{R}(V, \widehat{V})$ by the cone on its vertices defines a graph T with a quasi-isometry $T \hookrightarrow \mathcal{R}(V, \widehat{V})$, and T is a tree by [Akb08, Thm. 1] and [Cho08, §6]. Furthermore, T is the Bass–Serre tree for the splitting of \mathbb{G} in (1) (indeed Akbas and Cho obtain the splitting for \mathbb{G} from T).

Claim 2. The tree T is quasi-isometric to $\text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_P \cap \Gamma_E)$.

Proof of Claim 2. As mentioned above, T is a Bass–Serre tree for the splitting $G = A *_C B$ in (1). In particular, T can be obtained from $\Gamma_{\mathbb{G}}$ by collapsing cosets of the vertex groups to points. Up to quasi-isometry, this is the same as coning off cosets of the vertex groups. From this one deduces that T is quasi-isometric to $\text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_P)$ (since \mathbb{G}_P is one of the vertex groups, and the other vertex group is finite).

Recall that \mathbb{G}_P is generated by $\{\alpha, \beta, \gamma\}$ by [Sch04, §2], and \mathbb{G}_E is generated by $\{\alpha, \beta, \gamma\delta\}$ by Proposition 2.5. It follows from the presentation (1) for \mathbb{G} that $\mathbb{G}_E \cap \mathbb{G}_P = \langle \alpha, \beta \rangle$. In particular, $\mathbb{G}_E \cap \mathbb{G}_P$ has finite index in \mathbb{G}_P , so $\text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_P)$ and $\text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_E \cap \mathbb{G}_P)$ are quasi-isometric. This proves Claim 2.

Finally, observe that $\text{Cone}(\text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_E \cap \mathbb{G}_P), \mathbb{G}_E)$ and $\text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_E)$ are quasi-isometric (this holds generally). Combining all of the above quasi-isometries, gives the desired conclusion:

$$\begin{aligned} \mathcal{P}(V) &\sim \text{Cone}(\mathcal{R}, \mathbb{G}_E) \sim \text{Cone}(T, \mathbb{G}_E) \\ &\sim \text{Cone}(\text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_E \cap \mathbb{G}_P), \mathbb{G}_E) \sim \text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_E). \quad \square \end{aligned}$$

5.2 Step 2: the orbit map $H \rightarrow \text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_E)$

In this step we show (Proposition 5.3) that the orbit map $G \rightarrow \text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_E)$ is a q.i. embedding. Combined with Step 1, this proves Theorem A.

To show $G \rightarrow \text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_E)$ is a q.i. embedding, we first show $G \hookrightarrow \Gamma_{\mathbb{G}}$ is a q.i. embedding.

Lemma 5.2.

(i) Every finitely generated subgroup $H < \mathbb{G}$ is quasi-isometrically embedded.

(ii) *The Goeritz group \mathbb{G} is virtually free.*

Proof. First we explain why (ii) implies (i). This is a well-known consequence of Marshall Hall's theorem [Hal49, Thm. 5.1], which states that if K is a finitely generated subgroup of a free group F , then K is a free factor of a finite-index subgroup, i.e. there exists a finite-index $F' < F$ so that $F' = K * L$ (for some subgroup L). By this result, K is isometrically embedded in F' (e.g. use the retract $F' \twoheadrightarrow K$), and hence $K \hookrightarrow F$ is a q.i. embedding. Clearly a similar argument works for virtually free groups.

Now we prove (ii). A finitely generated group is virtually free if and only if it can be expressed as the fundamental group of a finite graph of finite groups; see [Ser80, Ch. II.2.6] and [SW79, Thm. 7.3 ff.]. We show that \mathbb{G} has this structure.

By the presentations of \mathbb{G} given by [Akb08, Thm. 2] and [Cho08, §5], \mathbb{G} is an amalgamated product as in (1). Any such group is the fundamental group of a graph of groups of the form pictured in Figure 29. This proves the lemma. \square

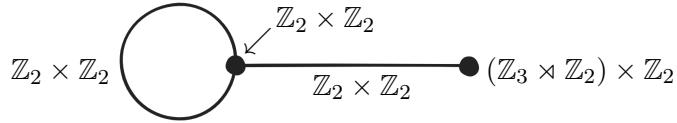


Figure 29: A graph of groups description of \mathbb{G} .

We use Lemma 5.2 to prove Proposition 5.3, which implies that $G \rightarrow \text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_E)$ is a q.i. embedding (since G is purely pseudo-Anosov).

Proposition 5.3. *Fix a subgroup $H < \mathbb{G}$. Assume that H acts freely on the set of primitive disks. Then the orbit map $H \rightarrow \text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_E)$ is a q.i. embedding.*

Proof. By Lemma 5.2 any finitely generated $H < \mathbb{G}$ is quasi-isometrically embedded in \mathbb{G} , hence quasi-convex [BH99, III.Γ.3.6]. By [AM21, Corollary 6.13], to show that $H \rightarrow \text{Cone}(\Gamma_{\mathbb{G}}, \mathbb{G}_E)$ is a q.i. embedding it suffices to show that the limit set $\Lambda(H) \subset \partial\mathbb{G}$ is disjoint from the limit sets $\Lambda(g\mathbb{G}_Eg^{-1}) = \Lambda(g\mathbb{G}_E)$ for each $g \in \mathbb{G}$.

Since $\Lambda(G_1) \cap \Lambda(G_2) = \Lambda(G_1 \cap G_2)$ in $\partial\mathbb{G}$ for quasi-convex subgroups $G_1, G_2 < \mathbb{G}$ of a hyperbolic group \mathbb{G} [GMRS98, Lem. 2.6], we conclude that $\Lambda(H) \cap \Lambda(g\mathbb{G}_Eg^{-1})$ must be empty, since otherwise $H \cap g\mathbb{G}_Eg^{-1}$ would be nontrivial for some $g \in \mathbb{G}$, which contradicts the assumption that H acts freely on the set of primitive disks. \square

6 Characterization of Goeritz pseudo-Anosov elements

This section contains the proof of Theorem B. For this, we fix $g \in \mathbb{G}$ and assume g is not pseudo-Anosov. If g has finite order, then g is conjugate into one of the vertex groups in (1) by the theory of graphs of groups [SW79, Cor. 3.8]. These vertex groups appear as (ii) and (iii) in the statement of Theorem B. Therefore, the most interesting case of Theorem B is when g has infinite order.

Assuming $g \in \mathbb{G}$ is reducible and infinite order, to prove Theorem B, it suffices to show that g (or a power) stabilizes either a finite set of primitive disks or an infinite-diameter genus-1 subsurface (these are classified in §3). We prove this with a case-by-case analysis of the canonical reduction system $\text{CRS}(g)$. As such, the proof will also immediately prove Corollary 1.2.

Recall that a *reduction system* of a reducible mapping class g is a multicurve that is preserved by g . Reduction systems for g are partially ordered by inclusion, and the *canonical reduction system* $\text{CRS}(g)$ is defined as the intersection of all maximal reduction systems of g . See [FM12, §13.2] for more information and also [BLM83, Lem. 2.6] for basic properties of $\text{CRS}(g)$.

In general, there are the following possibilities for a multicurve, like $\text{CRS}(g)$, on a genus-2 surface:

- (A) a separating curve
- (B) a nonseparating curve
- (C) two curves: one separating, one nonseparating
- (D) two nonseparating curves
- (E) three curves: one separating, two nonseparating
- (F) three nonseparating curves

To analyze these possibilities we use some results of Oertel [Oer02] about the structure of homeomorphisms of handlebodies:

- Fact 1. A multitwist of $S = \partial V$ extends to the handlebody V if and only if there is a collection of essential disks and annuli whose boundary is the multicurve [Oer02, Thm. 1.11] (an essential annulus is incompressible and not boundary parallel).
- Fact 2. If $X \subset S$ is an essential subsurface that is preserved by a homeomorphism $g : V \rightarrow V$, then the *characteristic compression body* $Q_X \subset V$ is also preserved by g . See [Oer02, Cor. 2.2] for this statement and [Oer02, §2] for

the definition of the characteristic compression body; the main example we will use is that if $X \cong \Sigma_{1,1}$ supports a disk, then Q_X is a solid torus.

Case (A). Suppose that $\text{CRS}(g) = \{c\}$, where c is separating. We are done if c is a reducing curve (since then we are in case (ii) of Theorem B), so assume not. Let $X_1, X_2 \cong \Sigma_{1,1}$ be the subsurfaces that c bounds in S . We can write $g^2 = h_1 \circ h_2 \circ T_c^n$, where h_i is supported on X_i , and $h_i|_{X_i} \in \text{Mod}(X_i)$ is either the identity or pseudo-Anosov. Suppose that X_1 is primitively compressible in V . Then the characteristic compression body Q_{X_1} is a solid torus. From this we conclude that X_1 supports a unique primitive disk D , and hence either $\{D\}$ or $\{D, gD\}$ is preserved by g (depending on whether $g(X_1) = X_1$ or $g(X_1) = X_2$). We conclude similarly if X_1 (or X_2) is primitively compressible in \widehat{V} . Thus it remains to consider the case X_1 and X_2 are primitively incompressible in both V and \widehat{V} . Since c is not reducing, we can assume (without loss of generality) that c does not bound a disk in V . Since $T_c^n \notin \mathbb{G}$ (by Fact 1), we can assume that h_1 , say, is pseudo-Anosov. Then X_1 is an infinite-diameter hole for $\mathcal{P}(V)$ that is preserved by g^2 .

Case (B). Suppose that $\text{CRS}(g) = \{c\}$, where c is nonseparating. Let $X = S \setminus n(c) \cong \Sigma_{1,2}$. We can write $g^2 = h \circ T_c^n$, where h is supported on X and preserves each component of ∂X , and $h|_X \in \text{Mod}(X)$ is either the identity or a pseudo-Anosov. By Fact 1, $T_c^n \notin \mathbb{G}$, so $h \neq \text{Id}$. If c does not bound a primitive disk, then X is a hole for $\mathcal{P}(V)$, and X has infinite diameter since $g|_X = h$ is pseudo-Anosov. Since $X \cong \Sigma_{1,2}$, this contradicts Theorem D (classification of holes). Therefore, c bounds a primitive disk, which is fixed by g .

Case (C). Suppose that $\text{CRS}(g) = \{c_1, c_2\}$, where c_1 is separating and c_2 is nonseparating. Let $X, Y \cong \Sigma_{1,1}$ be the subsurfaces that c_1 bounds, and assume Y contains c_2 . We can write $g^2 = h \circ T_{c_1}^{n_1} \circ T_{c_2}^{n_2}$, where h preserves X , and $h|_X \in \text{Mod}(X)$ is either pseudo-Anosov or the identity. We consider these cases separately.

First assume that h is pseudo-Anosov. Then X has infinite diameter with respect to $\mathcal{P}(V)$. Observe that Y must support a primitive disk; otherwise X is an infinite diameter hole for $\mathcal{P}(V)$, which implies (by Theorem D and §3.6) that g acts as a pseudo-Anosov on both X and Y contrary to our assumption about $\text{CRS}(g)$. Let D be a primitive disk supported in Y . The canonical compression body Q_Y is a solid torus preserved by g , and D is the unique disk (up to isotopy) in this solid torus, so $g(D) = D$.

If $h|_X \in \text{Mod}(X)$ is the identity, then g^2 is a multitwist. By Fact 1, c_1 and c_2 either bound disks in V or bound an essential annulus in V , and the same holds for \widehat{V} . In any case, (after properly orienting c_1, c_2) the homology class $c_1 - c_2 \in H_1(S)$ is a

nonzero element in the kernel of the homomorphism $H_1(S) \rightarrow H_1(V) \oplus H_1(\widehat{V})$. By Lemma 2.21 this implies that $H_2(S^3) \neq 0$, a contradiction.

Case (D). Suppose that $\text{CRS}(g) = \{c_1, c_2\}$, where both c_1 and c_2 are nonseparating. Since g preserves $\text{CRS}(g)$, we are done if at least one of c_1 and c_2 bounds a primitive disk in V or \widehat{V} , so we assume that neither bounds a primitive disk. Denoting $X = S \setminus n(c_1 \cup c_2)$, we can write $g^4 = h \circ T_{c_1}^{n_1} \circ T_{c_2}^{n_2}$, where h is supported on X and preserves each component of ∂X , and $h|_X \in \text{Mod}(X)$ is either pseudo-Anosov or the identity. In fact, $h|_X$ is not pseudo-Anosov because otherwise $X \cong \Sigma_{0,4}$ would be an infinite-diameter hole for $\mathcal{P}(V)$, contrary to Theorem D. Then g^4 is a multitwist, and by applying Fact 1 in the same way as in Case (C), we conclude that $H_2(S^3) \neq 0$, a contradiction.

Cases (E) and (F). Suppose that $\text{CRS}(g) = \{c_1, c_2, c_3\}$, as in (E) or (F). Each complementary component of $\text{CRS}(g)$ in S is a pair of pants, which has finite mapping class group. Thus a power of g is a nontrivial multitwist about the curves in $\text{CRS}(g)$. Then, as in Cases (C) and (D), we apply Fact 1 arrive at a contradiction. Here there are two cases. In the first case c_1 , say, bounds disks in both V and \widehat{V} , and c_2 and c_3 bound annuli in each of V and \widehat{V} ; this implies that c_1 is separating, so both c_2, c_3 are nonseparating, and this implies that $c_2 - c_3$ is a nontrivial element in the kernel of $H_1(S) \rightarrow H_1(V) \oplus H_1(\widehat{V})$. In the second case, without loss of generality, c_1 bounds a disk in V , c_3 bounds a disk in \widehat{V} , c_2 and c_3 bound an annulus in V , and c_1 and c_2 bound an annulus in \widehat{V} . At most one of the c_i is separating, so $c_1 - c_2 + c_3$ is a nontrivial element of the kernel of $H_1(S) \rightarrow H_1(V) \oplus H_1(\widehat{V})$.

This completes the proof of Theorem B. \square

A Classification of genus-1 fibered knots

Recall that a knot $K \subset S^3$ is *genus-1 fibered* if $S^3 \setminus K$ fibers as a once-punctured torus bundle over the circle. In this appendix we give a proof of the following classical result, first proved in [GAn70].

Theorem A.1 (González-Acuña). *The genus-1 fibered knots $K \subset S^3$ are the trefoil, its mirror, and the figure-8 knot.*

The author has not been able to access [GAn70] and does not know the proof given there. The proof below is almost surely different because (for example) we use Casson's invariant, which was defined in 1985, after the publication of [GAn70]. One might consider the argument below to be a “modern proof”.

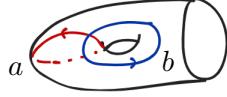


Figure 30: Surface $F = \Sigma_{1,1}$ and a homology basis a, b .

First we reformulate the result. Let $F = \Sigma_{1,1}$, and denote $\mathring{F} = F \setminus \partial F$. By the discussion in §2.6, given a genus-1 fibering

$$\mathring{F} \rightarrow S^3 \setminus K \rightarrow S^1$$

there is a Heegaard decomposition

$$S^3 = (F \times I) \cup_\phi (F \times I)$$

whose gluing, which matches horizontal boundaries of the two I -bundles, is encoded by a mapping class $\phi \in \text{Mod}(F)$. See Figure 12. Under a homeomorphism $M_\phi \cong S^3$, the circle $(\partial F) \times 1$ corresponds to a fibered knot. To prove Theorem A.1, we consider all 3-manifolds of the form $M_\phi = (F \times I) \cup_\phi (F \times I)$, determine when $M_\phi \cong S^3$, and determine the knot $(\partial F) \times 1$ in S^3 .

For clarity, we denote the two I -bundles by $F \times I$ and $\widehat{F} \times I$. Fix curves a, b giving a homology basis for $F = \Sigma_{1,1}$ as pictured in Figure 30. We write T_a, T_b for the (right) Dehn twists about these curves.

We deduce Theorem A.1 from Theorem A.2.

Theorem A.2 (Building S^3 from I -bundles). *Fix $F = \Sigma_{1,1}$ and $\phi \in \text{Mod}(F)$. Let $M_\phi = (F \times I) \cup_\phi (\widehat{F} \times I)$ be the closed 3-manifold described above. If $M_\phi \cong S^3$ then ϕ is conjugate in $\text{Mod}(\Sigma_{1,1})$ to either $T_a T_b^{-1}$ or $(T_a T_b)^{\pm 1}$. The respective fibered knots are the figure-8 and the trefoil and its mirror.*

Remark A.3. The I -bundle structures on $M_{T_a T_b}$ and $M_{(T_a T_b)^{-1}}$ on S^3 differ by an orientation-reversing homeomorphism (Remark 2.20). This explains why the corresponding fibered knots are mirrors.

Remark A.4. Recall (e.g. [FM12, §3.6]) that $\text{Mod}(F)$ is a central extension

$$1 \rightarrow \mathbb{Z} \rightarrow \text{Mod}(F) \xrightarrow{p} \text{Mod}(\mathring{F}) \rightarrow 1.$$

The kernel of p is generated by the boundary Dehn twist $T_{\partial F}$. Furthermore, there is an isomorphism $\text{Mod}(\mathring{F}) \cong \text{SL}_2(\mathbb{Z})$ given by the action on $H_1(\mathring{F}) \cong \mathbb{Z}^2$. In particular, working in the basis a, b ,

$$p(T_a T_b^{-1}) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad p(T_a T_b) = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}.$$

The matrix $p(T_a T_b)$ has order 6, and in $\text{Mod}(F)$ we have the relation $(T_a T_b)^6 = T_{\partial F}$.

Proof of Theorem A.2.

Step 1 (homology 3-spheres). To reduce our search, we first show that if $H_1(M_\phi) = 0$, then the action of ϕ on $H_1(F)$ is conjugate in $\mathrm{SL}_2(\mathbb{Z})$ to one of

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}^{\pm 1}. \quad (9)$$

Apply Mayer–Vietoris to the decomposition $M = (F \times I) \cup (\widehat{F} \times I)$. Denoting $\Sigma \cong \Sigma_2$ be the common boundary of $F \times I$ and $\widehat{F} \times I$, we have the following exact sequence

$$H_1(\Sigma) \xrightarrow{j} H_1(F \times I) \oplus H_1(\widehat{F} \times I) \xrightarrow{s} H_1(M_\phi) \xrightarrow{0} H_0(\Sigma).$$

It is not hard to show that the connecting homomorphism $H_1(M_\phi) \rightarrow H_0(\Sigma)$ is zero. Then $H_1(M_\phi) = 0$ if and only if j is surjective, and since j is a homomorphism $\mathbb{Z}^4 \rightarrow \mathbb{Z}^4$, j is surjective if and only if j is an isomorphism.

Next we express j in coordinates. Denote by $a_1, b_1 \subset \widehat{F} \times 1 \cong \widehat{F} = F$ the curves in Figure 30, which give a basis for $H_1(\widehat{F} \times I)$. We use $\phi^{-1}(a_1), \phi^{-1}(b_1) \subset F \times 1$ as basis for $H_1(F \times I)$. Let $a_0, b_0 \subset \widehat{F} \times 0$ denote the parallel copies of a_1, b_1 , and observe that $a_0, b_0, a_1, b_1 \subset \partial(\widehat{F} \times I) = \Sigma$ is a basis for $H_1(\Sigma)$. With respect to these choices, j has matrix

$$\begin{pmatrix} I & I \\ \tau & I \end{pmatrix},$$

where I denotes the 2×2 identity matrix, and τ is the matrix for the action of ϕ on $H_1(F)$. Then j is an isomorphism if and only if

$$\pm 1 = \det \begin{pmatrix} I & I \\ \tau & I \end{pmatrix} = \det(I - \tau) = 2 - \mathrm{tr}(\tau).$$

Consequently, either $\mathrm{tr}(\tau) = 3$ or $\mathrm{tr}(\tau) = 1$. It is well-known that this implies τ is conjugate to one of the matrices in (9). We briefly explain this below.

If $\mathrm{tr}(\tau) = 3$, then τ is conjugate in $\mathrm{SL}_2(\mathbb{Z})$ to $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. This is because τ is a hyperbolic matrix with characteristic polynomial $x^2 - 3x + 1$, and conjugacy classes of these matrices correspond to narrow ideal classes in $\mathbb{Z}[\lambda]$, where λ is a root of the characteristic polynomial [Wal84]. Here $\mathbb{Z}[\lambda] = \mathbb{Z}\left[\frac{1+\sqrt{5}}{2}\right]$ is the ring of integers in $\mathbb{Q}(\sqrt{5})$. It is well-known that the narrow class number of this ring is 1.

If $\mathrm{tr}(\tau) = 1$, then since τ satisfies its characteristic polynomial $x^2 - x + 1$, it follows that τ has order 6. From the action of $\mathrm{SL}_2(\mathbb{Z})$ on the dual tree to the Farey graph, it follows τ is conjugate to either $\begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}$ or its inverse.

Step 2 (S^3 recognition). As noted in Remark A.4, elements of $\text{Mod}(F)$ that act on the same on $H_1(F)$ differ by the central Dehn twist $T_{\partial F}$. We want to show two statements:

- (i) If $\phi = T_a T_b^{-1}$ or $(T_a T_b)^{\pm 1}$, then $M_\phi \cong S^3$. The fibered knots in these cases are the figure-8 and the trefoil, respectively.
- (ii) If ϕ is as in (i) and $\psi = \phi \circ T_{\partial F}^n$ where $n \neq 0$, then $M_\psi \not\cong S^3$.

(i) Since M_ϕ and M_ϕ^{-1} are homeomorphic (Remark A.3), it suffices to prove (i) for $\phi = T_a T_b$ and $\phi = T_a T_b^{-1}$. We can conclude quickly by showing that M has a Heegaard diagram that is equivalent to the standard Heegaard diagram for S^3 .

We first explain this for $\phi = T_a T_b$. We start with the Heegaard diagram, given by the collection of simple closed curves $(x_1, x_2; y_1, y_2)$ on Σ pictured in Figure 32 (left). Now observe that the curves $(z, x_2; y_1, y_2)$ provide another Heegaard diagram for M_ϕ , one that is equivalent to the standard Heegaard for S^3 . To properly understand these claims and Figure 32, see Figure 31, which explains how we are translating between a curve the “standard” genus-2 surface with curves on $\Sigma = \partial(F \times I)$. (In particular, be careful not to confuse Σ with $S \subset S^3$.) In Figure 32, the curves y_1, y_2 are obtained by applying ϕ to x_1, x_2 on the genus-1 subsurface on the “right” side of Σ . Then y_1, y_2 bound vertical disks in $\widehat{F} \times I$ by construction. The curve z bounds a disk in $F \times I$ because it is disjoint from x_1 and x_2 .

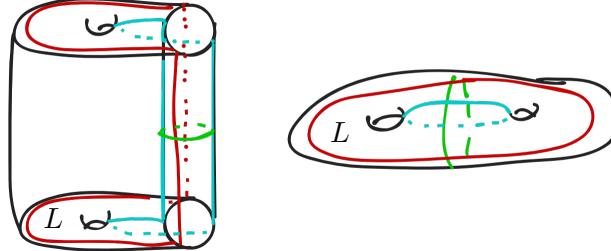


Figure 31: Curves on $\Sigma = \partial(F \times I)$. These curves bound vertical disks.

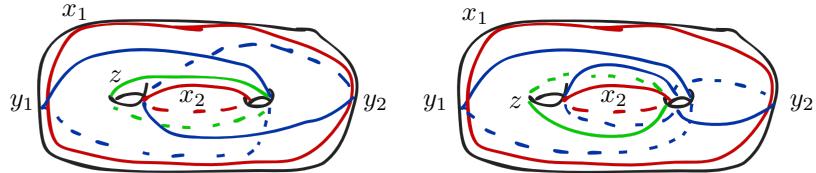


Figure 32: Curves on Σ (not S !) that give Heegaard diagrams for $(F \times I) \cup (\widehat{F} \times I)$. The cases $\phi = T_a T_b$ (left) and $\phi = T_a T_b^{-1}$ (right).

The argument for $\phi = T_a T_b^{-1}$ is similar. We start with the Heegaard diagram $(x_1, x_2; y_1, y_2)$ in Figure 32 (right). Then we observe that $(x_1, x_2; y_1, z)$ is also a Heegaard diagram for M_ϕ , and it is equivalent to the standard Heegaard for S^3 .

The last sentence of (i) will be proved in the process of proving (ii).

(ii) Fix $\phi \in \{T_a T_b^{-1}, (T_a T_b)^{\pm 1}\}$ and $n \neq 0$ and define $\psi = \phi \circ T_{\partial F}^n$. We show M_ψ is not homeomorphic to S^3 . This can be deduced using Casson's invariant $\lambda(M) \in \mathbb{Z}$ for homology 3-spheres and its interpretation in terms of the Torelli group, as we now explain. Replacing the ϕ by ψ changes the gluing changes the gluing of the handlebodies $F \times I, \widehat{F} \times I$ by a separating twist in $\text{Mod}(\Sigma)$. By [Mor89, Lem. 3.4] this implies that M_ψ is obtained from $M_\phi = S^3$ by (-1) -surgery on $K := (\partial F) \times \{1\}$. Then by a theorem of Casson (see [AM90, pg. xii]), one has

$$\lambda(M_\psi) = \frac{1}{2} \Delta''_K(1),$$

where Δ''_K is the second derivative of the Alexander polynomial. Since $\lambda(S^3) = 0$, to show $M_\psi \neq S^3$, it suffices to determine K and show $\Delta''_K(1) \neq 0$.

To identify K , we find a homeomorphism from $M_\phi = (F \times I) \cup (\widehat{F} \times I)$ to the standard Heegaard splitting $S^3 = V \cup \widehat{V}$, and determine the image of K as a curve on $S = \partial V$.

First we treat the case $\phi = T_a T_b^{-1}$. To draw K on V in this case, we first draw K on $\Sigma = \partial(F \times I)$ together with the curves in Figure 32 that give a Heegaard diagram that is equivalent to the standard one. Using this, we can we can easily transport K to $S = \partial V$; the resulting curve is the figure-8 knot. See Figures 33 and 2.

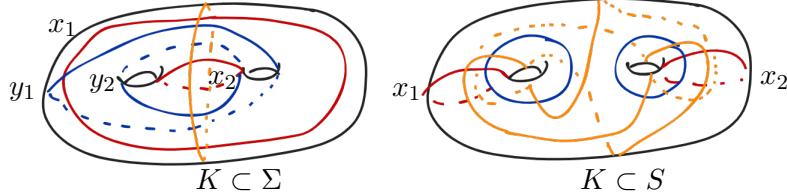


Figure 33: Left: $\Sigma = \partial(F \times I)$, a Heegaard diagram (equivalent to the standard one on S^3), and the curve $K = (\partial F) \times 1$. Right: $S = \partial V$, the standard Heegaard diagram, and the corresponding curve K .

For the K the figure-8 knot, $\Delta_K = 3 - x - x^{-1}$, and $\Delta''_K(1) = -2 \neq 0$. This shows that $M_\psi \not\cong S^3$ when $\psi = \phi \circ T_{\partial}^n$ ($n \neq 0$) and $\phi = T_a T_b^{-1}$.

The case $\phi = (T_a T_b)^{\pm 1}$ can be argued similarly. If we draw $K \subset \Sigma$ on S we arrive at the curve Figure 34, which is the trefoil knot. For K the trefoil knot $\Delta_K = x + x^{-1} - 1$, and $\Delta''_K(1) = 2 \neq 0$ as desired. This shows that $M_\psi \not\cong S^3$ in the case $\psi = \phi \circ T_{\partial}^n$ ($n \neq 0$) and $\phi = (T_a T_b)^{\pm 1}$.

This completes the proof of Theorem A.2. \square



Figure 34: Left: $K = \partial X$ in the case $\phi = (T_a T_b)^{\pm 1}$, pictured on $S = \partial V$. Right: the trefoil knot.

References

- [Akb08] E. Akbas. A presentation for the automorphisms of the 3-sphere that preserve a genus two Heegaard splitting. *Pacific J. Math.*, 236(2):201–222, 2008.
- [AM90] S. Akbulut and J. D. McCarthy. *Casson’s invariant for oriented homology 3-spheres*, volume 36 of *Mathematical Notes*. Princeton University Press, Princeton, NJ, 1990. An exposition.
- [AM21] C. Abbott and J. Manning. Boundaries of coned-off hyperbolic spaces, 2021.
- [BBF15] M. Bestvina, K. Bromberg, and K. Fujiwara. Constructing group actions on quasi-trees and applications to mapping class groups. *Publ. Math. Inst. Hautes Études Sci.*, 122:1–64, 2015.
- [BBKL20] M. Bestvina, K. Bromberg, A. Kent, and C. Leininger. Undistorted purely pseudo-Anosov groups. *J. Reine Angew. Math.*, 760:213–227, 2020.
- [Bes04] M. Bestvina. Questions in geometric group theory. <http://www.math.utah.edu/~bestvina>, July 2004.
- [BH99] M. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*, volume 319 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.
- [BLM83] J. Birman, A. Lubotzky, and J. McCarthy. Abelian and solvable subgroups of the mapping class groups. *Duke Math. J.*, 50(4):1107–1120, 1983.
- [Cho08] S. Cho. Homeomorphisms of the 3-sphere that preserve a Heegaard splitting of genus two. *Proc. Amer. Math. Soc.*, 136(3):1113–1123, 2008.
- [CKL20] S. Cho, Y. Koda, and J. H. Lee. Disk surgery and the primitive disk complexes of the 3-sphere. *Topology Appl.*, 272:107092, 6, 2020.
- [DKL14] S. Dowdall, A. Kent, and C. Leininger. Pseudo-Anosov subgroups of fibered 3-manifold groups. *Groups Geom. Dyn.*, 8(4):1247–1282, 2014.

- [FM02] B. Farb and L. Mosher. Convex cocompact subgroups of mapping class groups. *Geom. Topol.*, 6:91–152, 2002.
- [FM12] B. Farb and D. Margalit. *A primer on mapping class groups*, volume 49 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 2012.
- [GAn70] F. González-Acuña. Dehn’s construction on knots. *Bol. Soc. Mat. Mexicana (2)*, 15:58–79, 1970.
- [GMRS98] R. Gitik, M. Mitra, E. Rips, and M. Sageev. Widths of subgroups. *Trans. Amer. Math. Soc.*, 350(1):321–329, 1998.
- [Goe33] L. Goeritz. Die abbildungen der brezelfläche und der vollbrezel vom geschlecht 2. *Abh. Math. Sem. Univ. Hamburg*, 9(1):244–259, 1933.
- [Gor87] C. McA. Gordon. On primitive sets of loops in the boundary of a handlebody. *Topology Appl.*, 27(3):285–299, 1987.
- [Hal49] M. Hall, Jr. Coset representations in free groups. *Trans. Amer. Math. Soc.*, 67:421–432, 1949.
- [Ham05] U. Hamenstädt. Word hyperbolic extensions of surface groups. Preprint, arXiv:math.GT/0505244, 2005.
- [Hat76] A. Hatcher. Homeomorphisms of sufficiently large P^2 -irreducible 3-manifolds. *Topology*, 15(4):343–347, 1976.
- [KL07] A. Kent and C. Leininger. Subgroups of mapping class groups from the geometrical viewpoint. In *In the tradition of Ahlfors-Bers. IV*, volume 432 of *Contemp. Math.*, pages 119–141. Amer. Math. Soc., Providence, RI, 2007.
- [KL08] A. Kent and C. Leininger. Shadows of mapping class groups: capturing convex cocompactness. *Geom. Funct. Anal.*, 18(4):1270–1325, 2008.
- [KLS09] A. Kent, C. Leininger, and S. Schleimer. Trees and mapping class groups. *J. Reine Angew. Math.*, 637:1–21, 2009.
- [KMT17] T. Koberda, J. Mangahas, and S. Taylor. The geometry of purely loxodromic subgroups of right-angled Artin groups. *Trans. Amer. Math. Soc.*, 369(11):8179–8208, 2017.
- [Min11] H. Min. Hyperbolic graphs of surface groups. *Algebr. Geom. Topol.*, 11(1):449–476, 2011.
- [MM00] H. Masur and Y. Minsky. Geometry of the complex of curves. II. Hierarchical structure. *Geom. Funct. Anal.*, 10(4):902–974, 2000.
- [MM04] H. Masur and Y. Minsky. Quasiconvexity in the curve complex. In *In the tradition of Ahlfors and Bers, III*, volume 355 of *Contemp. Math.*, pages 309–320. Amer. Math. Soc., Providence, RI, 2004.
- [MMS12] H. Masur, L. Mosher, and S. Schleimer. On train-track splitting sequences. *Duke Math. J.*, 161(9):1613–1656, 2012.
- [Mor89] S. Morita. Casson’s invariant for homology 3-spheres and characteristic classes of surface bundles. I. *Topology*, 28(3):305–323, 1989.

- [Mos97] L. Mosher. A hyperbolic-by-hyperbolic hyperbolic group. *Proc. Amer. Math. Soc.*, 125(12):3447–3455, 1997.
- [Mos06] L. Mosher. Problems in the geometry of surface group extensions. In *Problems on mapping class groups and related topics*, volume 74 of *Proc. Sympos. Pure Math.*, pages 245–256. Amer. Math. Soc., Providence, RI, 2006.
- [MS13] H. Masur and S. Schleimer. The geometry of the disk complex. *J. Amer. Math. Soc.*, 26(1):1–62, 2013.
- [MT16] J. Mangahas and S. Taylor. Convex cocompactness in mapping class groups via quasiconvexity in right-angled Artin groups. *Proc. Lond. Math. Soc. (3)*, 112(5):855–881, 2016.
- [Oer02] U. Oertel. Automorphisms of three-dimensional handlebodies. *Topology*, 41(2):363–410, 2002.
- [Pig06] A. Piggott. Palindromic primitives and palindromic bases in the free group of rank two. *J. Algebra*, 304(1):359–366, 2006.
- [Sch82] M. Scharlemann. The complex of curves on nonorientable surfaces. *J. London Math. Soc. (2)*, 25(1):171–184, 1982.
- [Sch04] M. Scharlemann. Automorphisms of the 3-sphere that preserve a genus two Heegaard splitting. *Bol. Soc. Mat. Mexicana (3)*, 10(Special Issue):503–514, 2004.
- [Sch20] S. Schleimer. Notes on the complex of curves. Online lecture notes, 2020.
- [Ser80] J.-P. Serre. *Trees*. Springer-Verlag, Berlin-New York, 1980. Translated from the French by John Stillwell.
- [SW79] P. Scott and T. Wall. Topological methods in group theory. In *Homological group theory (Proc. Sympos., Durham, 1977)*, volume 36 of *London Math. Soc. Lecture Note Ser.*, pages 137–203. Cambridge Univ. Press, Cambridge-New York, 1979.
- [Wal84] D. I. Wallace. Conjugacy classes of hyperbolic matrices in $\mathrm{Sl}(n, \mathbf{Z})$ and ideal classes in an order. *Trans. Amer. Math. Soc.*, 283(1):177–184, 1984.
- [Whi00] K. Whittlesey. Normal all pseudo-Anosov subgroups of mapping class groups. *Geom. Topol.*, 4:293–307, 2000.
- [Zie65] H. Zieschang. Simple path systems on full pretzels. *Mat. Sb. (N.S.)*, 66 (108):230–239, 1965.

Notation guide

V, \hat{V}	genus-2 handlebodies in Heegaard splitting of S^3
S	genus-2 Heegaard surface embedded in S^3
$\Sigma_{g,n}$	compact surface of genus g with n boundary components
$\mathcal{C}(S)$	curve complex of S

$\mathcal{D}(V)$	disk complex of V
$\mathcal{P}(V)$	complex of primitive disks in V
$\mathcal{R}(V, \hat{V})$	reducing sphere complex
D, E	primitive disks in V
\hat{D}, \hat{E}	primitive disks in \hat{V}
\mathbb{G}	the genus-2 Goeritz group
$n(c)$	regular neighborhood of (multi)curve $c \subset S$
T_c	the right Dehn twist about simple closed curve c
X, Y, Z	essential subsurfaces of S
$i(a, b)$	geometric intersection number of simple closed curves a, b
(B, a, b)	surgery bigon/boundary compression
$\alpha, \beta, \gamma, \delta$	standard generators for the genus-2 Goeritz group
$d_{\mathcal{C}(S)}(a, b)$	distance in the 1-skeleton of the curve complex $\mathcal{C}(S)$
$d_X(a, b)$	subsurface projection distance
T	an I -bundle embedded in V
$\partial_h T, \partial_v T$	horizontal and vertical boundaries of I -bundle T

Bena Tshishiku
 Department of Mathematics, Brown University
Email address: `bena_tshishiku@brown.edu`