Algebraic & Geometric Topology

Volume 25 (2025)

Issue 1 (pages 1–644)

msp

ALGEBRAIC & GEOMETRIC TOPOLOGY

msp.org/agt

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

The subscription price for 2025 is US $760/year for the electronic version, and $1110/year (+$75, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP. Algebraic & Geometric Topology is indexed by Mathematical Reviews, Zentralblatt MATH, Current Mathematical Publications and the Science Citation Index.

Algebraic & Geometric Topology (ISSN 1472-2747 printed, 1472-2739 electronic) is published 9 times per year and continuously online, by Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840. Periodical rate postage paid at Oakland, CA 94615-9651, and additional mailing offices. POSTMASTER: send address changes to Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840.

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

https://msp.org/

Algebraic & Geometric Topology **25**:1 (2025) 1–38
DOI: 10.2140/agt.2025.25.1
Published: 5 March 2025

Cutting and pasting in the Torelli subgroup of $\mathrm{Out}(F_n)$

JACOB LANDGRAF

Using ideas from 3-manifolds, Hatcher–Wahl defined a notion of automorphism groups of free groups with boundary. We study their Torelli subgroups, adapting ideas introduced by Putman for surface mapping class groups. Our main results show that these groups are finitely generated, and also that they satisfy an appropriate version of the Birman exact sequence.

57K20, 57K30, 57M07

1 Introduction

Let $F_n = \langle x_1, \dots, x_n \rangle$ be the free group on n letters, and let $\mathrm{Out}(F_n)$ be the group of outer automorphisms of F_n. In many ways, $\mathrm{Out}(F_n)$ behaves very similarly to $\mathrm{Mod}(\Sigma_{g,b})$, the mapping class group of the surface $\Sigma_{g,b}$ of genus g with b boundary components. For an overview of some of these similarities, see [7] by Bridson and Vogtmann.

One such connection is that they both contain a Torelli subgroup. In the mapping class group, the Torelli subgroup $\mathcal{I}(\Sigma_{g,b}) \subset \mathrm{Mod}(\Sigma_{g,b})$ is defined to be the kernel of the action on $H_1(\Sigma_{n,b}; \mathbb{Z})$ for $b = 0, 1$. In $\mathrm{Out}(F_n)$, we define a similar subgroup,[1] denoted IO_n, as the kernel of the action of $\mathrm{Out}(F_n)$ on $H_1(F_n; \mathbb{Z}) = \mathbb{Z}^n$.

On surfaces with multiple boundary components, there are many possible definitions one might use to define a Torelli subgroup of $\mathrm{Mod}(\Sigma_{g,b})$. Putman [22] defines a Torelli subgroup $\mathcal{I}(\Sigma_{g,b}, P)$ for $b > 1$ requiring the additional data of a partition P of the boundary components. The goal of the current paper is to mirror Putman's procedure to define an "IO_n with boundary".

Let $M_{n,b} = \#_n(S^1 \times S^2) \setminus (b \text{ open 3-disks})$. For simplicity, we will write M_n if $b = 0$. A key property of $M_{n,b}$ is that it has fundamental group F_n. Fix such an identification. The *mapping class group* $\mathrm{Mod}(M_{n,b})$ is the group of orientation-preserving diffeomorphisms of $M_{n,b}$ fixing the boundary pointwise modulo isotopies fixing the boundary pointwise. Letting $\mathrm{Diff}^+(M_{n,b}, \partial M_{n,b})$ be the topological group of orientation-preserving diffeomorphisms fixing the boundary pointwise, we can also write $\mathrm{Mod}(M_{n,b}) = \pi_0(\mathrm{Diff}^+(M_{n,b}, \partial M_{n,b}))$. By a theorem of Laudenbach [19], there is an exact sequence

$$1 \to (\mathbb{Z}/2)^n \to \mathrm{Mod}(M_n) \to \mathrm{Out}(F_n) \to 1, \tag{1}$$

where the map $\mathrm{Mod}(M_n) \to \mathrm{Out}(F_n)$ is given by the action (up to conjugation) on $\pi_1(M_n)$, and the

[1] It is also common to see this group denoted by IA_n, but we wish to reserve this notation for the analogous subgroup of $\mathrm{Aut}(F_n)$.

$(\mathbb{Z}/2)^n$ is generated by sphere twists about n disjointly embedded 2-spheres (see Section 2 for the definition and relevant properties of sphere twists). Recent work of Brendle, Broaddus, and Putman [6] shows that this sequence actually splits as a semidirect product. This exact sequence implies that, modulo a finite group, $\mathrm{Out}(F_n)$ acts on M_n up to isotopy. Therefore, M_n plays almost the same role for $\mathrm{Out}(F_n)$ that $\Sigma_{g,b}$ plays for $\mathrm{Mod}(\Sigma_{g,b})$.

Adding boundary components From Laudenbach's sequence (1), we see that

$$\mathrm{Out}(F_n) \cong \mathrm{Mod}(M_n)/\mathrm{STwist}(M_n),$$

where $\mathrm{STwist}(M_n) \cong (\mathbb{Z}/2)^n$ is the subgroup of $\mathrm{Mod}(M_n)$ generated by sphere twists. Now that we have related $\mathrm{Out}(F_n)$ to a geometrically defined group, we can start introducing boundary components. Extending the relationship given by Laudenbach's sequence, we define "$\mathrm{Out}(F_n)$ with boundary" as

$$\mathrm{Out}(F_{n,b}) := \mathrm{Mod}(M_{n,b})/\mathrm{STwist}(M_{n,b}).$$

When $b = 1$, Laudenbach [19] also shows that $\mathrm{Out}(F_{n,1}) \cong \mathrm{Aut}(F_n)$. Hatcher and Wahl [14] introduced a more general version of $\mathrm{Out}(F_{n,b})$, which they denoted by $A^s_{n,k}$. The original definition of $A^s_{n,k}$ has to do with classes of self-homotopy equivalences of a certain graph. However, in [14] the authors give an equivalent definition, which says that $A^s_{n,k}$ is the mapping class group of M_n with s spherical and k toroidal boundary components, modulo sphere twists. With this definition, we see that $\mathrm{Out}(F_{n,b}) = A^b_{n,0}$. Similar groups have been examined in the work of Jensen and Wahl [16] and Wahl [26]. Their versions, however, involve only toroidal boundary components, and thus are distinct from $\mathrm{Out}(F_{n,b})$.

Torelli subgroups An important feature of sphere twists (discussed in Section 2) is that they act trivially on homotopy classes of embedded loops, and thus act trivially on $H_1(M_n)$. Therefore, the action of $\mathrm{Mod}(M_{n,b})$ on $H_1(M_{n,b})$ induces an action of $\mathrm{Out}(F_{n,b})$ on $H_1(M_{n,b})$. We can then define the Torelli subgroup $\mathrm{IO}_{n,b} \subset \mathrm{Out}(F_{n,b})$ to be the kernel of this action. However, this definition does not capture all homological information when $b > 1$, especially when $M_{n,b}$ is being embedded in $M_{m,c}$. To see why, consider the scenario depicted in Figure 1, in which $M_{2,2}$ has been embedded into M_4. This embedding induces a homomorphism $\iota_M\colon \mathrm{Mod}(M_{2,2}) \to \mathrm{Mod}(M_4)$ obtained by extending by the identity. This map sends sphere twists to sphere twists, and so we get an induced map $\iota_*\colon \mathrm{Out}(F_{2,2}) \to \mathrm{Out}(F_4)$. However, this does *not* restrict to a map $\mathrm{IO}_{2,2} \to \mathrm{IO}_4$ under this definition of $\mathrm{IO}_{n,b}$ since elements of $\mathrm{IO}_{2,2}$ are not required to fix the homology class of the subarc of α lying inside $M_{2,2}$. To address this issue, we will use a slightly modified homology group.

Definition Fix a partition P of the boundary components of $M_{n,b}$.

(a) Two boundary components ∂_1, ∂_2 of $M_{n,b}$ are *P-adjacent* if there is some $p \in P$ such that $\{\partial_1, \partial_2\} \subset p$.

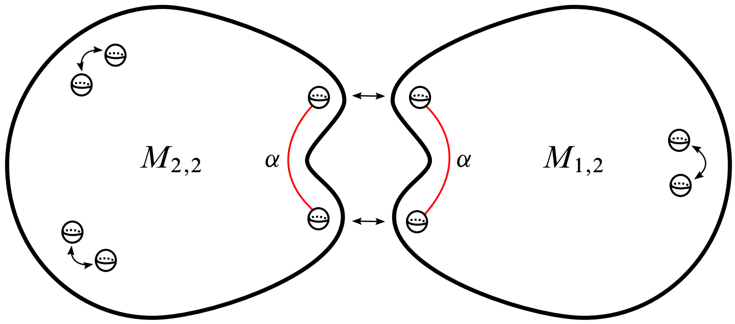


Figure 1: A copy of $M_{2,2}$ and $M_{1,2}$ glued together to obtain M_4. We realize $M_{2,2}$ as a 3-sphere with the six indicated open balls removed, then the boundaries of these removed balls are identified according to the arrows (and similarly for $M_{1,2}$). The class $[\alpha]$ need not be fixed by elements of $\mathrm{IO}_{2,2}$ with the naive definition.

(b) Let $H_1^P(M_{n,b})$ be the subgroup of $H_1(M_{n,b}, \partial M_{n,b})$ spanned by

$$\big\{[h] \in H_1(M_{n,b}, \partial M_{n,b}) \mid \text{ either } h \text{ is a simple closed curve or } h \text{ is a properly embedded arc with endpoints in distinct } P\text{-adjacent boundary components}\big\}.$$

(c) There is a natural action of $\mathrm{Out}(F_{n,b})$ on $H_1^P(M_{n,b})$, and we define the Torelli subgroup $\mathrm{IO}_{n,b}^P \subset \mathrm{Out}(F_{n,b})$ to be the kernel of this action.

Returning to Figure 1, let P be the trivial partition of the boundary components of $M_{2,2}$ with a single P-adjacency class. With this choice of partition, we see that $[\alpha \cap M_{2,2}] \in H_1^P(M_{2,2})$. If $f \in \mathrm{IO}_{2,2}^P$, then it follows that $\iota_*(f) \in \mathrm{Out}(F_4)$ preserves the homology class of α. Therefore, $\iota_*(f) \in \mathrm{IO}_4$, and so ι_* restricts to a map $\mathrm{IO}_{2,2}^P \to \mathrm{IO}_4$.

Restriction As we discussed in the last paragraph, given an embedding $\iota\colon M_{n,b} \hookrightarrow M_m$, we can extend by the identity to get a map $\iota_*\colon \mathrm{Out}(F_{n,b}) \to \mathrm{Out}(F_m)$. In general, ι_* may not be injective. However, it is injective if no connected component of $M_m \setminus M_{n,b}$ is diffeomorphic to D^3 (see Appendix A). Moreover, such an embedding induces a natural partition of the boundary components of $M_{n,b}$ as follows.

Definition Fix an embedding $\iota\colon M_{n,b} \hookrightarrow M_m$. Let N be a connected component of $M_m \setminus \mathrm{int}(M_{n,b})$, and let p_N be the set of boundary components of $M_{n,b}$ shared with N. Then the partition P of the boundary components of $M_{n,b}$ induced by ι is defined to be

$$P = \{p_N \mid N \text{ a connected component of } M_m \setminus M_{n,b}\}.$$

With this definition, one might guess that $\iota_*^{-1}(\mathrm{IO}_n) = \mathrm{IO}_{n,b}^P$. This turns out to be the case, and this is our first main theorem, which we prove in Section 3.

Theorem A (restriction theorem) *Let* $\iota\colon M_{n,b} \hookrightarrow M_m$ *be an embedding,* $\iota_*\colon \mathrm{Out}(F_{n,b}) \to \mathrm{Out}(F_m)$ *the induced map, and* P *the induced partition of the boundary components of* $M_{n,b}$. *Then* $\mathrm{IO}_{n,b}^P = \iota_*^{-1}(\mathrm{IO}_m)$.

Birman exact sequence From here, we move on to exploring the parallels between these Torelli subgroups and those of mapping class groups. There is a well-known relationship between the mapping class groups of surfaces with a different number of boundary components called the Birman exact sequence (see [12]):

$$1 \to \pi_1(\mathrm{UT}(\Sigma_{n,b-1})) \to \mathrm{Mod}(\Sigma_{g,b}) \to \mathrm{Mod}(\Sigma_{g,b-1}) \to 1.$$

Here, $\mathrm{UT}(\Sigma_{n,b-1})$ is the unit tangent bundle of $\Sigma_{n,b-1}$, the map $\pi_1(\mathrm{UT}(\Sigma_{n,b-1})) \to \mathrm{Mod}(\Sigma_{g,b})$ is given by pushing a boundary component around a loop, and the map $\mathrm{Mod}(\Sigma_{g,b}) \to \mathrm{Mod}(\Sigma_{g,b-1})$ is given by attaching a disk onto this boundary component. In Section 4, we will prove versions of the Birman exact sequence for $\mathrm{Mod}(M_{n,b})$ and $\mathrm{Out}(F_{n,b})$, all culminating in the following sequence for $\mathrm{IO}^P_{n,b}$.

Theorem B (Birman exact sequence) *Fix $n, b > 0$ such that $(n,b) \neq (1,1)$, and let $M_{n,b} \hookrightarrow M_{n,b-1}$ be an embedding obtained by gluing a ball to the boundary component ∂. Fix $x \in M_{n,b-1} \setminus M_{n,b}$. Let P be a partition of the boundary components of $M_{n,b}$, let P' be the induced partition of the boundary components of $M_{n,b-1}$, and let $p \in P$ be the set containing ∂. We then have an exact sequence*

$$1 \to L \to \mathrm{IO}^P_{n,b} \xrightarrow{\iota_*} \mathrm{IO}^{P'}_{n,b-1} \to 1,$$

where L is equal to

(a) $\pi_1(M_{n,b-1}, x) \cong F_n$ *if* $p = \{\partial\}$,

(b) $[\pi_1(M_{n,b-1}, x), \pi_1(M_{n,b-1}, x)] \cong [F_n, F_n]$ *if* $p \neq \{\partial\}$.

Moreover, this sequence splits if $b \geq 2$.

Remark This theorem may seem superficially similar to results proven by Day and Putman [9; 11]. However, we consider a very different notion of "automorphisms with boundary", and so these results are unrelated.

Finite generation Once we have established this version of the Birman exact sequence, in Section 5, we will define a generating set for $\mathrm{IO}^P_{n,b}$. This generating set will be inspired by the generating set for IO_n found by Magnus [21] in 1935.

Theorem 1.1 (Magnus) *Let $F_n = \langle x_1, \dots, x_n \rangle$. The group IO_n is generated by the $\mathrm{Out}(F_n)$-classes of the automorphisms*

$$M_{ij} : x_i \mapsto x_j x_i x_j^{-1}, \quad M_{ijk} : x_i \mapsto x_i [x_j, x_k],$$

for all distinct $i, j, k \in \{1, \dots, n\}$ with $j < k$. Here, the automorphisms are understood to fix x_ℓ for $\ell \neq i$.

Throughout this paper, we will use the convention $[a,b] = aba^{-1}b^{-1}$. Since we defined $\mathrm{IO}^P_{n,b}$ to be a subgroup of $\mathrm{Mod}(M_{n,b})/\mathrm{STwist}(M_{n,b})$, our generators will be defined geometrically rather than algebraically. However, in the case of $b = 0$, they will reduce directly to Magnus's generators. In Section 6, we will show that these elements do indeed generate $\mathrm{IO}^P_{n,b}$.

Theorem C *The group* $\mathrm{IO}^P_{n,b}$ *is finitely generated for* $n \geq 1,\ b \geq 0$*.*

This is rather striking because the analogous result for Torelli subgroups of mapping class groups with multiple boundary components is still open. We will prove this theorem by using the Birman exact sequence to reduce to $b = 0$ and applying Magnus's theorem. Unfortunately, the tools we have constructed do not seem strong enough to give a novel proof of Magnus's theorem. We will, however, prove a weaker version in Section 7. The original proof of Magnus's Theorem 1.1 comes in two steps: showing that the given automorphisms $\mathrm{Out}(F_n)$-normally generate IO_n, and then showing that the subgroup they generate is normal in $\mathrm{Out}(F_n)$. We will give a proof of the first step in our setting. For alternative proofs of the first step, as well as more information on the second step in this context, see work by Bestvina, Bux and Margalit [5] as well as Day and Putman [10].

Theorem D *The group* IO_n *is* $\mathrm{Out}(F_n)$*-normally generated by the automorphisms* M_{ij} *and* M_{ijk}*, where* $i, j, k \in \{1, \dots, n\}$ *and* $j < k$*.*

Abelianization Once we have a finite generating set for $\mathrm{IO}^P_{n,b}$, a natural question arises: how does the cardinality of this generating set compare to the rank of $H_1(\mathrm{IO}^P_{n,b})$? For $b \leq 1$, this question is answered by a result of Andreadakis [1] and Bachmuth [3].

Theorem 1.2 (Andreadakis, Bachmuth) *The abelianization of* $\mathrm{IO}_{n,b}$ *is torsion-free of rank* $n \cdot \binom{n}{2} - n$ *if* $b = 0$*, and rank* $n \cdot \binom{n}{2}$ *if* $b = 1$*.*

This theorem was proved using a version of the Johnson homomorphism

$$\tau \colon \mathrm{IA}_n \to \mathrm{Hom}\big(H, \textstyle\bigwedge^2 H\big),$$

where $H = H_1(F_n) = \mathbb{Z}^n$. We will recall the definition of this homomorphism in Section 8, along with the proof of Theorem 1.2. We then move on to computing the rank of $H_1(\mathrm{IO}^P_{n,b})$ for $b > 1$. To do this, we choose an embedding $M_{n,b} \hookrightarrow M_{m,1}$, which induces an injection $\mathrm{IO}^P_{n,b} \to \mathrm{IO}_{m,1} = \mathrm{IA}_m$. Composing this map with τ gives a map $\tau_* \colon \mathrm{IO}^P_{n,b} \to \mathrm{Hom}\big(H, \bigwedge^2 H\big)$. We then compute the image of our generators under τ_*, and use this to count the rank of $\tau_*(\mathrm{IO}^P_{n,b})$.

Theorem E *The abelianization of* $\mathrm{IO}^P_{n,b}$ *is torsion-free of rank*

$$n \cdot \binom{n}{2} + \left(b \cdot \binom{n}{2} - |P| \cdot \binom{n}{2}\right) + (|P| \cdot n - n).$$

Outline In Section 2, we will give a short overview of sphere twists. We then move on to proving Theorem A in Section 3. We will establish all of our versions of the Birman exact sequence (including Theorem B) in Section 4. In Section 5, we will define our candidate generators for $\mathrm{IO}^P_{n,b}$, and we will prove that they generate (Theorem C) in Section 6 using the Birman exact sequence and Magnus's Theorem 1.1. In Section 7, we will prove Theorem D. We then move on to Section 8, in which we recall the definition of the Johnson homomorphism for IA_n, and use it to compute the rank of the abelianization

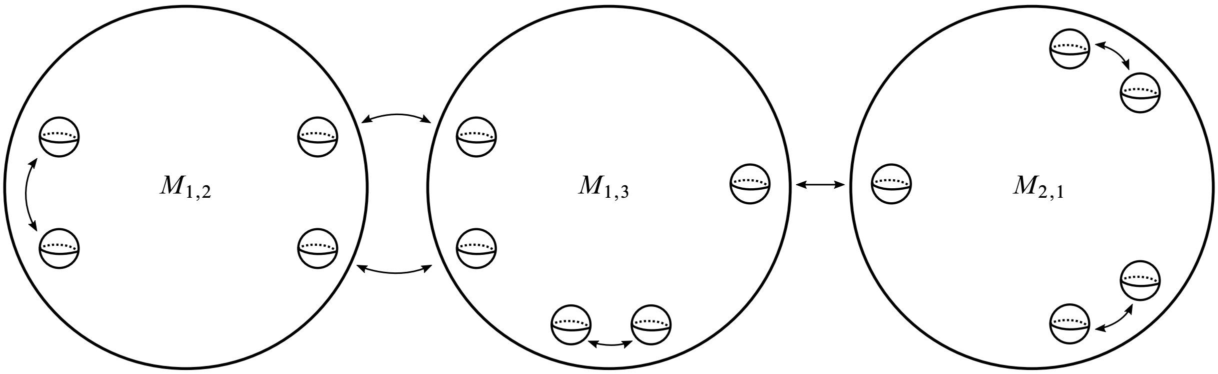


Figure 2: M_5 realized by gluing $M_{1,2}$, $M_{1,3}$, and $M_{2,1}$ together along their boundaries as indicated by the arrows.

of $\mathrm{IO}_{n,b}^P$, proving Theorem E. Finally, we conclude with two appendices. In Appendix A, we provide conditions for a map $\mathrm{Out}(F_{n,b}) \to \mathrm{Out}(F_m)$ induced by an inclusion to be injective, and in Appendix B we prove a lemma which allows us to realize bases of $H_2(M_m)$ as collections of disjoint oriented spheres.

Figure conventions We will frequently direct the reader to figures which are intended to give some geometric intuition for the manifold $M_{n,b}$. In order to assemble $M_{n,b}$, we begin with one or more copies of S^3, remove a collection of open balls, and then glue the resulting boundary components together in pairs. These gluings will be indicated by double-sided arrows connecting the boundary spheres being glued. As an example, see Figure 2.

Acknowledgements I would like to thank my advisor Andy Putman for directing me to $\mathrm{Out}(F_n)$ and its Torelli subgroup, and for his input during the revision process. I would also like to thank Patrick Heslin and Aaron Tyrrell for helpful conversations regarding diffeomorphism groups, as well as Dan Margalit for an enlightening question which resulted in the addition of Section 8.

This material is based upon work supported by the National Science Foundation under grant number DMS-1547292.

2 Preliminaries

Since sphere twists play a fundamental role throughout the remainder of the paper, we will give a brief overview of them here.

Sphere twists Fix a smoothly embedded 2-sphere $S \subset M_{n,b}$, and let $U \cong S \times [0,1]$ be a tubular neighborhood of S. Recall that $\pi_1(\mathrm{SO}(3), \mathrm{id}) \cong \mathbb{Z}/2\mathbb{Z}$, and the nontrivial element $\gamma\colon [0,1] \to \mathrm{SO}(3)$ is given by rotating $\mathbb{R}^3$ one full revolution about any fixed axis through the origin. Fix an identification $S = S^2 \subset \mathbb{R}^3$. Then, we define the sphere twist about S, denoted $T_S \in \mathrm{Mod}(M_{n,b})$, to be the class of the diffeomorphism which is the identity on $M_{n,b} \setminus U$ and is given by $(x,t) \mapsto (\gamma(t) \cdot x, t)$ on $U \cong S \times [0,1]$.

The isotopy class of T_S depends only on the isotopy class of S. In fact, more is true: Laudenbach [19] showed that the class of T_S depends only on the homotopy class of S.

Action on curves and surfaces Since $\pi_1(\mathrm{SO}(3), \mathrm{id}) \cong \mathbb{Z}/2\mathbb{Z}$, we see that sphere twists have order at most two. However, it is tricky to show that sphere twists are actually nontrivial because they act trivially on homotopy classes of embedded arcs and surfaces. To see why this is true, let $S \subset M_{n,b}$ be an embedded 2-sphere, and let $U = S \times [0, 1]$ be a tubular neighborhood of S. Suppose that α is either an arc or surface embedded in $M_{n,b}$ (we will handle both cases simultaneously). We can homotope α such that it is either disjoint from U or intersects U transversely. Let $p \in S$ be one of points in S which lies on the axis of rotation used to construct T_S. We can homotope α such that $\alpha \cap U$ collapses into $p \in [0, 1]$. Note that this process is not an isotopy, and α is no longer embedded in $M_{n,b}$. This is not an issue because a result of Laudenbach [19] shows that if α is fixed up to homotopy, then it is fixed up to isotopy. Since T_S fixes $p \times [0, 1]$ pointwise, it follows that T_S fixes α up to homotopy. The upshot of this is that a more sophisticated invariant must be constructed to detect the nontriviality of T_S. In [18; 19], Laudenbach uses framed cobordisms to show that for $b = 0, 1$, the sphere twist T_S is trivial if and only if S is separating. In the case of no boundary components, Brendle, Broaddus, and Putman [6] give another proof of this fact by showing that sphere twists act nontrivially on a trivialization of the tangent bundle of M_n up to isotopy.

Sphere twist subgroup Let $\mathrm{STwist}(M_{n,b}) \subset \mathrm{Mod}(M_{n,b})$ be the subgroup generated by sphere twists. Given $\mathfrak{f} \in \mathrm{Mod}(M_{n,b})$ and a sphere twists T_S, we have the "change of coordinates" formula

$$\mathfrak{f} T_S \mathfrak{f}^{-1} = T_{\mathfrak{f}(S)}.$$

This shows that $\mathrm{STwist}(M_{n,b})$ is a normal subgroup of $\mathrm{Mod}(M_{n,b})$. In fact, even more is true. Letting $\mathfrak{f} = T_{S'}$ in the above formula and using the fact that sphere twists act trivially on embedded surfaces up to isotopy, we find that

$$T_{S'} T_S T_{S'}^{-1} = T_{T_{S'}(S)} = T_S,$$

which implies $\mathrm{STwist}(M_{n,b})$ is actually abelian. Since nontrivial sphere twists have order two, it follows that $\mathrm{STwist}(M_{n,b})$ is isomorphic to a product of copies of $\mathbb{Z}/2\mathbb{Z}$. For $b = 0, 1$, another result of Laudenbach shows that $\mathrm{STwist}(M_{n,b}) \cong (\mathbb{Z}/2\mathbb{Z})^n$ and is generated by the sphere twists about the n core spheres $* \times S^2$ in each $S^1 \times S^2$ summand. For $b > 1$, one can show that $\mathrm{STwist}(M_{n,b}) \cong (\mathbb{Z}/2\mathbb{Z})^{n+b-1}$. The -1 in the exponent reflects the fact that the product of all the sphere twists about boundary components is trivial. Since we will need this fact later, we include a proof here.

Lemma 2.1 *If $S_1, \dots, S_b \subset M_{n,b}$ be spheres parallel to the b boundary components of $M_{n,b}$, then the element $T_{S_1} \cdots T_{S_b}$ is trivial in* $\mathrm{Mod}(M_{n,b})$.

Proof We will prove this by induction on n. As the base case, consider $M_{0,b}$. The argument in this case follows a proof of Hatcher and Wahl [15, pages 214–215], but we include the proof here as well for

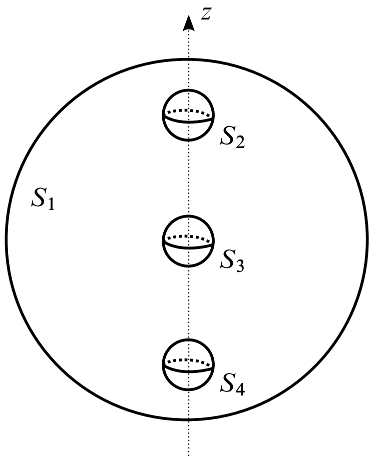


Figure 3: $M_{0,4}$ embedded in $\mathbb{R}^3$.

completeness. If $b = 0$, then the statement is trivial. If $b > 0$, then we can embed $M_{0,b}$ in $\mathbb{R}^3$ as the unit ball with $b-1$ smaller balls removed along the z-axis (see Figure 3). We may then use the z-axis as the axis of rotation for the sphere twists about all the boundary components. Taking S_1 to be the unit sphere, we then see that the product $T_2 \cdots T_b$ is isotopic to T_1. Since sphere twists have order two, this gives the desired relation, and so we have completed the base case.

Next, consider $M_{n,b}$ for $n > 0$. Since $n > 0$, there exists a nonseparating sphere $S \subset M_{n,b}$ which is disjoint from $S_1, \dots, S_b$. Splitting $M_{n,b}$ along S yields a submanifold diffeomorphic to $M_{n-1,b+2}$. Let $\iota_M : \mathrm{Mod}(M_{n-1,b+2}) \to \mathrm{Mod}(M_{n,b})$ be the map induced by inclusion. Let $T_1, \dots, T_{b+2}$ be the sphere twists about the boundary components of $M_{n-1,b+2}$, and order them such that $\iota_M(T_j) = T_{S_j}$ for $0 \le j \le b$. With this ordering, notice that $\iota_M(T_{b+1}) = \iota_M(T_{b+2}) = T_S$. Since sphere twists have order two,

$$\iota_M(T_1 \cdots T_{b+2}) = T_{S_1} \cdots T_{S_b} \cdot T_S^2 = T_{S_1} \cdots T_{S_b}.$$

By our induction hypothesis, $T_1 \cdots T_{b+2}$ is trivial in $\mathrm{Mod}(M_{n-1,b+2})$, and so we are done. □

If $b = 1$, this shows that the sphere twist about the boundary component is trivial. However, if $b > 1$, then the sphere twists about boundary components are nontrivial. We will also need this fact, so we prove it here.

Lemma 2.2 *Let $b > 1$, and let ∂ be a boundary component of $M_{n,b}$. Then $T_\partial \in \mathrm{Mod}(M_{n,b})$ is nontrivial.*

Proof Let ∂' be a boundary component of $M_{n,b}$ different from ∂. Then we get an embedding

$$\iota \colon M_{n,b} \hookrightarrow M_{n+1}$$

by attaching ∂ and ∂' with a copy of $S^2 \times I$, and capping off all the remaining boundary components. Let $\iota_M : \mathrm{Mod}(M_{n,b}) \to \mathrm{Mod}(M_{n+1})$ be the map induced by ι. Then $\iota_M(T_\partial)$ is a sphere twist about a nonseparating sphere. Earlier in this section, we saw that such sphere twists are nontrivial, and so we conclude that T_∂ is nontrivial as well. □

3 Restriction theorem

Fix $n, b \geq 0$, and let P be a partition of the boundary components of $M_{n,b}$. Recall that we have defined $H_1^P(M_{n,b})$ to be the submodule of $H_1(M_{n,b}, \partial M_{n,b})$ generated by

$$\big\{[h] \in H_1(M_{n,b}, \partial M_{n,b}) \mid \text{ either } h \text{ is a simple closed curve or } h \text{ is a properly embedded arc with endpoints in distinct } P\text{-adjacent boundary components}\big\},$$

and $\mathrm{IO}_{n,b}^P$ is the kernel of the action of $\mathrm{Out}(F_{n,b})$ on $H_1^P(M_{n,b})$ induced by the action of $\mathrm{Mod}(M_{n,b})$.

Remark This version of homology is simpler than the one used in [22]. There are two reasons for this.

- In our case, we can take homology relative to the entire boundary, whereas in [22], homology is taken relative to a set consisting of a single point from each boundary component. This is because in surfaces, the boundary components give nontrivial elements of H_1, and the arcs considered in $H_1^P(\Sigma_{g,b})$ can get "wrapped around" those boundary components. This is not a problem in our setting because loops in boundary components of $M_{n,b}$ are trivial in H_1.
- Next, suppose we have an embedding $\Sigma_{g,b} \hookrightarrow \Sigma_{g'}$ of surfaces. It is possible for a nontrivial element $a \in H_1(\Sigma_{g,b})$ to become trivial in $H_1(\Sigma_{g'})$ (for instance, if a boundary component is capped off). So, there could be elements of $\mathrm{Mod}(\Sigma_{g,b})$ which act trivially on $H_1(\Sigma_{g'})$, but not fix a. In other words, the Torelli group would not be closed under restrictions. To fix this, the author in [22] must mod out by the submodules of $H_1(\Sigma_{g,b})$ spanned by the $p \in P$ (with proper orientation chosen). This is not a problem in the 3-dimensional case however, since an inclusion $M_{n,b} \hookrightarrow M_m$ induces an injection $H_1(M_{n,b}) \to H_1(M_m)$.

Proof of Theorem A Let $\iota\colon M_{n,b} \hookrightarrow M_m$ be an embedding, and let $\iota_*\colon \mathrm{Out}(F_{n,b}) \to \mathrm{Out}(F_m)$ be the induced map. Recall that we must show that $\iota_*^{-1}(\mathrm{IO}_m) = \mathrm{IO}_{n,b}^P$, where P is the partition of the boundary components induced by ι as described in the introduction.

This proof will follow the proof of [22, Theorem 3.3]. Define the following subsets of $H_1(M_m)$ (we use $\cdot$ to denote concatenation of arcs):

$$Q_1 = \{[h] \in H_1(M_m) \mid h \text{ is a simple closed curve in } M_m \setminus M_{n,b}\},$$

$$Q_2 = \{[h] \in H_1(M_m) \mid h \text{ is a simple closed curve in } M_{n,b}\},$$

$$Q_3 = \big\{[h_1 \cdot h_2] \in H_1(M_m) \mid h_1 \text{ is a properly embedded arc in } M_{n,b} \text{ with endpoints in distinct } P\text{-adjacent boundary components and } h_2 \text{ is a properly embedded arc in } M_m \setminus M_{n,b} \text{ with the same endpoints as } h_1\big\}.$$

We claim that the homology group $H_1(M_m)$ is spanned by $Q_1 \cup Q_2 \cup Q_3$. To see why, let $[\alpha] \in H_1(M_m)$ be the class of a loop α. If α can be homotoped to lie entirely inside $M_{n,b}$ or $M_m \setminus M_{n,b}$, then we are

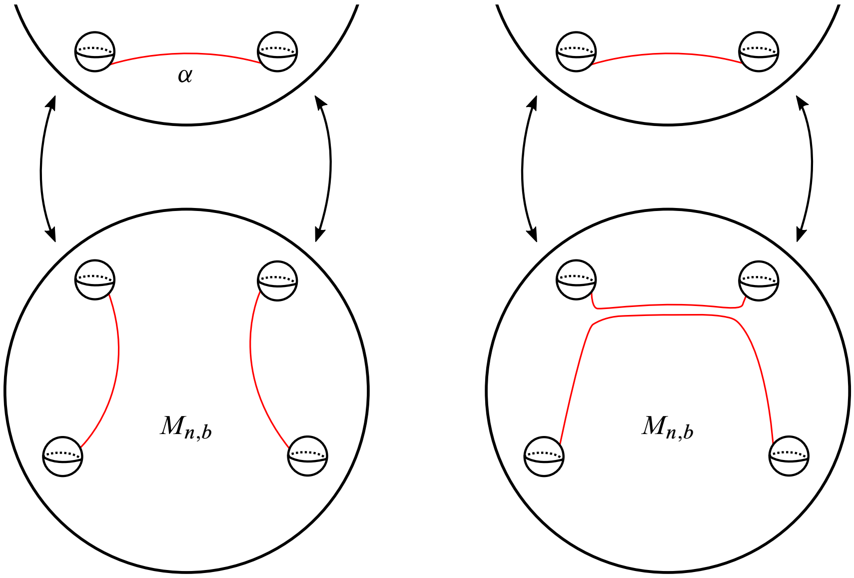


Figure 4: A loop can be surgered into a collection of loops which intersect $\partial M_{n,b}$ exactly twice.

done. On the other hand, suppose that crosses the boundary of $M_{n,b}$. Without loss of generality, we may assume that α crosses the boundary of $M_{n,b}$ exactly twice since any loop can be surgered into a collection of such loops (see Figure 4). It follows that α has the form $\alpha = \gamma \cdot \delta$, where $\gamma \subset M_{n,b}$ is an arc connecting boundary components of $M_{n,b}$, and $\delta \subset M_m \setminus M_{n,b}$ is a arc with the same endpoints as γ. Recall that under the partition P induced by the inclusion ι, two boundary components are P-adjacent if they lie on the same component of $M_m \setminus M_{n,b}$. Therefore, the existence of δ implies that the boundary components intersected by α are P-adjacent, and thus $[\alpha] \in Q_3$. This completes the proof of the claim.

Let $f \in \mathrm{IO}^{P}_{n,b}$. By the definition of $\mathrm{IO}^{P}_{n,b}$, the element $\iota_*(f)$ acts trivially on Q_2. Moreover, $\iota_*(f)$ acts trivially on Q_1 by the definition of ι_*. Lastly, suppose that $[h_1 \cdot h_2] \in Q_3$. Then $\iota_*(f)$ fixes the homology class of h_1 since $f \in \mathrm{IO}^{P}_{n,b}$, and fixes h_2 pointwise by the definition of ι_*. Therefore, $f \in \iota_*^{-1}(\mathrm{IO}_m)$.

Next, suppose that $f \in \iota_*^{-1}(\mathrm{IO}_m)$. By definition, $\iota_*(f)$ acts trivially on $H_1(M_m)$, and thus on Q_2 as well since the map $H_1(M_{n,b}) \to H_1(M_m)$ induced by ι is injective. This implies that f acts trivially on homology classes of simple closed curves in $M_{n,b}$. So, we only need to check that f preserves the homology classes of arcs in M_m connecting distinct P-adjacent boundary components. Suppose there is a class of arcs $[\alpha] \in H_1^P(M_{n,b})$. Since P is the partition of the boundary components induced by ι, $[\alpha]$ can be completed to a homology class $[\alpha \cdot \beta] \in H_1(M_m)$, where β is an arc in $M_m \setminus M_{n,b}$ connecting the endpoints of α. Then since $\iota_*(f) \in \mathrm{IO}_m$ and $\iota_*(f)$ fixes β pointwise, we have

$$0 = ([\alpha \cdot \beta]) - \iota_*(f)([\alpha \cdot \beta]) = [\alpha] - f([\alpha]).$$

This shows that f acts trivially on $[\alpha]$. Therefore, $f \in \mathrm{IO}^{P}_{n,b}$. □

4 Birman exact sequence

In this section, we give a version of the Birman exact sequence for the groups $\mathrm{IO}^{P}_{n,b}$. We will start by giving a Birman exact sequence on the level of mapping class groups. We note that Banks has proved a version of the Birman exact sequence for 3-manifolds (see [4]). However, this version involves forgetting a puncture rather than capping a boundary component, so we will prove our own version here. Once we have the sequence for mapping class groups, we will mod out by sphere twists to get a corresponding sequence for $\mathrm{Out}(F_{n,b})$, and finally restrict to get a sequence for $\mathrm{IO}^{P}_{n,b}$.

Remark In the following theorems, we exclude the case $(n,b)=(1,1)$. This is because boundary drags in $\mathrm{Mod}(M_{1,1})$ are trivial (see the proof of Theorem 4.1 for the definition of boundary drags). In this case, we have isomorphisms

$$\mathrm{Mod}(M_{1,1})\cong\mathrm{Mod}(M_1)\cong\mathbb{Z}/2\times\mathbb{Z}/2,\quad \mathrm{Out}(F_{1,1})\cong\mathrm{Out}(F_1)\cong\mathbb{Z}/2,\quad \mathrm{IO}^{\{\partial\}}_{1,1}\cong\mathrm{IO}_1\cong 1,$$

where one of the generators of $\mathrm{Mod}(M_1)=\mathrm{Mod}(S^1\times S^2)$ is a sphere twist about the sphere $*\times S^2$ and the other is the antipodal map in both coordinates.

Theorem 4.1 *Fix $n,b>0$ such that $(n,b)\neq(1,1)$. Glue a ball to a boundary component of $M_{n,b}$, and let $M_{n,b}\hookrightarrow M_{n,b-1}$ be the resulting embedding. Fix $x\in M_{n,b-1}\setminus M_{n,b}$.*

(a) *If $b>1$, choose a lift $\tilde{x}\in\mathrm{Fr}_x(M_{n,b-1})$ of x, where $\mathrm{Fr}(M_{n,b-1})$ is the oriented frame bundle of $M_{n,b-1}$. We then have an exact sequence*

$$1\to\pi_1(\mathrm{Fr}(M_{n,b-1}),\tilde{x})\to\mathrm{Mod}(M_{n,b})\to\mathrm{Mod}(M_{n,b-1})\to 1.$$

(b) *If $b=1$ and $n>1$, then we have an exact sequence*

$$1\to\pi_1(M_{n,b-1},x)\to\mathrm{Mod}(M_{n,b})\to\mathrm{Mod}(M_{n,b-1})\to 1.$$

Proof Let $\mathrm{Diff}^+(M_{n,b-1})$ denote the space of orientation-preserving diffeomorphisms of $M_{n,b-1}$ which restrict to the identity on $\partial M_{n,b-1}$, and let $\mathrm{Diff}^+(M_{n,b-1},\tilde{x})$ be the subspace of $\mathrm{Diff}^+(M_{n,b-1})$ consisting of diffeomorphisms which fix the framing $\tilde{x}$. It is standard that there is a fiber bundle

$$\mathrm{Diff}^+(M_{n,b-1},\tilde{x})\to\mathrm{Diff}^+(M_{n,b-1})\xrightarrow{p}\mathrm{Fr}(M_{n,b-1}),$$

where the map $p\colon\mathrm{Diff}^+(M_{n,b-1})\to\mathrm{Fr}(M_{n,b-1})$ is given by $\varphi\mapsto d\varphi(\tilde{x})$. Passing to the long exact sequence of homotopy groups associated to this fiber bundle, we find the segment

$$\pi_1(\mathrm{Fr}(M_{n,b-1}))\to\pi_0(\mathrm{Diff}^+(M_{n,b-1},\tilde{x}))\to\pi_0(\mathrm{Diff}^+(M_{n,b-1}))\to\pi_0(\mathrm{Fr}(M_{n,b-1})).\tag{2}$$

Since $\mathrm{Fr}(M_{n,b-1})$ is the *oriented* frame bundle, it is connected, and so $\pi_0(\mathrm{Fr}(M_{n,b-1}))$ is trivial.

Next, we claim that $\pi_0(\mathrm{Diff}^+(M_{n,b-1},\tilde{x}))$ is isomorphic to $\mathrm{Mod}(M_{n,b})$. For a proof of this fact in the surface case, see [12, page 102]. We will give an analogous argument here. Let B be the ball glued to the boundary component of $M_{n,b}$, and let $\mathrm{Diff}^+(M_{n,b-1}, B) \subset \mathrm{Diff}^+(M_{n,b-1})$ be the subspace of diffeomorphisms fixing B pointwise. Then $\pi_0(\mathrm{Diff}^+(M_{n,b-1}, B)) \cong \mathrm{Mod}(M_{n,b})$, where the isomorphism is obtained by simply removing B. Letting $\mathrm{Emb}^+((B, M_{n,b-1}), \tilde{x})$ be the space of smooth, orientation-preserving embeddings $B \hookrightarrow M_{n,b-1}$ which fix the framing $\tilde{x}$, we get a fiber bundle

$$\mathrm{Diff}^+(M_{n,b-1}, B) \hookrightarrow \mathrm{Diff}^+(M_{n,b-1}, \tilde{x}) \to \mathrm{Emb}^+((B, M_{n,b-1}), \tilde{x}),$$

where the second map is given by restriction to B. Again, we pass to the long exact sequence of homotopy groups, and find the segment

$$\begin{aligned}\pi_1(\mathrm{Emb}^+((B, M_{n,b-1}), \tilde{x})) &\to \pi_0(\mathrm{Diff}^+(M_{n,b-1}, B)) \\ &\to \pi_0(\mathrm{Diff}^+(M_{n,b-1}, \tilde{x})) \to \pi_0(\mathrm{Emb}^+((B, M_{n,b-1}), \tilde{x})).\end{aligned}$$

Since B is contractible, so is $\mathrm{Emb}^+((B, M_{n,b-1}), \tilde{x})$. This gives an isomorphism

$$\pi_0(\mathrm{Diff}^+(M_{n,b-1}, \tilde{x})) \cong \pi_0(\mathrm{Diff}^+(M_{n,b-1}, B)) \cong \mathrm{Mod}(M_{n,b}),$$

as desired.

With these identifications, the sequence (2) then becomes

$$\pi_1(\mathrm{Fr}(M_{n,b-1})) \to \mathrm{Mod}(M_{n,b}) \to \mathrm{Mod}(M_{n,b-1}) \to 1. \tag{3}$$

To get a short exact sequence, we must understand the kernel of the map

$$\widetilde{\mathrm{Push}}\colon \pi_1(\mathrm{Fr}(M_{n,b-1})) \to \mathrm{Mod}(M_{n,b}).$$

We remark here that the map $\widetilde{\mathrm{Push}}$ is given by pushing and rotating a small ball containing x about a loop based at x. This is in analogy with the "disk pushing maps" seen in the case of surfaces. Since $M_{n,b-1}$ is a compact, orientable 3-manifold, it is parallelizable, and hence we have

$$\pi_1(\mathrm{Fr}(M_{n,b-1})) \cong \pi_1(M_{n,b-1}) \times \pi_1(SO(3)) = \pi_1(M_{n,b-1}) \times \mathbb{Z}/2.$$

Consider the map $\mathrm{Mod}(M_{n,b}) \to \mathrm{Aut}(\pi_1(M_{n,b}, y))$, where the basepoint y is on the boundary component ∂ being capped off. As is shown in Figure 5, the composition

$$\pi_1(\mathrm{Fr}(M_{n,b-1})) \cong \pi_1(M_{n,b-1}) \times \mathbb{Z}/2 \xrightarrow{\widetilde{\mathrm{Push}}} \mathrm{Mod}(M_{n,b}) \to \mathrm{Aut}(\pi_1(M_{n,b}, y))$$

is given by conjugation about the loop being pushed around. Since $\mathrm{Aut}(\pi_1(M_{n,b}, y)) \cong \mathrm{Aut}(F_n)$ is centerless for $n > 1$, the entire kernel of $\widetilde{\mathrm{Push}}$ must be contained in $1 \times \mathbb{Z}/2 \subset \pi_1(M_{n,b-1}) \times \mathbb{Z}/2$. However, the image of the generator of this subgroup in $\mathrm{Mod}(M_{n,b})$ is the sphere twist T_∂. By Theorems 2.1 and 2.2, this sphere twist is nontrivial if and only if $b > 1$. If $b > 1$, this shows that Push is injective, and (3) gives us the desired exact sequence. On the other hand, if $b = 1$, then $\ker(\widetilde{\mathrm{Push}}) = 1 \times \mathbb{Z}/2$. Therefore, the image of $\widetilde{\mathrm{Push}}$ in $\mathrm{Mod}(M_{n,b})$ is isomorphic to

$$\pi_1(\mathrm{Fr}(M_{n,b-1}))/\langle T_\partial \rangle \cong \pi_1(M_{n,b-1})$$

as desired. □

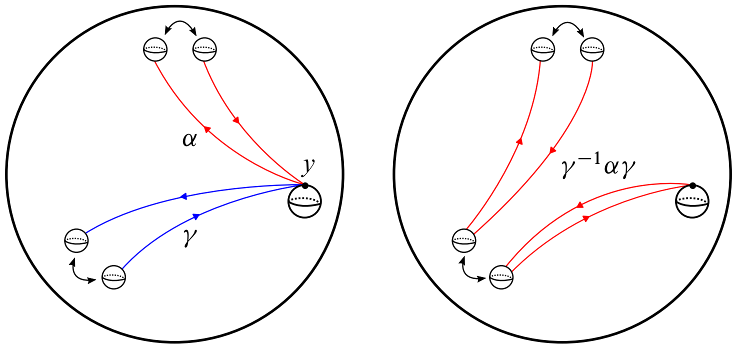


Figure 5: The image of α under $\widetilde{\mathrm{Push}}(\gamma, T)$ is $\gamma^{-1}\alpha\gamma$. Here, T can be either T_∂ or trivial.

Modding out by sphere twists Now that we have a Birman exact sequence for $\mathrm{Mod}(M_{n,b})$, we can mod out by sphere twists to get a Birman exact sequence for $\mathrm{Out}(F_{n,b})$. Consider the map

$$i_M : \mathrm{Mod}(M_{n,b}) \to \mathrm{Mod}(M_{n,b-1})$$

given by capping off a boundary component ∂. Since ι_M takes sphere twists to sphere twists, this map descends to a map $\iota_* : \mathrm{Out}(F_{n,b}) \to \mathrm{Out}(F_{n,b-1})$. Since ι_M is surjective, ι_* is as well. Let K be the kernel of ι_*, and let $\psi : \mathrm{Mod}(M_{n,b}) \to \mathrm{Out}(F_{n,b})$ be the quotient map. If $b > 1$, then the kernel of ι_M is $\pi_1(\mathrm{Fr}(M_{n,b-1}), \tilde{x})$ by Theorem 4.1. Let $\widetilde{\mathrm{Push}} : \pi_1(\mathrm{Fr}(M_{n,b-1}), \tilde{x}) \to \mathrm{Mod}(M_{n,b})$ be the map defined in the proof of Theorem 4.1, and fix an identification $\pi_1(\mathrm{Fr}(M_{n,b-1}), \tilde{x}) = \pi_1(M_{n,b-1}, x) \times \mathbb{Z}/2$. Since

$$\iota_*\big(\psi(\widetilde{\mathrm{Push}}(\gamma, T))\big) = \psi\big(\iota_M(\widetilde{\mathrm{Push}}(\gamma, T))\big) = \psi(\mathrm{id}) = \mathrm{id}$$

for all $(\gamma, T) \in \pi_1(M_{n,b-1}, x) \times \mathbb{Z}/2\mathbb{Z}$, the image of $\pi_1(\mathrm{Fr}(M_{n,b-1}), \tilde{x})$ under $\psi \circ \widetilde{\mathrm{Push}}$ is contained in K. In other words, we have the following commutative diagram:

$$\begin{tikzcd}
1 \arrow[r] & \pi_1(\mathrm{Fr}(M_{n,b-1}), \tilde{x}) \arrow[r, "\widetilde{\mathrm{Push}}"] \arrow[d, "\psi_P"] & \mathrm{Mod}(M_{n,b}) \arrow[r, "\iota_M"] \arrow[d, "\psi"] & \mathrm{Mod}(M_{n,b-1}) \arrow[r] \arrow[d] & 1 \\
1 \arrow[r] & K \arrow[r] & \mathrm{Out}(F_{n,b}) \arrow[r, "\iota_*"] & \mathrm{Out}(F_{n,b-1}) \arrow[r] & 1
\end{tikzcd}$$

where $\psi_P = \psi \circ \widetilde{\mathrm{Push}}$. Next, we claim that the map $\psi_P : \pi_1(\mathrm{Fr}(M_{n,b-1}), \tilde{x}) \to K$ is surjective. To see this, let $f \in K$, and choose a lift $\mathfrak{f} \in \mathrm{Mod}(M_{n,b})$ of f. Since $\iota_*(f) = \mathrm{id}$, the image $\iota_M(\mathfrak{f})$ is a product of sphere twists $T_{S_1} \cdots T_{S_j}$. For each $T_{S_i} \in \mathrm{Mod}(M_{n,b-1})$, choose a preimage $T'_{S_i} \in \mathrm{Mod}(M_{n,b})$ which is also a sphere twist. Then

$$\iota_M(T'_{S_1} \cdots T'_{S_j} \cdot \mathfrak{f}) = \mathrm{id},$$

which implies that $T'_{S_1} \cdots T'_{S_j} \cdot \mathfrak{f} = \widetilde{\mathrm{Push}}(\gamma, T)$ for some $(\gamma, T) \in \pi_1(M_{n,b-1}, x) \times \mathbb{Z}/2\mathbb{Z}$. Moreover, $\psi(T'_{S_1} \cdots T'_{S_j} \cdot \mathfrak{f}) = f$, which verifies our claim that $\psi_P : \pi_1(\mathrm{Fr}(M_{n,b-1}), \tilde{x}) \to K$ is surjective.

Now, we wish to identify the kernel of ψ_P. Let $(\gamma, T) \in \pi_1(M_{n,b-1}, x) \times \mathbb{Z}/2\mathbb{Z}$, and fix a basepoint y on the boundary component being capped off. At the end of the proof of Theorem 4.1, we saw that

$\widetilde{\mathrm{Push}}(\gamma, T)$ acts nontrivially on $\pi_1(M_{n,b}, y)$ if and only if γ is trivial. Since sphere twists act trivially on homotopy classes of curves, it follows that $\psi_P(\gamma, T)$ is nontrivial if γ is nontrivial. Therefore, the kernel of ψ_P must lie inside $1 \times \mathbb{Z}/2\mathbb{Z} \subset \pi_1(M_{n,b-1}, x) \times \mathbb{Z}/2\mathbb{Z}$. However, the generator of $1 \times \mathbb{Z}/2\mathbb{Z}$ gets mapped to T_∂ under $\widetilde{\mathrm{Push}}$, which is killed in $\mathrm{Out}(F_{n,b})$. Therefore, $\ker(\psi) = 1 \times \mathbb{Z}/2\mathbb{Z}$, and so it follows that $K \cong \pi_1(M_{n,b-1}, x)$.

On the other hand, if $b = 1$ and $n > 1$, then the kernel of the map $\iota_M : \mathrm{Mod}(M_{n,b}) \to \mathrm{Mod}(M_{n,b-1})$ is $\pi_1(M_{n,b-1}, x)$ by Theorem 4.1. Almost exactly the same argument used above shows that the quotient map restricts to a surjective map $\psi_P : \pi_1(M_{n,b-1}, x) \to K$. However, in this case, ψ_P is injective since the sphere twist T_∂ has already been killed off. Thus, we find that $K \cong \pi_1(M_{n,b-1}, x)$ in this case as well.

From now on, we will identify the kernel of the map $\iota_* : \mathrm{Out}(F_{n,b}) \to \mathrm{Out}(F_{n,b-1})$ with $\pi_1(M_{n,b-1}, x)$. The map $\pi_1(M_{n,b-1}, x) \to \mathrm{Out}(F_{n,b})$ will play a significant role throughout the remainder of the paper, and so we give a formal definition here.

Definition The map $\mathrm{Push} : \pi_1(M_{n,b-1}, x) \to \mathrm{Out}(F_{n,b})$ is defined as $\mathrm{Push}(\gamma) = \psi(\widetilde{\mathrm{Push}}(\gamma, T))$, where $T \in \mathbb{Z}/2\mathbb{Z}$ is arbitrary. Since sphere twists become trivial in $\mathrm{Out}(F_{n,b})$, this element depends only on γ.

The upshot of this is that we have proven the Birman exact sequence for $\mathrm{Out}(F_{n,b})$.

Theorem 4.2 *Fix $n, b > 0$ such that $(n, b) \neq (1, 1)$, and let $M_{n,b} \hookrightarrow M_{n,b-1}$ be an embedding obtained by gluing a ball to a boundary component. Fix $x \in M_{n,b-1} \setminus M_{n,b}$. Then the following sequence is exact:*

$$1 \to \pi_1(M_{n,b-1}, x) \xrightarrow{\mathrm{Push}} \mathrm{Out}(F_{n,b}) \xrightarrow{\iota_*} \mathrm{Out}(F_{n,b-1}) \to 1.$$

Restrict to Torelli We now move on to proving Theorem B, which gives a Birman exact sequence for $\mathrm{IO}_{n,b}^P$. We start by recalling its statement. Let P be a partition of the boundary components of $M_{n,b}$, and fix a boundary component ∂. Let $p \in P$ be the set containing ∂, and let $\iota : M_{n,b} \hookrightarrow M_{n,b-1}$ be the inclusion obtained by capping off ∂. The partition P induces a partition P' of the boundary components of $M_{n,b-1}$ by removing ∂ from p. With this definition of P', the map $\iota_* : \mathrm{Out}(F_{n,b}) \to \mathrm{Out}(F_{n,b-1})$ restricts to a map $\mathrm{IO}_{n,b}^P \to \mathrm{IO}_{n,b-1}^{P'}$, which we will also call ι_*. The sequence from Theorem 4.2 then restricts to

$$1 \to \pi_1(M_{n,b-1}) \cap \mathrm{IO}_{n,b}^P \to \mathrm{IO}_{n,b}^P \xrightarrow{\iota_*} \mathrm{IO}_{n,b-1}^{P'}.$$

Theorem B asserts that ι_* is surjective, and identifies its kernel $\pi_1(M_{n,b-1}) \cap \mathrm{IO}_{n,b}^P$. We start with surjectivity.

Lemma 4.3 *The induced map $\iota_* : \mathrm{IO}_{n,b}^P \to \mathrm{IO}_{n,b-1}^{P'}$ is surjective for any embedding $\iota : M_{n,b} \hookrightarrow M_{n,b-1}$.*

Proof Consider an element $g \in \mathrm{IO}_{n,b-1}^{P'}$. Our goal is to find some $f \in \mathrm{IO}_{n,b}^P$ such that $\iota_*(f) = g$. There are two cases.

First, suppose that $p=\{\partial\}$. Then the inclusion ι induces an isomorphism $\iota_H : H_1^P(M_{n,b}) \to H_1^{P'}(M_{n,b-1})$ which is equivariant with respect to the actions of $\mathrm{Out}(F_{n,b})$ and $\mathrm{Out}(F_{n,b-1})$. In other words, for any $[h] \in H_1^P(M_{n,b})$ and $f \in \mathrm{Out}(F_{n,b})$, we have

$$\iota_H(f \cdot [h]) = \iota_*(f) \cdot \iota_H([h]). \tag{4}$$

By Theorem 4.2, there exists some $f \in \mathrm{Out}(F_{n,b})$ such that $\iota_*(f)=g$. We claim that $f \in \mathrm{IO}^P_{n,b}$. To see this, let $[h] \in H_1^P(M_{n,b})$. Then, by (4), we see that

$$\iota_H(f \cdot [h]) = \iota_*(f) \cdot \iota_H([h]) = g \cdot \iota_H([h]) = \iota_H([h]).$$

Since ι_H is an isomorphism, this implies that $f \cdot [h] = [h]$, and so $f \in \mathrm{IO}^P_{n,b}$, as desired.

Next, suppose that $p \neq \{\partial\}$. Again, choose $f \in \mathrm{Out}(F_{n,b})$ such that $\iota_*(f)=g$. In this case, there is no longer a well-defined map $H_1^P(M_{n,b}) \to H_1^{P'}(M_{n,b-1})$. However, there is a subgroup of $H_1^P(M_{n,b})$ which projects isomorphically onto $H_1^{P'}(M_{n,b-1})$. Let $A \subset H_1^P(M_{n,b})$ be the subgroup generated by

$$\bigl\{[a] \in H_1(M_{n,b}, \partial M_{n,b}) \mid \text{ either } a \text{ is a simple closed curve or } a \text{ is a properly embedded arc with neither endpoint on } \partial\bigr\}.$$

It is clear that $A \cong H_1^{P'}(M_{n,b-1})$.

Let $[k] \in H_1^P(M_{n,b})$ be the class of an arc k which has an endpoint on ∂. We claim that $H_1^P(M_{n,b})$ is generated by A and $[k]$. To establish this claim, it suffices to show that $[\ell] \in \langle A, [k]\rangle$, where $[\ell] \in H_1^P(M_{n,b})$ is the class of any arc with an endpoint on ∂ and the other elsewhere. Such an ℓ exists since $p \neq \{\partial\}$. Fix such a class $[\ell]$, and let $\alpha \subset \partial$ be an arc connecting the endpoints of ℓ and k on ∂. Orient ℓ, α, and k such that the curve $\ell \cdot \alpha \cdot k$ is well-defined.

If the endpoints of ℓ and k which are not on ∂ lie on distinct boundary components, then $\ell \cdot \alpha \cdot k$ is an arc connecting P'-adjacent boundary components. Therefore, $[\ell]+[\alpha]+[k] \in A$. Since $[\alpha]=0$ in $H_1^P(M_{n,b})$, it follows that $[\ell] \in \langle A, [k]\rangle$. On the other hand, if the endpoints of ℓ and k which are not on ∂ lie on the same boundary component ∂', then we can complete $\ell \cdot \alpha \cdot k$ to a loop $\ell \cdot \alpha \cdot k \cdot \beta$, where $\beta \subset \partial'$ is an arc connecting the endpoints of ℓ and k. Then

$$[\ell]+[k] = [\ell]+[\alpha]+[k]+[\beta] = [\ell \cdot \alpha \cdot k \cdot \beta] \in A,$$

and so $[\ell] \in \langle A, [k]\rangle$. This completes the proof of the claim that $H_1^P(M_{n,b})$ is generated by A and $[k]$.

Since A projects isomorphically onto $H_1^{P'}(M_{n,b-1})$, and this projection is equivariant with respect to the actions of $\mathrm{Out}(F_{n,b})$ and $\mathrm{Out}(F_{n,b-1})$, we have $f \cdot [a] = [a]$. It follows that f acts trivially on A. Therefore, if f fixes $[k]$, then $f \in \mathrm{IO}^P_{n,b}$ by the discussion in the preceding paragraph, and so we are done. On the other hand, if f does not fix $[k]$, then $\gamma = k \cdot f(k)^{-1}$ is a nontrivial loop based at a point on ∂. So, the element $\mathrm{Push}(\gamma)^{-1} \cdot f \in \mathrm{Out}(F_{n,b})$ fixes $[k]$. Moreover, $\mathrm{Push}(\gamma)$ acts trivially on A, and so $\mathrm{Push}(\gamma)^{-1} \cdot f$ does as well. Thus, $\mathrm{Push}(\gamma)^{-1} \cdot f \in \mathrm{IO}^P_{n,b}$. Finally, since $\mathrm{Push}(\gamma) \in \ker(i_*)$, we have that $\iota_*(\mathrm{Push}(\gamma)^{-1} \cdot f) = g$, and so we are done. □

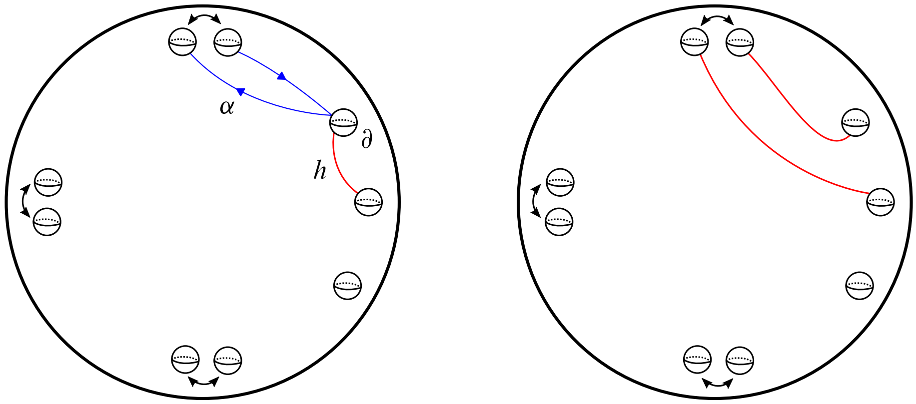


Figure 6: Dragging ∂ around α takes $[h]$ to $[\alpha]+[h]$.

Proof of Theorem B Recall that we want to show that we have an exact sequence

$$1 \to L \xrightarrow{\text{Push}} \mathrm{IO}^{P}_{n,b} \xrightarrow{\iota_*} \mathrm{IO}^{P'}_{n,b-1} \to 1,$$

where L is equal to

(a) $\pi_1(M_{n,b-1}, x) \cong F_n$ if $p = \{\partial\}$,

(b) $[\pi_1(M_{n,b-1}, x), \pi_1(M_{n,b-1}, x)] \cong [F_n, F_n]$ if $p \neq \{\partial\}$.

By Lemma 4.3 and the discussion preceding it, all that is left to show is that $\pi_1(M_{n,b-1}) \cap \mathrm{IO}^{P}_{n,b}$ agrees with the subgroups L given above.

We begin with the case $p = \{\partial\}$. Recall that $\pi_1(M_{n,b-1})$ acts on $M_{n,b}$ by pushing the boundary component ∂ about a given loop. Since ∂ is not P-adjacent to any other boundary components, it follows that $\pi_1(M_{n,b-1})$ acts trivially on $H_1^P(M_{n,b})$. Therefore, $\pi_1(M_{n,b-1}) \subset \mathrm{IO}^{P}_{n,b}$, and so $\pi_1(M_{n,b-1}) \cap \mathrm{IO}^{P}_{n,b} = \pi_1(M_{n,b-1})$. This completes this case.

Next, suppose that $p \neq \{\partial\}$. In this case, not all elements of $\pi_1(M_{n,b})$ are contained in $\mathrm{IO}^{P}_{n,b}$. This is because dragging ∂ about loops may change the homology class of arcs connected to ∂. In particular, if $\gamma \in \pi_1(M_{n,b-1})$ and $[h] \in H_1^P(M_{n,b})$ is the class of arc with an endpoint in ∂ and the other elsewhere, then $\mathrm{Push}(\gamma)$ acts on $[h]$ via

$$\mathrm{Push}(\gamma) \cdot [h] = [\gamma] + [h].$$

See Figure 6 for an illustration. This implies that an element $\mathrm{Push}(\gamma)$ is in $\mathrm{IO}^{P}_{n,b}$ if and only if $[\gamma] = 0$ in $H_1(M_{n,b-1})$. Thus,

$$\pi_1(M_{n,b-1}) \cap \mathrm{IO}^{P}_{n,b} = [\pi_1(M_{n,b-1}), \pi_1(M_{n,b-1})],$$

which is what we wanted to show. □

5 Generators

In this section, we will define our generators of $\mathrm{IO}^P_{n,b}$. The definition of these generators will involve splitting and dragging boundary components, so we will discuss these processes in more detail first, then move on to the definitions.

Splitting along spheres Let $S \subset M_{n,b}$ be an embedded 2-sphere. By *splitting along* S, we mean removing an open tubular neighborhood N of S from $M_{n,b}$. If S is nonseparating, the resulting manifold will be diffeomorphic to $M_{n-1,b+2}$ and if S is separating, the result will be diffeomorphic to $M_{m_1,c_1} \sqcup M_{m_2,c_2}$, where $m_1 + m_2 = n$ and $c_1 + c_2 = b + 2$. Notice that the resulting manifold is a submanifold of $M_{n,b}$, and so we get a corresponding map $\mathrm{Mod}(M_{n-1,b+2}) \to \mathrm{Mod}(M_{n,b})$ if S is nonseparating, or $\mathrm{Mod}(M_{m_1,c_1}) \times \mathrm{Mod}(M_{m_2,c_2}) \to \mathrm{Mod}(M_{n,b})$ if S is separating. In either case, this map sends sphere twists to sphere twists, and thus induces a map $\iota_*\colon \mathrm{Out}(F_{n-1,b+2}) \to \mathrm{Out}(F_{n,b})$ or $\iota_*\colon \mathrm{Out}(F_{m_1,c_1}) \times \mathrm{Out}(F_{m_2,c_2}) \to \mathrm{Out}(F_{n,b})$, depending on whether or not S separates $M_{n,b}$.

Dragging boundary components Let ∂ be a boundary component of $M_{n,b}$, and let $\iota\colon M_{n,b} \hookrightarrow M_{n,b-1}$ be the embedding obtained by capping off ∂. By Theorem 4.2, we have an exact sequence

$$1 \to \pi_1(M_{n,b-1}, x) \xrightarrow{\mathrm{Push}} \mathrm{Out}(F_{n,b}) \xrightarrow{\iota_*} \mathrm{Out}(F_{n,b-1}) \to 1,$$

where $x \in M_{n,b-1} \setminus M_{n,b}$. Given $\gamma \in \pi_1(M_{n,b-1}, x)$, recall that the element $\mathrm{Push}(\gamma) \in \mathrm{Out}(M_{n,b})$ is given by pushing ∂ about the loop γ. In the remainder of this section, we will be dragging multiple boundary components at a time. So, from now on we will write $\mathrm{Push}_\partial(\gamma)$ in order to keep track of which boundary component is being pushed.

Magnus generators We now move on to defining our generators for $\mathrm{IO}^P_{n,b}$. In the $b = 0$ case, we have that $\mathrm{IO}^P_{n,0} = \mathrm{IO}_n$, where IO_n is the subgroup of $\mathrm{Out}(F_n)$ acting trivially on homology. In [21], Magnus found the following generating set for IO_n.

Theorem 5.1 (Magnus) *Let $F_n = \langle x_1, \dots, x_n \rangle$. The group IO_n is generated by the $\mathrm{Out}(F_n)$-classes of the automorphisms*

$$M_{ij}\colon x_i \mapsto x_j x_i x_j^{-1}, \quad M_{ijk}\colon x_i \mapsto x_i[x_j, x_k],$$

for all distinct $i, j, k \in \{1, \dots, n\}$ with $j < k$. Here, the automorphisms are understood to fix x_ℓ for $\ell \neq i$.

Our generating set will be inspired by Magnus's, and will indeed reduce to it when $b = 0$. In order to choose a concrete collection of elements, we will need to make some choices. First, fix a basepoint $* \in \mathrm{int}(M_{n,b})$ and a collection of n disjointly embedded oriented 2-spheres $S_1, \dots, S_n \subset M_{n,b} \setminus *$. We will call such a collection a *sphere basis* for $M_{n,b}$. In addition, choose a corresponding *geometric free*

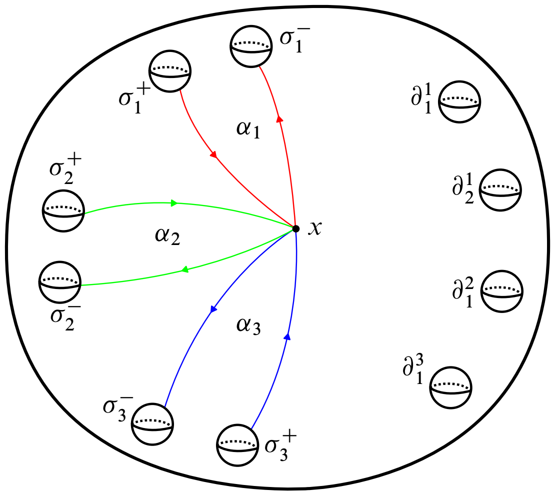


Figure 7: $M_{3,4}$ split along $S_1 \cup S_2 \cup S_3$ with the partition $P = \{\{\partial_1^1, \partial_2^1\}, \{\partial_1^2\}, \{\partial_1^3\}\}$.

basis; that is, a set $\{\alpha_1, \dots, \alpha_n\}$ of oriented simple closed curves intersecting only at $*$ such that

- α_i intersects S_i exactly once with positive orientation for all i,
- α_i is disjoint from S_j if $i \neq j$.

Notice that the homotopy classes of $\{\alpha_1, \dots, \alpha_n\}$ necessarily forms a free basis for $\pi_1(M_{n,b}, *)$. Splitting $M_{n,b}$ along the S_i reduces it to a 3-sphere $\mathcal{Z} \subset M_{n,b}$ with $b+2n$ boundary components. The submanifold $\mathcal{Z}$ will play a significant role throughout the remainder of this section because it will allow all of our choices made in the definitions to be unique. For each S_i, let σ_i^+ and σ_i^- be the boundary components of $\mathcal{Z}$ arising from the split along S_i, where σ_i^+ (resp. σ_i^-) is the component lying on the positive (resp. negative) side of S_i. We will also choose an ordering $P = \{p_1, \dots, p_{|P|}\}$ and an ordering $p_r = \{\partial_1^r, \dots, \partial_{b_r}^r\}$ for each $r \in \{1, \dots, |P|\}$. See Figure 7.

The following lemma will be helpful in showing that our generators lie in $\mathrm{IO}_{n,b}^P$.

Lemma 5.2 *Let $\mathcal{Z}$ be as above, and suppose that $h \subset M_{n,b}$ is a properly embedded oriented arc connecting P-adjacent boundary components of $M_{n,b}$. Then the homology class of $[h] \in H_1^P(M_{n,b})$ has the form*

$$[h] = [\alpha] + [h_0],$$

where α is a loop based at $*$, and h_0 is the unique arc (up to isotopy) in $\mathcal{Z}$ which has the same endpoints as h.

Proof We may homotope h such that it has the form $h = \gamma_s \cdot \alpha \cdot \gamma_e$, where (see Figure 8)

- $\gamma_s \subset \mathcal{Z}$ is the unique arc (up to isotopy) from the initial point of h to the basepoint $*$ of $M_{n,b}$,

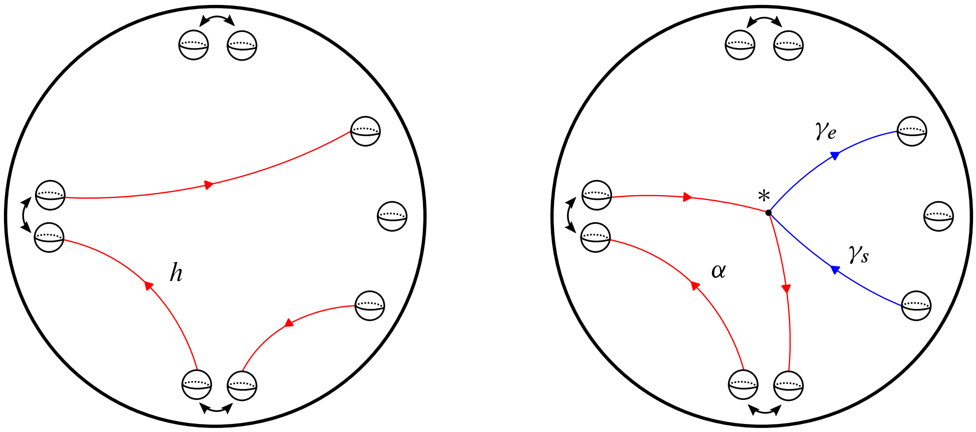


Figure 8: The arc h homotoped to be put in the form $\gamma_s \cdot \alpha \cdot \gamma_e$.

- $\gamma_e \subset \mathcal{Z}$ is the unique arc from $*$ to the endpoint of h,
- $\alpha \in \pi_1(M_{n,b}, *)$.

Then

$$[h] = [\gamma_s \cdot \alpha \cdot \gamma_e] = [\alpha] + [\gamma_s \cdot \gamma_e] = [\alpha] + [h_0],$$

as desired. □

Handle drags Let $i \in \{1, \ldots, n\}$, and let h_i be the unique (up to isotopy) properly embedded arc in $\mathcal{Z}$ connecting σ_i^+ and σ_i^- which is disjoint from the α_k. Choose a tubular neighborhood N_i of $\sigma_i^+ \cup h_i \cup \sigma_i^-$ that does not intersect any α_k for $k \neq i$. Let Σ_i be the boundary component of N_i which is not isotopic to σ_i^+ or σ_i^- (notice that Σ_i is diffeomorphic to a 2-sphere). Splitting $M_{n,b}$ along Σ_i yields $M_{n-1,b+1} \sqcup M_{1,1}$. Let $\Sigma_i' \subset \partial M_{n-1,b+1}$ be the boundary component coming from this split, and fix a basepoint $y_i \in \Sigma_i'$. Fix an oriented arc $\delta_i \subset Z$ from y_i to $*$ which only intersects Σ_i' at y_i. Since $\mathcal{Z}$ is a 3-sphere with spherical boundary components, δ_i is unique up to isotopy. The arc δ_i induces an isomorphism $\pi_1(M_{n-1,b+1}, *) \to \pi_1(M_{n-1,b+1}, y_i)$ given by $\gamma \mapsto \delta_i \gamma \delta_i^{-1}$. Define $\beta_j^i = \delta_i \alpha_j \delta_i^{-1}$. Then we define the *handle drag* $\mathrm{HD}_{ij} := \iota_*(\mathrm{Push}_{\Sigma_i'}(\beta_j^i), \mathrm{id}) \in \mathrm{Out}(F_{n,b})$ for $j \neq i$, where ι_* is the map $\mathrm{Out}(F_{n-1,b+1}) \times \mathrm{Out}(F_{1,1}) \to \mathrm{Out}(F_{n,b})$ induced by splitting along Σ_i.

To see that $\mathrm{HD}_{ij} \in \mathrm{IO}_{n,b}^P$, notice that HD_{ij} acts trivially on α_k for $k \neq i$, and acts on α_i via $\alpha_i \mapsto \alpha_j \alpha_i \alpha_j^{-1}$. See Figure 9. This shows that HD_{ij} acts trivially on homology classes of simple closed curves. Additionally, this shows that HD_{ij} reduces to M_{ij} of the Magnus generators if $b = 0$.

Next, suppose that h is an arc connecting P-adjacent boundary components. By Lemma 5.2, we may write $[h] = [\alpha] + [h_0]$, where α is a loop based at $*$, and h_0 is the unique arc (up to isotopy) in $\mathcal{Z}$ which has the same endpoints as h. We have seen that HD_{ij} fixes the homology class of α. Moreover, we may homotope HD_{ij} such that it fixes the arc h_0. Thus, HD_{ij} fixes the homology class of h, and we conclude that $\mathrm{HD}_{ij} \in \mathrm{IO}_{n,b}^P$.

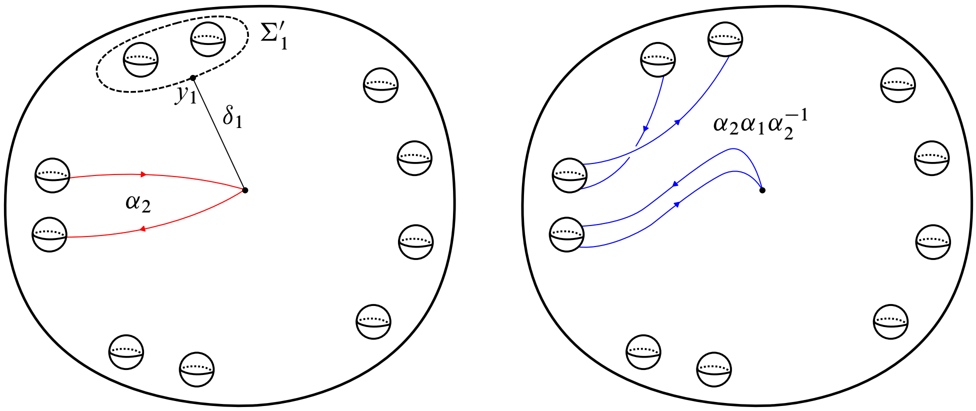


Figure 9: Setup of the handle drag HD_{12} and the image of α_1 under HD_{12}

Commutator drags Let $i,j,k \in \{1,\dots,n\}$ be distinct with $j<k$. Split $M_{n,b}$ along S_i to get $M_{n,b+2}$, where $\mathcal{Z} \subset M_{n,b+2} \subset M_{n,b}$. Fix basepoint $y_i \in \sigma_i^+$ and $z_i \in \sigma_i^-$, and choose oriented arcs $\delta_i, \varepsilon_i \subset \mathcal{Z}$ connecting y_i and z_i to $*$, respectively. Just as in the construction of handle drags, δ_i and ε_i are unique up to isotopy. Let $\beta_\ell^i = \delta_i\alpha_\ell\delta_i^{-1}$ and $\gamma_\ell^i = \varepsilon_i\alpha_\ell\varepsilon_i^{-1}$ for $\ell \in \{j,m\}$. Then, we define the *commutator drags* $\mathrm{CD}_{ijk}^+, \mathrm{CD}_{ijk}^- \in \mathrm{Out}(F_{n,b})$ as $\iota_*(\mathrm{Push}_{\sigma_i^+}([\beta_j^i,\beta_k^i]))$ and $\iota_*(\mathrm{Push}_{\sigma_i^-}([\gamma_j^i,\gamma_k^i]))$, respectively, where $\iota_*\colon \mathrm{Out}(F_{n,b+2}) \to \mathrm{Out}(F_{n,b})$ is the map induced by splitting along S_i. See Figure 10.

Again, we see that $\mathrm{CD}_{ijk}^\pm$ acts trivially on α_ℓ for $\ell \neq i$, the commutator drag CD_{ijk}^+ sends α_i to $\alpha_i[\alpha_j,\alpha_k]^{-1}$, and CD_{ijk}^- sends α_i to $[\alpha_j,\alpha_k]\alpha_i$. This shows that CD_{ijk}^+ reduces to M_{ijk}^{-1} of the Magnus generators when $b=0$.

Now, suppose that h is an arc connecting P-adjacent boundary components of $M_{n,b}$. By Lemma 5.2, we may express $[h]$ in the form $[h]=[\alpha]+[h_0]$. We just saw that $\mathrm{CD}_{ijk}^\pm$ fixes $[\alpha]$. We may also homotope $\mathrm{CD}_{ijk}^\pm$ such that it fixes h_0. Thus, $\mathrm{CD}_{ijk}^\pm$ fixes $[h]$, and so $\mathrm{CD}_{ijk}^\pm \in \mathrm{IO}_{n,b}^P$.

Boundary commutator drags Let $p_r \in P$ and $\partial_s^r \in p_r$. Fix $i,j \in \{1,\dots,n\}$ such that $i<j$. Choose a basepoint $y_s^r \in \partial_s^r$. Let $\gamma_s^r \subset \mathcal{Z}$ be the unique arc (up to isotopy) from y_s^r to $*$. Let $\beta_k^{rs} = \gamma_s^r\alpha_k(\gamma_s^r)^{-1}$ for $k \in \{i,j\}$. Then, we define the *boundary commutator drags* $\mathrm{BCD}_{ij}^{rs} = \mathrm{Push}_{\partial_s^r}([\beta_i^{rs},\beta_j^{rs}]) \in \mathrm{Out}(F_{n,b})$.

It is clear from the definition that BCD_{ij}^{rs} acts trivially on $\alpha_1,\dots,\alpha_n$ and arcs that do not have an endpoint on ∂_s^r. Suppose that h is an oriented arc with an endpoint on ∂_s^r. Without loss of generality, suppose the terminal endpoint of h lies on ∂_s^r. Applying Lemma 5.2, we may write $[h]=[\alpha]+[h_0]$, where α is a loop based at $*$ and $h_0 \subset \mathcal{Z}$ is the unique arc (up to isotopy) which shares endpoints with h. We just saw that BCD_{ij}^{rs} fixes the α_k, and thus fixes the homology class $[\alpha]$. Therefore,

$$\begin{aligned}\mathrm{BCD}_{j\ell m}([h]) = \mathrm{BCD}_{ij}^{rs}([\alpha]+[h_0]) = [\alpha]+\mathrm{BCD}_{ij}^{rs}([h_0]) &= [\alpha]+[h_0]+[\alpha_i\cdot\alpha_j\cdot\alpha_i^{-1}\cdot\alpha_j^{-1}]\\ &= [\alpha]+[h_0]=[h].\end{aligned}$$

So, it follows that $\mathrm{BCD}_{ij}^{rs} \in \mathrm{IO}_{n,b}^P$ as well.

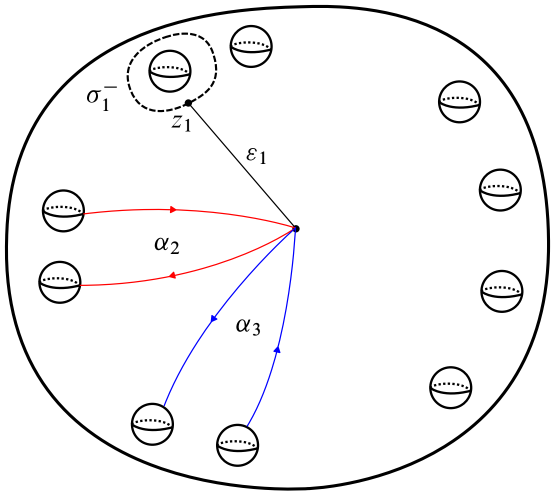


Figure 10: Setup of the commutator drag CD^-_{123}.

***P*-drags** The final type of elements we will define are called P-drags, where P is a partition of the boundary components of $M_{n,b}$. Let $p_r \in P$ and $i \in \{1,\dots,n\}$. Let $\Sigma_r \subset \mathcal{Z}$ be the unique 2-sphere (up to isotopy) which separates the boundary components of p_r from the remaining boundary components and the $\sigma_j^{\pm}$. Splitting $M_{n,b}$ along Σ_r gives $M_{n,b-c+1} \sqcup M_{0,c+1}$, where c is the number of boundary components in p. Let $\Sigma'_r \subset \partial M_{n,b-c+1}$ be the boundary component coming from this splitting. Just as in the construction of the other drags, fix a basepoint $y_r \in \Sigma'_r$ and an oriented arc γ_r from y_r to $*$ to get a basis $\{\beta^r_1,\dots,\beta^r_n\}$ of $\pi_1(M_{n,b-c+1}, y_r)$. See Figure 11. Then, we define the *P-drag* $\mathrm{PD}^r_i := \iota_*(\mathrm{Push}_{\Sigma'_r}(\beta_i), \mathrm{id}) \in \mathrm{Out}(F_{n,b})$, where $\iota_*\colon \mathrm{Out}(F_{n,b-c+1}) \times \mathrm{Out}(F_{0,c+1}) \to \mathrm{Out}(F_{n,b})$ is the map induced by splitting along Σ_r.

To see why $\mathrm{PD}^r_i \in \mathrm{IO}^P_{n,b}$, first notice that we can isotope PD^r_i to fix all the α_j. Next, if h is an arc connecting P-adjacent boundary components, we write $[h] = [\alpha] + [h_0]$ as in Lemma 5.2. As we just noted, PD^r_i fixes $[\alpha]$, so it suffices to show that PD^r_i fixes the homology class of h_0. If the endpoints of h lie on boundary components in p_r, then we may homotope h_0 such that it never crosses Σ_r. Then, PD^r_i fixes h_0. On the other hand, if the endpoints of h lie on boundary components which are not in p_r, then again we can homotope h_0 such that it does not cross Σ_r, and then homotope PD^r_i such that it fixes h_0. In either case, PD^r_i fixes the homology class of h_0, and so we conclude that $\mathrm{PD}^r_i \in \mathrm{IO}^P_{n,b}$.

Images under capping Suppose we have an embedding $\iota\colon M_{n,b} \hookrightarrow M_{n,b-1}$ given by capping off the boundary component ∂. Let $\iota_*\colon \mathrm{IO}^P_{n,b} \to \mathrm{IO}^{P'}_{n,b-1}$ be the induced map, where P' is the partition of the boundary components of $M_{n,b-1}$ induced by P. Using the sphere basis $\{S_1,\dots,S_n\}$ and geometric free basis $\{\alpha_1,\dots,\alpha_n\}$ for $M_{n,b}$, we get a corresponding sphere basis $\{\iota(S_1),\dots,\iota(S_n)\}$ and geometric free basis $\{\iota(\alpha_1),\dots,\iota(\alpha_n)\}$ for $M_{n,b-1}$. Moreover, the ordering on P (and each $p_r \in P$) induces an ordering on P'. We can repeat the process described throughout this section to define handle drags,

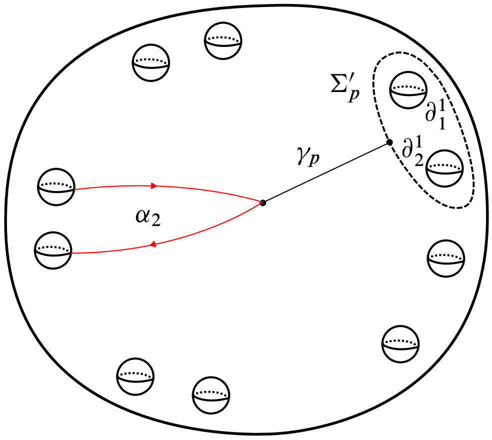


Figure 11: Setup of the P-drag PD^p_2, where $p = \{\partial^1_1, \partial^1_2\} \in P$.

commutator drags, boundary commutator drag, and P'-drags in $\mathrm{IO}^{P'}_{n,b-1}$, which we will denote by $\overline{\mathrm{HD}}_{ij}$, $\overline{\mathrm{CD}}^{\pm}_{ijk}$, $\overline{\mathrm{BCD}}^{rs}_{ij}$, and $\overline{\mathrm{PD}}^{r'}_i$, respectively. With this setup, we find that

$$\begin{aligned}
\iota_*(\mathrm{HD}_{ij}) &= \overline{\mathrm{HD}}_{ij},\\
\iota_*(\mathrm{CD}^{\pm}_{ijk}) &= \overline{\mathrm{CD}}^{\pm}_{ijk},\\
\iota_*(\mathrm{BCD}^{rs}_{ij}) &= \begin{cases}\mathrm{id} & \text{if } \partial = \partial^r_s,\\ \overline{\mathrm{BCD}}^{rs}_{ij} & \text{otherwise,}\end{cases}\\
\iota_*(\mathrm{PD}^r_i) &= \begin{cases}\mathrm{id} & \text{if } p_r = \{\partial^r_1\},\\ \overline{\mathrm{PD}}^r_i & \text{otherwise.}\end{cases}
\end{aligned}$$

6 Finite generation

Now that we have defined our collection of candidate generators for $\mathrm{IO}^{P}_{n,b}$, we now move on to proving that they do in fact generate. The first step in this proof will be an induction on b to reduce to the case of $b = 0$. This induction will rely on the following theorem of Tomaszewski [25] (see [24] for a geometric proof).

Theorem 6.1 (Tomaszewski) *Let F_n be the free group on n letters $\{x_1, \dots, x_n\}$. The commutator subgroup $[F_n, F_n]$ of F_n is freely generated by the set*

$$\bigl\{[x_i, x_j]^{x_i^{d_i} \cdots x_n^{d_n}},\ 1 \le i < j \le n,\ d_\ell \in \mathbb{Z},\ i \le \ell \le n\bigr\},$$

where the superscript denotes conjugation.

We will also need the following lemma from group theory.

Lemma 6.2 *Consider an exact sequence of groups*

$$1 \to K \to G \to Q \to 1.$$

Let S_Q be a generating set for Q. Moreover, assume that there are sets $S_K \subset K$ and $S_G \subset G$ such that K is contained in the subgroup of G generated by S_K and S_G. Then G is generated by the set $S_K \cup S_G \cup \widetilde{S}_Q$, where $\widetilde{S}_Q$ is a set consisting of one lift $\widetilde{q} \in G$ for every element $q \in S_q$.

Proof of lemma Let $G' \subset G$ be the subgroup generated by $S_K \cup S_G \cup \widetilde{S}_Q$, and let $K' = G' \cap K$. Then the following diagram commutes and has exact rows:

$$\begin{array}{ccccccccc} 1 & \longrightarrow & K' & \longrightarrow & G' & \longrightarrow & Q & \longrightarrow & 1 \\ & & \Big\downarrow{\scriptstyle\varphi} & & \Big\downarrow & & \Big\| & & \\ 1 & \longrightarrow & K & \longrightarrow & G & \longrightarrow & Q & \longrightarrow & 1 \end{array}$$

The vertical maps are all inclusions, and hence injective. Also, by assumption, the map φ is surjective. Therefore, by the five lemma, all of the vertical maps are isomorphisms, and so we are done. $\square$

We now prove the following result, which implies Theorem C since the collections of handle, commutator, boundary commutator, and P-drags are all finite. In Section 8, we will compute the number of each type of drag, and show how many become redundant in the abelianization of IA_n.

Theorem 6.3 *The group $\mathrm{IO}_{n,b}^{P}$ is generated by handle, commutator, boundary commutator, and P-drags for $b \geq 0$, $n > 0$.*

Proof As mentioned above, we will prove this by induction on b. The base case $b = 0$ follows directly from Magnus's Theorem 1.1.

If $b > 0$, fix a boundary component ∂ of $M_{n,b}$ and let $p \in P$ be the partition containing ∂. Let $\iota\colon M_{n,b} \hookrightarrow M_{n,b-1}$ be an embedding obtained by capping off ∂, and choose a basepoint $* \in M_{n,b-1} \setminus M_{n,b}$. By Theorem B, there is an exact sequence

$$1 \to L \xrightarrow{\mathrm{Push}} \mathrm{IO}_{n,b}^{P} \xrightarrow{\iota_*} \mathrm{IO}_{n,b-1}^{P'} \to 1,$$

where $L = \pi_1(M_{n,b}, *)$ if $p = \{\partial\}$ and $L = [\pi_1(M_{n,b}, *), \pi_1(M_{n,b}, *)]$ otherwise. As we saw in the discussion at the end of Section 5, we can define the drags of $\mathrm{IO}_{n,b}^{P}$ and $\mathrm{IO}_{n,b-1}^{P'}$ in a consistent way; that is, we can define our drags in such a way that ι_* takes handle drags to handle drags, commutator drags to commutator drags, and so on. By induction, $\mathrm{IO}_{n,b-1}^{P'}$ is generated by the desired drags. Therefore, it suffices to show that $\mathrm{Push}(L)$ is generated by our drags as well. If $p = \{\partial\}$, then $\mathrm{Push}(L)$ is precisely the subgroup of $\mathrm{IO}_{n,b}^{P}$ generated by the P-drags, and so we are done in this case.

The case of $p \neq \{\partial\}$ is less straightforward since the commutator subgroup of a free group is not finitely generated when $n \geq 2$. However, this is not necessary for $\mathrm{IO}_{n,b}^{P}$ to be finitely generated by our collection of drags. We will appeal to Lemma 6.2. Suppose that $p \neq \{\partial\}$. Then, by Theorem 6.1, the kernel $L = [\pi_1(M_{n,b}, *), \pi_1(M_{n,b}, *)]$ of the Birman exact sequence is generated by elements of the form $[x_i, x_j]^{x_i^{d_i} \cdots x_n^{d_n}}$. First, notice that $\mathrm{Push}([x_i, x_j])$ is the boundary commutator drag BCD_{ij}^{rs}, where $\partial_s^r = \partial$

is the boundary component of $M_{n,b}$ being capped off. Moreover, we have seen that the handle drag $\mathrm{HD}_{k\ell}$ acts on x_k by $x_k \mapsto x_\ell x_k x_\ell^{-1}$. It follows that $\mathrm{HD}_{ik} \cdot \mathrm{HD}_{jk}([x_i, x_j]) = [x_i, x_j]^{x_k}$. Continuing this pattern, we see that

$$[x_i, x_j]^{x_i^{d_i} \cdots x_n^{d_n}} = (\mathrm{HD}_{in} \cdot \mathrm{HD}_{jn})^{d_n} \cdots (\mathrm{HD}_{ii} \cdot \mathrm{HD}_{ji})^{d_i}([x_i, x_j]),$$

where HD_{ii} is taken to be trivial. Let $H = (\mathrm{HD}_{in} \cdot \mathrm{HD}_{jn})^{d_n} \cdots (\mathrm{HD}_{ii} \cdot \mathrm{HD}_{ji})^{d_i}$. Then,

$$\mathrm{Push}([x_i, x_k]^{x_i^{d_i} \cdots x_n^{d_n}}) = \mathrm{Push}(H([x_i, x_j])) = H \cdot \mathrm{Push}([x_i, x_j]) \cdot H^{-1} = H \cdot \mathrm{BCD}_{ij}^{rs} \cdot H^{-1}.$$

This shows that $\mathrm{Push}(L)$ is contained in the subgroup of $\mathrm{IO}_{n,b}^{P}$ generated by boundary commutator and handle drags. Applying Lemma 6.2 (taking $S_G = \{\text{handle drags}\}$ and $S_K = \{\mathrm{BCD}_{i,j}^{r,s} \mid 1 \le i, j \le n\}$), we conclude that $\mathrm{IO}_{n,b}^{P}$ is generated by the desired drags. □

7 Partial proof of Magnus's theorem

In this section, we will give a partial proof of Magnus's Theorem 1.1, which constituted the base case in the proof of Theorem 6.3. As stated in the introduction, the original proof of Magnus's theorem involved two steps: showing that the elements M_{ij} and M_{ijk} *normally* generate IO_n, and then showing that the subgroup generated by these elements is normal. We will give a proof of the first step here (Theorem D).

In order to establish this fact, we will examine the action of $\mathrm{IO}_{n,0}^{\{\}} = \mathrm{IO}_n$ on a certain simplicial complex, and apply the following theorem of Armstrong [2]. We say that a group G acts on a simplicial complex X *without rotations* if every simplex s is fixed pointwise by every element of its stabilizer, which we will denote by G_s.

Theorem 7.1 (Armstrong) *Suppose the group G acts on a simply connected simplicial complex X without rotations. If X/G is simply connected, then G is generated by the set*

$$\bigcup_{v \in X^{(0)}} G_v.$$

Here $X^{(0)}$ is the 0-skeleton of X.

Remark In [2], Armstrong proves the converse of this theorem as well. For a modern discussion of the proof of Theorem 7.1, along with some generalizations, we refer the reader to [23, Section 3].

The nonseparating sphere complex The complex to which we will apply Theorem 7.1 will be the *nonseparating sphere complex* $\mathbb{S}_n^{\mathrm{ns}}$. Vertices of $\mathbb{S}_n^{\mathrm{ns}}$ are isotopy classes of smoothly embedded nonseparating 2-spheres in M_n, and $\mathbb{S}_n^{\mathrm{ns}}$ has a k-simplex $\{S_0, \dots, S_k\}$ if the spheres $S_0, \dots, S_k$ can be realized pairwise disjointly and their union does not separate M_n. This is a subcomplex of the more ubiquitous *sphere complex*, which was introduced by Hatcher in [13] as a tool to explore the homological stability of $\mathrm{Out}(F_n)$ and $\mathrm{Aut}(F_n)$. In [13, Proposition 3.1], Hatcher proves the following connectivity result about $\mathbb{S}_n^{\mathrm{ns}}$.

Proposition 7.2 (Hatcher) *The complex* $\mathbb{S}_n^{\mathrm{ns}}$ *is* $(n-2)$*-connected.*

In particular, $\mathbb{S}_n^{\mathrm{ns}}$ is simply connected for $n \geq 3$. Recall that sphere twists act trivially on isotopy classes of embedded surfaces, and so we get an action of IO_n on $\mathbb{S}_n^{\mathrm{ns}}$. Notice that spheres in a simplex of $\mathbb{S}_n^{\mathrm{ns}}$ necessarily represent distinct H_2-classes. By Poincaré duality, elements of IO_n act trivially on $H_2(M_n)$, and so this implies that IO_n acts on $\mathbb{S}_n^{\mathrm{ns}}$ without rotations. Thus, in order to apply Theorem 7.1, we must show that $\mathbb{S}_n^{\mathrm{ns}}/\mathrm{IO}_n$ is simply connected.

To do this, we will give a description of $\mathbb{S}_n^{\mathrm{ns}}/\mathrm{IO}_n$ in terms of linear algebra. Fix an identification $H_2(M_n) = \mathbb{Z}^n$. Let $\mathrm{FS}(\mathbb{Z}^n)$ be the simplicial complex whose vertices are rank 1 summands of $\mathbb{Z}^n$, and there is a ℓ-simplex $\{A_0, \dots, A_\ell\}$ if $A_0 \oplus \cdots \oplus A_\ell$ is a rank $\ell+1$ summand of $\mathbb{Z}^n$. There is a map $\varphi\colon \mathbb{S}_n^{\mathrm{ns}}/\mathrm{IO}_n \to \mathrm{FS}(\mathbb{Z}^n)$ defined as follows. Let $s \in \mathbb{S}_n^{\mathrm{ns}}/\mathrm{IO}_n$ be a vertex, and choose a sphere $S \subset M_n$ which represents s. As noted above, elements of IO_n act trivially on $H_2(M_n)$. Therefore, the homology class $[S] \in H_2(M_n)$ does not depend on the choice of representative S. We then define $\varphi(s)$ to be the span of $[S]$ in $H_2(M_n)$. It is clear that φ extends to simplices.

Lemma 7.3 *The map* $\varphi\colon \mathbb{S}_n^{\mathrm{ns}}/\mathrm{IO}_n \to \mathrm{FS}(\mathbb{Z}^n)$ *is an isomorphism of simplicial complexes.*

Proof Let $\sigma = \{A_0, \dots, A_\ell\}$ be an ℓ-simplex of $\mathrm{FS}(\mathbb{Z}^n)$. We must show that, up to the action of IO_n, there exists a unique ℓ-simplex $\widetilde{\sigma}$ of $\mathbb{S}_n^{\mathrm{ns}}$ which projects to σ.

Let $v_j \in H_2(M_n)$ be a primitive element generating A_j for $0 \leq j \leq \ell$, and extend this to a basis $\{v_0, \dots, v_{n-1}\}$ for $H_2(M_{n,b}) = \mathbb{Z}^n$. In Appendix B, we will prove Lemma B.2, which says that there exists a collection $\{S_0, \dots, S_{n-1}\}$ of disjoint embedded 2-spheres such that $[S_j] = v_j$ for $0 \leq j \leq n-1$. Then the simplex $\widetilde{\sigma} = \{S_0, \dots, S_\ell\}$ of $\mathbb{S}_n^{\mathrm{ns}}$ maps to the σ under the composition

$$\mathbb{S}_n^{\mathrm{ns}} \to \mathbb{S}_n^{\mathrm{ns}}/\mathrm{IO}_n \xrightarrow{\varphi} \mathrm{FS}(\mathbb{Z}^n).$$

We will now show that $\widetilde{\sigma}$ is unique up to the action of IO_n. Suppose that $\widetilde{\sigma}' = \{S_0', \dots, S_\ell'\}$ is another simplex of $\mathbb{S}_n^{\mathrm{ns}}$ which projects to σ. Since $\widetilde{\sigma}$ and $\widetilde{\sigma}'$ both project to σ, we may order and orient the spheres such that $[S_j] = [S_j']$ for $0 \leq j \leq \ell$. Again by Lemma B.2, we can extend $\{S_1, \dots, S_\ell\}$ and $\{S_1', \dots, S_\ell'\}$ to collections of spheres $\{S_1, \dots, S_n\}$ and $\{S_1', \dots, S_n'\}$ such that $[S_j] = [S_j'] = v_j$ for $0 \leq j \leq n-1$. Notice that splitting M_n along either of these collections reduces M_n to a sphere with $2n$ boundary components. Therefore, there exists some $\mathfrak{f} \in \mathrm{Mod}(M_n)$ such that $\mathfrak{f}(S_j) = S_j'$ for all j. Let $f \in \mathrm{Out}(F_n)$ be the image of f. By construction, $f(\widetilde{\sigma}) = \widetilde{\sigma}'$. Furthermore, f fixes a basis for homology, and so $f \in \mathrm{IO}_n$. This completes the proof. □

This description of $\mathbb{S}_n^{\mathrm{ns}}/\mathrm{IO}_n$ is advantageous because $\mathrm{FS}(\mathbb{Z}^n)$ is known to be $(n-2)$-connected, and hence simply connected for $n \geq 3$. The first proof of this fact is due to Maazen [20] in his unpublished thesis (see [8, Theorem E] for a published proof). Thus, we have shown that $\mathbb{S}_n^{\mathrm{ns}}/\mathrm{IO}_n$ is sufficiently connected.

Corollary 7.4 (Maazen) *The complex* $\mathbb{S}_n^{\mathrm{ns}}/\mathrm{IO}_n$ *is simply connected for* $n \geq 3$.

As indicated in Theorem 7.1, the stabilizers of spheres play an important role in the proof of Theorem D, and so we introduce notation for them here. If S is an isotopy class of embedded sphere in M_n, we denote by $\mathrm{Out}(F_n, S)$ the stabilizer of S in $\mathrm{Out}(F_n)$, and define $\mathrm{IO}_n(S) = \mathrm{Out}(F_n, S) \cap \mathrm{IO}_n$. We now move on to the proof of Theorem D.

Proof of Theorem D We will induct on n. The base cases are easy; IO_1 and IO_2 are both trivial. Suppose now that IO_{n-1} is $\mathrm{Out}(F_{n-1})$-normally generated by handle and commutator drags. We must now show that IO_n is $\mathrm{Out}(F_n)$-normally generated by handle and commutator drags as well. By Theorem 7.1, Proposition 7.2, and Corollary 7.4, it suffices to show that $\mathrm{IO}_n(S)$ is generated by $\mathrm{Out}(F_n)$-conjugates of these drags for all S. Let $S_1, \dots, S_n \subset M_{n,b}$ be a sphere basis, and choose a corresponding geometric free basis $\{\alpha_1, \dots, \alpha_n\}$. Identify the homotopy classes of $\alpha_1, \dots, \alpha_n$ with our fixed basis $x_1, \dots, x_n$ for F_n. Use these bases to construct the handle and commutator drags as in Section 5. Recall that handle drags correspond to the automorphisms M_{ij} of Magnus's generators, and commutator drags correspond to M_{ijk}. We will first show that $\mathrm{IO}_n(S_1)$ is $\mathrm{Out}(F_n, S_1)$-normally generated by handle and commutator drags.

Splitting M_n along S_1 yields a copy of $M_{n-1,2}$. Let N be the tubular neighborhood of S_1 removed in this splitting, and let ∂_1 and ∂_2 be the boundary components of $M_{n-1,2}$. Then this splitting induces a surjective map $\mathrm{Out}(F_{n-1,2}) \to \mathrm{Out}(F_n, S_1)$, which restricts to a map $\iota_*\colon \mathrm{IO}^P_{n-1,2} \to \mathrm{IO}_n(S_1)$, where $P = \{p_1\} = \{\{\partial_1, \partial_2\}\}$. This map is also surjective.

Use the bases $\{\alpha_2, \dots, \alpha_n\}$ and $\{S_2, \dots, S_n\}$ to construct the handle, commutator, boundary commutator, and P-drags in $\mathrm{IO}^P_{n-1,2}$. By our induction hypothesis combined with the proof of Theorem 6.3, these drags $\mathrm{Out}(F_{n-1,2})$-normally generate $\mathrm{IO}^P_{n-1,2}$. Notice that with these choices of drags, the map ι_* takes handle and commutator drags to handle and commutator drags. Moreover, ι_* takes boundary commutator drags in $\mathrm{IO}^P_{n-1,2}$ to commutator drags in $\mathrm{IO}_n(S)$, and takes the P-drag PD^P_i to the handle drag HD_{1i}. Thus, $\mathrm{IO}_n(S_1)$ is $\mathrm{Out}(F_n, S_1)$-normally generated by handle and commutator drags.

Finally, let S be an arbitrary vertex of $\mathbb{S}^{\mathrm{ns}}_n$. Since S is nonseparating, there exists some $f \in \mathrm{Out}(F_n)$ such that $f(S_1) = S$. It follows that

$$\mathrm{IO}_n(S) = f \cdot \mathrm{IO}_n(S_1) \cdot f^{-1}.$$

Since $\mathrm{IO}_n(S_1)$ is $\mathrm{Out}(F_n, S_1)$-normally generated by handle and commutator drags, it follows that $\mathrm{IO}_n(S)$ is generated by $\mathrm{Out}(F_n)$-conjugates of handle and commutator drags, which is what we wanted to show. □

8 Computing the abelianization

In this section, we compute the abelianization of the group $\mathrm{IO}^P_{n,b}$, proving Theorem E. For the Torelli subgroup of the mapping class group of a surface, this was done by Johnson [17]. Some key tools used in this computation are the *Johnson homomorphisms*

$$\tau_{\Sigma_{g,1}}\colon \mathcal{I}(\Sigma_{g,1}) \to \textstyle\bigwedge^3 H \quad \text{and} \quad \tau_{\Sigma_g}\colon \mathcal{I}(\Sigma_g) \to \big(\textstyle\bigwedge^3 H\big)/H,$$

where $H = H_1(\Sigma_{g,b})$. Johnson showed that these homomorphisms are the abelianization maps modulo torsion if $g \geq 3$. For $\mathrm{IA}_n = \mathrm{IO}_{n,1}$, Andreadakis [1] and Bachmuth [3] used an analogous homomorphism $\tau\colon \mathrm{IA}_n \to \mathrm{Hom}\big(H, \bigwedge^2 H\big)$ (where now $H = H_1(F_n) = \mathbb{Z}^n$) to show that

$$H_1(\mathrm{IA}_n) \cong \mathrm{Hom}\big(H, \textstyle\bigwedge^2 H\big) \cong \mathbb{Z}^{n\cdot\binom{n}{2}}.$$

We will begin by recalling the definition of τ, along with the computation of the ranks of $H_1(\mathrm{IA}_n)$ and $H_1(\mathrm{IO}_n)$, and then proceed to the case of multiple boundary components.

The Johnson homomorphism Recall that $\mathrm{Out}(F_{n,1}) \cong \mathrm{Aut}(F_n)$, and the subgroup IA_n is precisely those automorphisms which act trivially on $H_1(F_n) = \mathbb{Z}^n$. The goal is to construct a homomorphism $\tau\colon \mathrm{IA}_n \to \mathrm{Hom}\big(H, \bigwedge^2 H\big)$, where $H = H_1(F_n) = \mathbb{Z}^n$.

First, we claim that the group $[F_n, F_n]/[F_n, [F_n, F_n]]$ is isomorphic to $\bigwedge^2 H$, where $[F_n, F_n]$ denotes the commutator subgroup of F_n. To see this, consider the short exact sequence

$$1 \to [F_n, F_n] \to F_n \to \mathbb{Z}^n \to 1.$$

Passing to the five-term exact sequence in homology, we get the sequence

$$0 \to H_2(\mathbb{Z}^n) \to H_1([F_n, F_n])_{\mathbb{Z}^n} \to H_1(F_n) \to H_1(\mathbb{Z}^n) \to 0,$$

where $H_1([F_n, F_n])_{\mathbb{Z}^n} = [F_n, F_n]/[F_n, [F_n, F_n]]$ denotes the group of coinvariants of $H_1([F_n, F_n])$ with respect to the action of $\mathbb{Z}^n$ (induced by the conjugation action of F_n on $[F_n, F_n]$). The map $H_1(F_n) \to H_1(\mathbb{Z}^n)$ is clearly an isomorphism, and so it follows that the map $H_2(\mathbb{Z}^n) \to [F_n, F_n]/[F_n, [F_n, F_n]]$ is an isomorphism as well. This proves our claim because $H_2(\mathbb{Z}^n) \cong \bigwedge^2 \mathbb{Z}^n$. Let $\rho\colon [F_n, F_n] \to \bigwedge^2 \mathbb{Z}^n$ be the projection. Following the definitions above, we see that ρ is defined by

$$\rho([x, y]) = [x] \wedge [y],$$

where $[x]$ and $[y]$ denote the classes of x and y in H, respectively.

Next, let $f \in \mathrm{IA}_n$. Then $f(x)x^{-1}$ is nullhomologous for all $x \in F_n$, and therefore lies in $[F_n, F_n]$. We define the map $\hat{\tau}_f\colon F_n \to \bigwedge^2 H$ via

$$\hat{\tau}_f(x) = \rho(f(x)x^{-1}).$$

We now check that $\hat{\tau}_f$ is a homomorphism. Let $x, y \in F_n$. Applying the relation $ab = [a, b]ba$, we get

$$\begin{aligned}
\hat{\tau}_f(xy) &= \rho(f(x)f(y)y^{-1}x^{-1}) \\
&= \rho\big([f(x), f(y)y^{-1}]\cdot(f(y)y^{-1})\cdot(f(x)x^{-1})\big) \\
&= [f(x)] \wedge [f(y)y^{-1}] + \hat{\tau}_f(y) + \hat{\tau}_f(y) \\
&= \hat{\tau}_f(y) + \hat{\tau}_f(y),
\end{aligned}$$

since $[f(y)y^{-1}]=0$. This shows that $\hat{\tau}$ is indeed a homomorphism. Furthermore, since $\bigwedge^2 H$ is abelian, the map $\hat{\tau}\colon F_n \to \bigwedge^2 H$ factors through the abelianization, inducing a map $\tau_f\colon H \to \bigwedge^2 H$. Therefore, we have a map

$$\tau\colon \mathrm{IA}_n \to \mathrm{Hom}\big(H, \textstyle\bigwedge^2 H\big)$$

sending f to τ_f. We now check that τ is a homomorphism. Let $f,g \in \mathrm{IA}_n$. Then

$$\tau_{fg}([x]) = \rho(f(g(x))x^{-1}) = \rho\big(f(g(x))(g(x))^{-1}g(x)x^{-1}\big) = \tau_f([g(x)]) + \tau_g([x]) = \tau_f([x]) + \tau_g([x])$$

since g fixes $[x]$. Thus, τ is the desired homomorphism.

We now move on to computing the image of our generators under τ. Since we are dealing with the case of one boundary component, boundary commutator drags are unnecessary since they are products of P-drags. Fix a basepoint $* \in \partial M_{n,1}$, and choose a basis $\{x_1,\dots,x_n\}$ of $\pi_1(M_{n,1},*)$. Construct the handle, commutator, and P-drags with respect to this basis.

Handle drags Recall that the handle drag HD_{ij} acts on $\pi_1(M_{n,1})$ by sending x_i to $x_j x_i x_j^{-1}$, and fixing the remaining basis elements. Therefore,

$$\tau(\mathrm{HD}_{ij})([x_\ell]) = \rho(\mathrm{HD}_{ij}(x_\ell)x_\ell^{-1}) = \begin{cases} 0 & \text{if } \ell \neq i, \\ \rho(x_j x_i x_j^{-1} x_i^{-1}) & \text{if } \ell = i. \end{cases}$$

Thus, $\tau(\mathrm{HD}_{ij})$ is the homomorphism $[x_i] \mapsto [x_j] \wedge [x_i]$ (and all other generators are sent to 0).

Commutator drags Notice that the product of commutator drags $\mathrm{CD}^+_{ijk} \cdot \mathrm{CD}^-_{ijk}$ is equal to a commutator of handle drags. Therefore, only the CD^-_{ijk} are needed in our generating set, and we can disregard the CD^+_{ijk} from now on. Recall that CD^-_{ijk} acts on $\pi_1(M_{n,1})$ by sending x_i to $[x_j,x_k]x_i$. Therefore,

$$\tau(\mathrm{CD}^-_{ijk})([x_\ell]) = \rho(\mathrm{CD}^-_{ijk}(x_\ell)x_\ell^{-1}) = \begin{cases} 0 & \text{if } \ell \neq i, \\ \rho([x_j,x_k]) & \text{if } \ell = i. \end{cases}$$

It follows that $\tau(\mathrm{CD}^-_{ijk})$ is the map $[x_i] \mapsto [x_j] \wedge [x_k]$.

***P*-drags** Next, we note that the product

$$\mathrm{PD}_j \cdot \mathrm{HD}_{1j} \cdots \mathrm{HD}_{nj} \tag{5}$$

is trivial in IA_n. For a justification of this fact, see the proof of the claim at the end of Theorem A.2. It follows that the P-drags are also redundant in our generating set for IA_n, and can be removed.

Abelianization of IA_n To compute the rank of the abelianization of IA_n, we use the following lemma.

Lemma 8.1 *Let G be a group and S a finite generating set for G. Suppose that $\varphi\colon G \to \mathbb{Z}^{|S|}$ is a surjective homomorphism. Then $H_1(G) \cong \mathbb{Z}^{|S|}$.*

Proof Let $F(S)$ denote the free group on the set S. Since $\mathbb{Z}^{|S|}$ is abelian, the homomorphism φ factors through the abelianization to give a map $\bar{\varphi}\colon H_1(G) \to \mathbb{Z}^n$, which is also surjective. Additionally, by the universal property of free groups, we have a map $\psi\colon F(S) \to G$. Passing to the abelianizations induces a map $\bar{\psi}\colon H_1(F(S)) \to H_1(G)$. Since S is a generating set for G, this map is also surjective. It follows that $\bar{\varphi} \circ \bar{\psi}$ is a surjective map between free abelian groups of equal rank, and is hence an isomorphism. Thus, $\bar{\varphi}$ is an isomorphism as well. □

From the discussion in the preceding paragraphs, we have a generating set for IA_n of size

$$\#(\text{handle drags}) + \#(\text{commutator drags}) = n(n-1) + n \cdot \binom{n-1}{2} = n \cdot \binom{n}{2}$$

(since P-drags can be written as a product of handle drags), and the image of this generating set spans $\mathrm{Hom}\bigl(H, \bigwedge^2 H\bigr)$, which also has dimension $n \cdot \binom{n}{2}$. Therefore, by Lemma 8.1, the group $H_1(\mathrm{IA}_n)$ has rank $n \cdot \binom{n}{2}$.

To compute the rank of $H_1(\mathrm{IO}_n)$, consider the quotient map $\mathrm{IA}_n \to \mathrm{IO}_n$, whose kernel is the subgroup of inner automorphisms (or P-drags under our geometric interpretation of IA_n). We compute the image of a P-drag under τ:

$$\tau(\mathrm{PD}_i)([x_\ell]) = \rho(\mathrm{PD}_i(x_\ell)x_\ell^{-1}) = \rho(x_i^{-1}x_\ell x_i x_\ell^{-1}) = [x_\ell] \wedge [x_i].$$

Since $\tau(\mathrm{PD}_i)$ is nontrivial, τ does not descend to a map $\mathrm{IO}_n \to \mathrm{Hom}\bigl(H, \bigwedge^2 H\bigr)$. However, the images $\{\tau(\mathrm{PD}_i)\}$ span a subgroup of $\mathrm{Hom}\bigl(H, \bigwedge^2 H\bigr)$ isomorphic to H (where the isomorphism is given by $[h] \mapsto ([x_\ell] \mapsto [x_\ell] \wedge [h])$). So, τ induces a map $\mathrm{IO}_n \to \mathrm{Hom}\bigl(H, \bigwedge^2 H\bigr)/H$. Just as the element given in (5) is trivial in IA_n, the product $\mathrm{HD}_{1j} \cdots \mathrm{HD}_{nj}$ is trivial in IO_n for all $j \in \{1, \dots, n\}$. Thus, we may throw out n handle drags from our generating set to obtain a generating set for IO_n of size $n \cdot \binom{n}{2} - n$. Since $\mathrm{Hom}\bigl(H, \bigwedge^2 H\bigr)$ has rank $n \cdot \binom{n}{2} - n$, Lemma 8.1 implies that $H_1(\mathrm{IO}_n)$ has rank $n \cdot \binom{n}{2} - n$ as well. This verifies Theorem 1.2 from the introduction.

Multiple boundary components We now move on to the case of multiple boundary components. Just as we did when constructing our drags in Section 5, fix an ordering $P = \{p_1, \dots, p_{|P|}\}$ and an ordering $p_r = \{\partial_1^r, \dots, \partial_{b_r}^r\}$ for each $p_r \in P$. We cap off the boundary components of each $p \in P$ as follows:

- If $|p| = 1$, we attach a copy of $M_{1,1}$ to the single boundary component of p.
- If $|p| = k > 1$, we attach a copy of $M_{0,k}$ to these k boundary components.
- If $p = p_1$, we follow the same rules as above, except we introduce an additional boundary component in the piece glued to p.

Capping off the boundary components in this way gives an embedding

$$\iota\colon M_{n,b} \hookrightarrow M_{m,1},$$

where the boundary component of $M_{m,1}$ lies in the piece attached to p_1. This embedding induces a map $\iota_*\colon \mathrm{IO}_{n,b}^P \to IA_m$. We obtain a similar map $\mathrm{IA}_m \to \mathrm{IO}_m$ by attaching a disk to the boundary component of $M_{m,1}$. In Appendix A, we will prove Theorem A.2, which says that the composition

$$\mathrm{IO}_{n,b}^P \xrightarrow{\iota_*} \mathrm{IA}_m \to \mathrm{IO}_m$$

is injective. It follows that ι_* is injective as well. Therefore, to compute the rank of the abelianization of $\mathrm{IO}_{n,b}^P$, it suffices to compute the rank of the abelianization of its image in IA_m. Let $H = H_1(M_{m,1})$, and let $\tau_*\colon \mathrm{IO}_{n,b}^P \to \mathrm{Hom}\big(H, \bigwedge^2 H\big)$ denote the composition

$$\mathrm{IO}_{n,b}^P \xrightarrow{\iota_*} \mathrm{IA}_m \xrightarrow{\tau} \mathrm{Hom}\big(H, \textstyle\bigwedge^2 H\big).$$

Our goal now becomes computing the images of handle, commutator, boundary commutator, and P-drags under τ_*.

Choosing a basis To carry out this computation, it will be helpful to choose bases for $\pi_1(M_{n,b})$ and $\pi_1(M_{m,1})$ carefully. For simplicity, we will assume that $|p_1| > 1$. The case of $|p_1| = 1$ is more straightforward. Fix a basepoint $z_1^1 \in \partial_1^1$, a sphere basis $\{S_1, \dots, S_n\}$, and a geometric free basis $\{x_1, \dots, x_n\}$ for $\pi_1(M_{n,b}, z_1^1)$. We define our drags in $\mathrm{IO}_{n,b}^P$ with respect to these bases.

Next, choose a basepoint $z \in \partial M_{m,1}$, and an oriented arc $\alpha_1 \subset M_{m,1} \setminus \mathrm{int}(M_{n,b})$ from z to z_1^1 (this is possible since the boundary component of $M_{m,1}$ lies on the piece attached to p_1). For $i \in \{1, \dots, n\}$, let $y_i = \alpha_1 x_i \alpha_1^{-1}$. Then $\{y_1, \dots, y_n\}$ is a partial basis for $\pi_1(M_{m,1}, z)$. We wish to extend this to a full basis. Throughout the definition of this extended basis, we encourage the reader to follow along in Figure 12.

For each boundary component ∂_s^r of $M_{n,b}$, fix a point $z_s^r \in \partial_s^r$ (leaving z_1^1 as before). For each $z_s^r \neq z_1^1$, let β_s^r be the unique oriented arc (up to isotopy) in $M_{n,b} \setminus \bigcup S_i$ from z_1^1 to z_s^r. For $s \in \{2, \dots, b_1\}$, define $y_s^1 = \alpha_1 \beta_s^1 \alpha_s^{-1}$. In Figure 12, the loops y_1^1 and y_2^1 are of this form.

Next, let $r > 1$. If $|p_r| = 1$, let γ_1^r be an oriented loop based at z_1^r which generates the fundamental group of the copy of $M_{1,1}$ attached to p_r. Then we define $y_1^r = \alpha_1 \beta_1^r \gamma_1^1 (\beta_1^r)^{-1} (\alpha_1)^{-1}$. In Figure 12, the curve y_1^3 is an example of such a loop.

On the other hand, suppose $|p_r| > 1$. For $s \in \{2, \dots, b_r\}$, let γ_s^r be the unique (up to isotopy) oriented curve in $M_{m,1} \setminus \mathrm{int}(M_{n,b})$ from z_1^r to z_s^r. Then, define

$$y_s^r = \alpha_1 \beta_1^r \gamma_s^r (\beta_s^r)^{-1} (\alpha_1)^{-1}.$$

The curve y_2^2 is an example of this type of loop in Figure 12.

Let $Y = \{y_1 \dots, y_n\}$. For $r \in \{1, \dots, |P|\}$, let $Y_r = \{y_1^r\}$ if $|p_r| = 1$, and $Y_r = \{y_2^r, \dots, y_{b_r}^r\}$ otherwise. Then the collection

$$Y \cup Y_1 \cup \dots \cup Y_{|P|}$$

forms a free basis for $\pi_1(M_{m,1}, z)$.

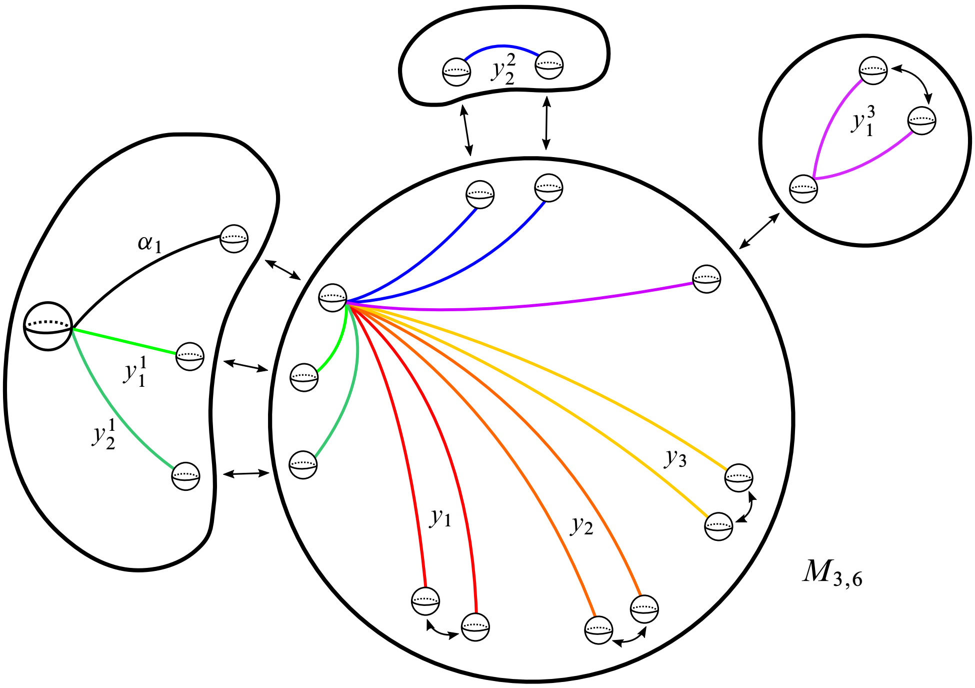


Figure 12: $M_{3,6}$ with the partition $P = \{\{\partial_1^1, \partial_2^1, \partial_3^1\}, \{\partial_1^2, \partial_2^2\}, \{\partial_1^3\}\}$ embedded into $M_{7,1}$. The loops $\{y_1, y_2, y_3\}$ are freely homotopic to a basis for $\pi_1(M_{3,6})$, and this basis has been extended to a basis $\{y_1, y_2, y_3, y_1^1, y_2^1, y_2^2, y_1^3\}$ of $\pi_1(M_{7,1}, z)$.

Computations and relations We now move on to the computation of the images of our collection of drags under τ_*. These computations are straightforward, and are summarized in Table 1. We see from these computations that there is a relation between the images of P-drags and Handle Drags. Namely,

$$\sum_{r=1}^{|P|} \tau_*(\mathrm{PD}_j^r) = -\sum_{i=1}^{n} \tau_*(\mathrm{HD}_{ij}) \tag{6}$$

for all $j \in \{1, \dots, n\}$. As we saw in the case of one boundary component, this is because

$$\mathrm{PD}_j^1 \cdots \mathrm{PD}_j^{|P|} \cdot \mathrm{HD}_{1j} \cdot \mathrm{HD}_{nj} = 1 \tag{7}$$

in $\mathrm{IO}_{n,b}^P$. Additionally, we see a relation between the image of boundary commutator drags:

$$\sum_{s=1}^{b_r} \tau_*(\mathrm{BCD}_{ij}^{rs}) = 0 \tag{8}$$

for all $r \in \{1, \dots, |P|\}$ and $i, j \in \{1, \dots, n\}$ with $i < j$. This relation holds because

$$\mathrm{BCD}_{ij}^{r1} \cdots \mathrm{BCD}_{ij}^{rb_r} = [\mathrm{PD}_i^r, \mathrm{PD}_j^r] \tag{9}$$

in $\mathrm{IO}_{n,b}^P$.

drag	action on π_1	image under τ_*
HD_{ij}	$y_i \mapsto y_j y_i y_j^{-1}$	$[y_i] \mapsto [y_j] \wedge [y_i]$
CD^-_{ijk}	$y_i \mapsto [y_j, y_k] y_i$	$[y_i] \mapsto [y_j] \wedge [y_k]$
BCD^{rs}_{jk} $(r, s > 1)$	$y^r_s \mapsto y^r_s [y_j, y_k]^{-1}$	$[y^r_s] \mapsto [y_k] \wedge [y_j]$
BCD^{r1}_{jk} $(r > 1)$	$y^r_s \mapsto [y_j, y_k] y^r_s$ $(s > 1)$	$[y^r_s] \mapsto [y_j] \wedge [y_k]$ $(s > 1)$
BCD^{1s}_{jk} $(s > 1)$	$y^1_s \mapsto y^1_s [y_j, y_k]$	$[y^1_s] \mapsto [y_j] \wedge [y_k]$
BCD^{11}_{jk}	$y \mapsto [y_j, y_k]^{-1} y [y_j, y_k]$ $(y \notin Y_1)$ $y^1_s \mapsto [y_j, y_k]^{-1} y^r_s$ $(s > 1)$	$[y] \mapsto 0$ $(y \notin Y_1)$ $[y^1_s] \mapsto [y_k] \wedge [y_j]$ $(s > 1)$
PD^r_j $(r > 1)$	$y^r_s \mapsto y_j y^r_s y_j^{-1}$ $(s > 1)$	$[y^r_s] \mapsto [y_j] \wedge [y^r_s]$ $(s > 1)$
PD^1_j	$y \mapsto y_j^{-1} y y_j$ $(y \notin Y_1)$	$[y] \mapsto [y] \wedge [y_j]$ $(y \notin Y_1)$

Table 1: Computing the image of drags under τ_*.

Contributions to abelianization From the computations and relations above, we see that the handle drags and commutator drags together still contribute $n \cdot \binom{n}{2}$ dimensions to the abelianization of $\mathrm{IO}^P_{n,b}$. There are $b \cdot \binom{n}{2}$ boundary commutator drags, but the relations given in (8) kill off $|P| \cdot \binom{n}{2}$ of these in the abelianization (though we can also remove this many elements from our generating set by using (9)). Finally, the number of P-drags is $|P| \cdot n$, but n of them are killed in the abelianization by (6) (and again, we may remove n elements from our generating set by (7)). Adding these all together, we find that the image of $\tau_* \colon \mathrm{IO}^P_{n,b} \to \mathrm{Hom}\bigl(H, \bigwedge^2 H\bigr)$ has rank

$$R = n \cdot \binom{n}{2} + \left(b \cdot \binom{n}{2} - |P| \cdot \binom{n}{2}\right) + (|P| \cdot n - n).$$

Moreover, we can reduce our generating set (using (7) and (9)) to a set of size R as well. Thus, by Lemma 8.1, the group $H_1(\mathrm{IO}^P_{n,b})$ has rank R, which proves Theorem E.

Appendix A Injectivity of the inclusion map

We end this paper with a proof of the following facts, which are surely known to experts, but for which we do not know a reference. They are significant because they allow us to realize the groups $\mathrm{Out}(F_{n,b})$ (and hence $\mathrm{IO}^P_{n,b}$) as subgroups of $\mathrm{Out}(F_m)$. We will begin with a low-genus case.

Lemma A.1 *The induced map $\iota_* \colon \mathrm{Out}(F_{1,1}) \to \mathrm{Out}(F_m)$ is injective for any embedding $\iota \colon M_{1,1} \hookrightarrow M_m$.*

Proof By Laudenbach [18], the group $\mathrm{Out}(F_{1,1}) \cong \mathrm{Aut}(F_1) \cong \mathbb{Z}/2$, where the nontrivial element $f \in \mathrm{Out}(F_{1,1})$ acts on $\pi_1(M_{1,1}, x) \cong \mathbb{Z}$ by inverting the generator. Therefore, $\iota_*(f) \in \mathrm{Out}(F_m)$ is the class of the automorphism

$$\begin{cases} x_1 \mapsto x_1^{-1}, \\ x_j \mapsto x_j & \text{if } j > 1. \end{cases}$$

This automorphism is not an inner automorphism for any $m \geq 1$, so i_* is injective. □

Theorem A.2 *Fix* $n, b \geq 1$ *such that* $(n, b) \neq (1, 1)$*, and let* $\iota\colon M_{n,b} \hookrightarrow M_m$ *be an embedding. The induced map* $\iota_*\colon \mathrm{Out}(F_{n,b}) \to \mathrm{Out}(F_m)$ *is injective if and only if no component of* $M_m \setminus \mathrm{int}(M_{n,b})$ *is diffeomorphic to a* 3*-disk.*

Proof Suppose first that some component of $M_m \setminus \mathrm{int}(M_{n,b})$ is diffeomorphic to a disk, and let ∂ be the boundary component of $M_{n,b}$ capped off by this disk. By the Birman exact sequence (Theorem 4.2), dragging this boundary component along any nontrivial loop will give a nontrivial element in the kernel of ι_*.

Suppose now that no component of $M_m \setminus \mathrm{int}(M_{n,b})$ is a disk. We will first prove the theorem in the case $b = 1$, and then move on to the general result.

Case 1 Suppose we have an embedding $\iota\colon M_{n,1} \hookrightarrow M_m$. Since no component of $M_m \setminus \mathrm{int}(M_{n,b})$ is a disk, $m > n$. If $n = 1$, then we are done by Lemma A.1, so we may assume that $n > 1$. Fix a basepoint x on the boundary of $M_{n,1}$, and choose a free basis $\{x_1, \dots, x_n\}$ of $\pi_1(M_{n,1}, x)$. The embedding ι induces an injection $\pi_1(M_{n,b}, x) \hookrightarrow \pi_1(M_m, x)$ which identifies $\pi_1(M_{n,1}, x)$ with a free summand of $\pi_1(M_m, x)$. This allows us to extend $\{x_1, \dots, x_n\}$ to a free basis $\{x_1, \dots, x_m\}$ of $\pi_1(M_m, x)$. Given $f \in \mathrm{Out}(F_{n,1}) \cong \mathrm{Aut}(F_n)$, the image $\iota_*(f) \in \mathrm{Out}(F_m)$ is the class of the automorphism $\varphi \in \mathrm{Aut}(F_m)$ generated by

$$\varphi\colon x_i \mapsto \begin{cases} f(x_i) & \text{if } 1 \leq i \leq n, \\ x_i & \text{if } n < i \leq m. \end{cases}$$

Suppose that φ is an inner automorphism. If $m > n + 1$, then φ fixes at least two generators of F_m, and thus must be trivial. It follows that f is trivial as well. On the other hand, if $m = n + 1$, then φ fixes x_m. Since φ is inner, φ must conjugate by a power of x_m. However, if φ conjugates by a nontrivial power of x_m, then f would not act as an automorphism on $\langle x_1, \dots, x_n \rangle \subset F_m$, which is a contradiction. Thus, φ is trivial, and so f is trivial as well.

In summary, we have shown that φ is an inner automorphism if and only if f is trivial, which implies that ι_* is injective.

Case 2 Next, suppose that $\iota\colon M_{n,b} \hookrightarrow M_m$ is an embedding, where $b > 1$. Let $\partial_1, \dots, \partial_b$ be the boundary components of $M_{n,b}$. Let $\Sigma \subset M_{n,b}$ be a 2-sphere which separates $M_{n,b}$ into $M_{n,1}$ and $M_{0,b+1}$ (see Figure 13). Then we have a composition of inclusions

$$M_{n,1} \hookrightarrow M_{n,b} \hookrightarrow M_m.$$

Let $\kappa_*\colon \mathrm{Out}(F_{n,1}) \to \mathrm{Out}(F_{n,b})$ be the map induced by inclusion. By the preceding case, $\iota_* \circ \kappa_*$ is injective. Let $f \in \mathrm{Out}(F_{n,b})$, and suppose that $\iota_*(f) = \mathrm{id}$. By repeated applications of the Birman exact sequence (Theorem 4.2), f has the form $f = p_1 p_2 \cdots p_b \cdot \kappa_*(g)$, where $g \in \mathrm{Out}(F_{n,1}) \cong \mathrm{Aut}(F_n)$ and $p_j \in \mathrm{Out}(F_{n,1})$ is a boundary drag of ∂_j along a loop β_j. Fix a basepoint $x \in \Sigma$, and let $\gamma_j \in \pi_1(M_{n,b}, x)$ be representative of the free homotopy class of β_j. Choose a free basis $\{x_1, \dots, x_n\}$ for $\pi_1(M_{n,1}, x)$. Extend this to a free basis $\{x_1, \dots, x_m\}$ for $\pi_1(M_m, x)$ such that for each $i > n$, the loop x_i intersects

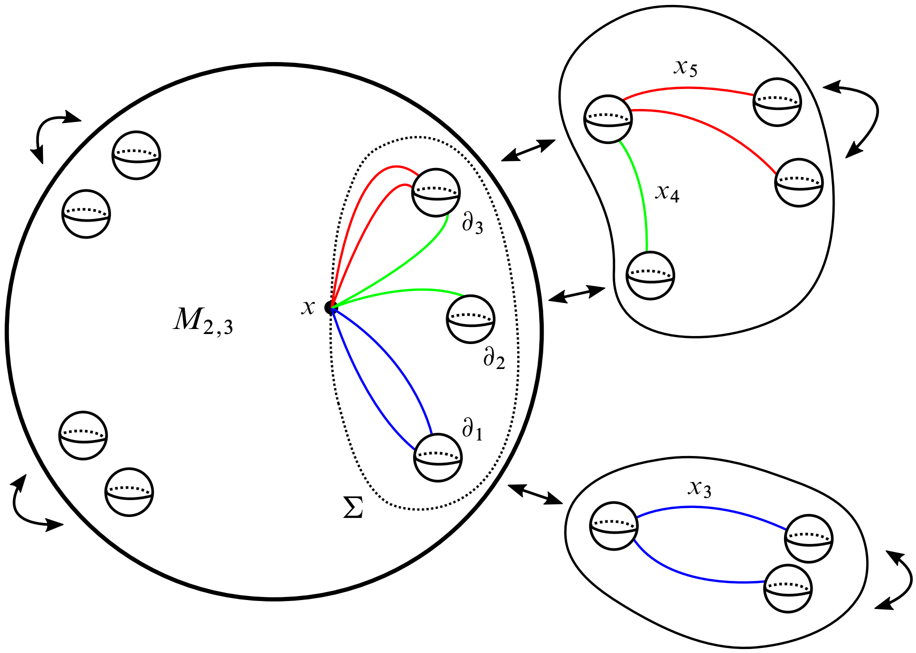


Figure 13: $M_{2,3}$ embedded inside M_5. For clarity, x_1 and x_2 are not shown, but they lie entirely on the opposite side of Σ from x_3, x_4, and x_5.

the set $\bigcup_{j=1}^{b} \partial_j$ exactly twice: once when exiting $M_{n,b}$, and once when reentering (see Figure 13). For $i > n$, let $\partial_{\ell(i)}$ be the boundary component through which α_i leaves $M_{n,b}$, and let $\partial_{r(i)}$ be the boundary component through which it returns. Then $\iota_*(f)$ is the class of the automorphism $\varphi \in \mathrm{Aut}(F_m)$ given by

$$\varphi\colon x_i \mapsto \begin{cases} g(x_i) & \text{for } 1 \le i \le n, \\ \gamma_{\ell(i)} x_i \gamma_{r(i)}^{-1} & \text{for } n < i \le m. \end{cases}$$

By assumption, this automorphism is an inner automorphism. Suppose that φ conjugates by a reduced word w in the x_i. Since g is an automorphism of $\langle x_1, \dots, x_n \rangle \subset F_m$, it follows that $w \in \langle x_1, \dots x_n \rangle$. We will show that this implies that f is trivial by induction on the reduced word length of w.

For the base case, suppose that the word length of w is 0. Then w and φ are both trivial. Since $\iota_* \circ \kappa_*$ is injective, g is trivial as well. Suppose now that some γ_j is not nullhomotopic. Since no component of $M_m \setminus \mathrm{int}(M_{n,b})$ is a disk, there exists some x_i which passes through ∂_j, where $i > n$. In other words, either $\ell(i) = j$ or $r(i) = j$. This is a contradiction because then $\varphi(x_i) = \gamma_{\ell(i)} x_i \gamma_{r(i)}^{-1} \neq x_i$. Thus, all γ_j are nullhomotopic, and so f is trivial. This completes the base case.

Next, suppose that w has positive word length, and let $x_i^{\pm 1}$ be the last letter in the reduced form of w. Then, $w = w' x_i^{\pm 1}$, where the length of w' is less than that of w. To avoid notational complexity, we will assume that $x_i^{\pm 1} = x_1$, but the same argument works for any other x_i. Consider the element

$$h := \mathrm{HD}_{21}\, \mathrm{HD}_{31} \cdots \mathrm{HD}_{n1} \cdot q_1 \cdots q_b \in \mathrm{Out}(F_{n,b}),$$

where HD_{i1} is the handle drag of the i^{th} handle about the first handle (see Section 5) and q_j is obtained by dragging ∂_j about a loop in the free homotopy class of x_1. By construction, $\iota_*(h) \in \mathrm{Out}(F_m)$ is the class of the automorphism which conjugates by x_1. Therefore, $\iota_*(h^{-1} f)$ is the class of the automorphism which conjugates by w'. By our induction hypothesis, this implies that $h^{-1} f$ is trivial.

Claim *The element h is trivial.*

Proof Let $\Sigma' \subset M_{n,b}$ be a 2-sphere which separates $M_{n,b}$ into $M_{1,1}$ and $M_{n-1,b+1}$, where the $M_{1,1}$ is the handle containing x_1. Let $\lambda_*\colon \mathrm{Out}(F_{1,1}) \to \mathrm{Out}(F_{n,b})$ be the map induced by this inclusion. Notice that $h = \lambda_*(q)$, where $q \in \mathrm{Out}(F_{1,1})$ drags the boundary component of $M_{1,1}$ about the nontrivial loop in the positive direction. We saw in the proof of Lemma A.1 that $\mathrm{Out}(F_{1,1}) \cong \mathbb{Z}/2$, and the nontrivial element acts on $\pi_1(M_{1,1})$ by inversion. However, the element q acts trivially on $\pi_1(M_{1,1})$, and is thus trivial itself. It follows that h is trivial as well. □

Combining the claim with the fact that $h^{-1} f$ is trivial, we find that f is trivial. This completes the induction, and so we conclude that ι_* is injective. □

Appendix B Realizing homology classes as spheres

In this section, we prove a result used in the proof of Lemma 7.3 which involves realizing bases of $H_2(M_n)$ as collections of 2-spheres. Recall that $H_2(M_n) = \mathbb{Z}^n$. This identification induces a homomorphism $\eta\colon \mathrm{Mod}(M_n) \to \mathrm{GL}_n(\mathbb{Z})$ which takes a mapping class to its action on homology.

Lemma B.1 *The map $\eta\colon \mathrm{Mod}(M_n) \to \mathrm{GL}_n(\mathbb{Z})$ is surjective.*

Proof First, notice that $H^1(M_n) = \mathbb{Z}^n$. This identification also induces a homomorphism

$$\eta'\colon \mathrm{Mod}(M_n) \to \mathrm{GL}_n(\mathbb{Z})$$

which is well-known to be surjective. Indeed, this map factors as

$$\mathrm{Mod}(M_n) \xrightarrow{q} \mathrm{Out}(F_n) \xrightarrow{\varphi} \mathrm{GL}_n(\mathbb{Z}),$$

where q is the quotient map, and φ sends an automorphism class to its action on H^1. Therefore, if we choose our identifications $H^1(M_n) = \mathbb{Z}^n$ and $H_2(M_n) = \mathbb{Z}^n$ to agree with Poincaré duality, then η and η' are the same map. Thus, η is surjective. □

Lemma B.2 *Let $\{v_1, \dots, v_n\}$ be a basis for $H_2(M_n) = \mathbb{Z}^n$, and let $A = \{S_1, \dots, S_\ell\}$ be a collection of disjoint embedded oriented 2-spheres in M_n which satisfy $[S_j] = v_j$ for $1 \le j \le \ell$. Then A can be extended to a collection $\bar{A} = \{S_1, \dots, S_n\}$ of disjoint embedded oriented 2-spheres such that $[S_j] = v_j$ for $1 \le j \le n$.*

Proof We will induct on n. The base case $n = 0$ is trivial. So assume $n > 0$, and let $\{v_1, \dots, v_n\}$ and $A = \{S_1, \dots, S_\ell\}$ be as stated. There are two cases.

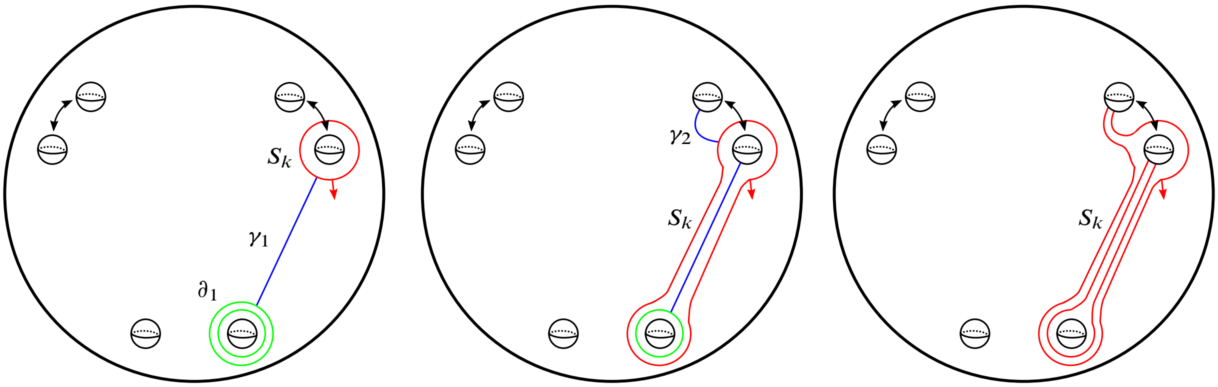


Figure 14: Surgering boundary spheres onto S_k.

First, suppose that $\ell = 0$. If we identity $H_2(M_n)$ with $\mathbb{Z}^n$, then by Lemma B.1 the resulting map $\eta\colon \mathrm{Mod}(M_n) \to \mathrm{GL}_n(\mathbb{Z})$ is surjective. Choose any collection $\Sigma_1, \dots, \Sigma_n \subset M_n$ of disjoint embedded 2-spheres such that $M_n \setminus (\Sigma_1 \cup \cdots \cup \Sigma_n)$ is connected. Then $\{[\Sigma_1], \dots, [\Sigma_n]\}$ is a basis for $H_2(M_n)$. Since $\mathrm{GL}_n(\mathbb{Z})$ acts transitively on ordered bases of $\mathbb{Z}^n$ and the map η is surjective (Lemma B.1), there exists some $\mathfrak{f} \in \mathrm{Mod}(M_n)$ such that $\eta(\mathfrak{f}) \cdot [\Sigma_j] = v_j$ for all $1 \leq j \leq n$. In other words, $[\mathfrak{f}(\Sigma_j)] = v_j$, and so $\{\mathfrak{f}(\Sigma_1), \dots, \mathfrak{f}(\Sigma_n)\}$ is the desired collection of spheres.

Next, suppose that $\ell > 0$. Splitting M_n along S_1 gives an embedding $\iota\colon M_{n-1,2} \hookrightarrow M_n$. Notice that the induced map $\iota_H\colon H_2(M_{n-1,2}) \to H_2(M_n)$ is an isomorphism. Let $w_j = \iota_H^{-1}(v_j)$ for $1 \leq j \leq n$, and let ∂ and ∂' be the boundary components of $M_{n-1,2}$. Capping the two boundary components of $M_{n-1,2}$ with disks D and D', we get another embedding $\iota'\colon M_{n-1,2} \hookrightarrow M_{n-1}$. This embedding induces a surjective map $\iota'_H\colon H_2(M_{n-1,2}) \to H_2(M_{n-1})$ whose kernel is generated by $[\partial]$. Let $w'_j = \iota'_H(w_j)$ for $2 \leq j \leq n$, and let $S'_k = \iota'(S_k)$ for $2 \leq k \leq \ell$. By our induction hypothesis, we can extend the collection $\{S'_2, \dots, S'_\ell\}$ to a collection $\{S'_2, \dots, S'_{n-1}\}$ of disjoint embedded oriented 2-spheres in M_{n-1} such that $[S'_j] = w'_j$ for $2 \leq j \leq n$. Moreover, since the disks D and D' used to cap the boundary components of $M_{n-1,2}$ are contractible, we may isotope $S'_{\ell+1}, \dots, S'_{n-1}$ such that they are disjoint from D and D'. Let $S_j = (\iota')^{-1}(S'_j)$ for $\ell + 1 \leq j \leq n$. If $[S_k] = w_k$ for all k, then $\{S_1, \dots, S_n\}$ is the desired collection, and we are done. However, since the kernel of ι'_H is generated by $[\partial]$, we have

$$[S_k] = w_k + c_k[\partial],$$

where $c_k \in \mathbb{Z}$. Note that $c_k = 0$ for $2 \leq k \leq \ell$. To fix this, we may surger parallel copies of ∂ or ∂' onto S_k such that it has the correct homology class. The process is as follows (see Figure 14):

(i) If $c_k > 0$, take c_k parallel copies of ∂', which we denote by $\partial_1, \dots, \partial_{c_k}$. If instead $c_k < 0$, take $\partial_1, \dots, \partial_{c_k}$ to be parallel copies of ∂. Order the ∂_j such that ∂_1 is furthest from its respective boundary component, then ∂_2, and so on.

(ii) Let γ_1 be a properly embedded arc connecting the positive side of S_k to ∂_1 which does not intersect any of the other S_j or ∂_j.

(iii) Surger S_k and ∂_1 together via a tube running along γ_1.

(iv) Repeat steps (ii) and (iii) for the remaining ∂_j.

Once we have carried out this process for all the S_k, we will have obtained a collection $\{S_2, \dots, S_n\}$ of spheres whose homology classes are exactly $w_2, \dots, w_n$. Thus, $\{S_1, \dots, S_n\}$ is the desired collection of 2-spheres. □

References

[1] **S Andreadakis**, *On the automorphisms of free groups and free nilpotent groups*, Proc. Lond. Math. Soc. 15 (1965) 239–268 MR Zbl

[2] **M A Armstrong**, *On the fundamental group of an orbit space*, Proc. Cambridge Philos. Soc. 61 (1965) 639–646 MR Zbl

[3] **S Bachmuth**, *Induced automorphisms of free groups and free metabelian groups*, Trans. Amer. Math. Soc. 122 (1966) 1–17 MR Zbl

[4] **J E Banks**, *The Birman exact sequence for* 3*-manifolds*, Bull. Lond. Math. Soc. 49 (2017) 604–629 MR Zbl

[5] **M Bestvina**, **K-U Bux**, **D Margalit**, *Dimension of the Torelli group for* $\mathrm{Out}(F_n)$, Invent. Math. 170 (2007) 1–32 MR Zbl

[6] **T Brendle**, **N Broaddus**, **A Putman**, *The mapping class group of connect sums of* $S^2 \times S^1$, Trans. Amer. Math. Soc. 376 (2023) 2557–2572 MR Zbl

[7] **M R Bridson**, **K Vogtmann**, *Automorphism groups of free groups, surface groups and free abelian groups*, from "Problems on mapping class groups and related topics", Proc. Sympos. Pure Math. 74, Amer. Math. Soc., Providence, RI (2006) 301–316 MR Zbl

[8] **T Church**, **B Farb**, **A Putman**, *Integrality in the Steinberg module and the top-dimensional cohomology of* $\mathrm{SL}_n\mathcal{O}_K$, Amer. J. Math. 141 (2019) 1375–1419 MR Zbl

[9] **M Day**, **A Putman**, *A Birman exact sequence for* $\mathrm{Aut}(F_n)$, Adv. Math. 231 (2012) 243–275 MR Zbl

[10] **M Day**, **A Putman**, *The complex of partial bases for* F_n *and finite generation of the Torelli subgroup of* $\mathrm{Aut}(F_n)$, Geom. Dedicata 164 (2013) 139–153 MR Zbl

[11] **M Day**, **A Putman**, *A Birman exact sequence for the Torelli subgroup of* $\mathrm{Aut}(F_n)$, Int. J. Algebra Comput. 26 (2016) 585–617 MR Zbl

[12] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Math. Ser. 49, Princeton Univ. Press (2012) MR Zbl

[13] **A Hatcher**, *Homological stability for automorphism groups of free groups*, Comment. Math. Helv. 70 (1995) 39–62 MR Zbl

[14] **A Hatcher**, **N Wahl**, *Stabilization for the automorphisms of free groups with boundaries*, Geom. Topol. 9 (2005) 1295–1336 MR Zbl

[15] **A Hatcher**, **N Wahl**, *Stabilization for mapping class groups of* 3*-manifolds*, Duke Math. J. 155 (2010) 205–269 MR Zbl

[16] **C A Jensen**, **N Wahl**, *Automorphisms of free groups with boundaries*, Algebr. Geom. Topol. 4 (2004) 543–569 MR Zbl

[17] **D Johnson**, *The structure of the Torelli group, III: The abelianization of $\mathcal{T}$*, Topology 24 (1985) 127–144 MR Zbl

[18] **F Laudenbach**, *Sur les 2-sphères d'une variété de dimension* 3, Ann. of Math. 97 (1973) 57–81 MR

[19] **F Laudenbach**, *Topologie de la dimension trois: homotopie et isotopie*, Astérisque 12, Soc. Math. France, Paris (1974) MR Zbl

[20] **H Maazen**, *Homology stability for the general linear group*, PhD thesis, Utrecht University (1951) Available at https://dspace.library.uu.nl/handle/1874/237657

[21] **W Magnus**, *Über n-dimensionale Gittertransformationen*, Acta Math. 64 (1935) 353–367 MR Zbl

[22] **A Putman**, *Cutting and pasting in the Torelli group*, Geom. Topol. 11 (2007) 829–865 MR Zbl

[23] **A Putman**, *Obtaining presentations from group actions without making choices*, Algebr. Geom. Topol. 11 (2011) 1737–1766 MR Zbl

[24] **A Putman**, *The commutator subgroups of free groups and surface groups*, Enseign. Math. 68 (2022) 389–408 MR Zbl

[25] **W Tomaszewski**, *A basis of Bachmuth type in the commutator subgroup of a free group*, Canad. Math. Bull. 46 (2003) 299–303 MR Zbl

[26] **N Wahl**, *From mapping class groups to automorphism groups of free groups*, J. Lond. Math. Soc. 72 (2005) 510–524 MR Zbl

Department of Mathematics, University of Notre Dame
Notre Dame, IN, United States

landgrafjacob@gmail.com

Received: 18 November 2021 Revised: 15 September 2023

Algebraic & Geometric Topology **25**:1 (2025) 39–54
DOI: 10.2140/agt.2025.25.39
Published: 5 March 2025

Hyperbolic groups with logarithmic separation profile

NIR LAZAROVICH
CORENTIN LE COZ

We prove that hyperbolic groups with logarithmic separation profiles split over cyclic groups. This shows that such groups can be inductively built from Fuchsian groups and free groups by amalgamations and HNN extensions over finite or virtually cyclic groups. However, we show that not all groups admitting such a hierarchy have logarithmic separation profile by providing an example of a surface amalgam over a cyclic group with superlogarithmic separation profile.

20E06, 20F65, 20F67, 20F69, 51F30; 14H30

1 Introduction

The separation profile was first introduced by Benjamini, Schramm and Timár [1] in 2012. It measures large scale connectivity of infinite graphs, in the spirit of the celebrated theorem of Lipton and Trajan for planar graphs [12].

Definition 1.1 (Benjamini, Schramm and Timár [1]) Given a finite graph $\Gamma = (V\Gamma, E\Gamma)$, we shall say that a set of vertices $C \subset V\Gamma$ *cuts* (or *separates*) the graph Γ if the connected components of the subgraph induced by $V\Gamma - C$ contain at most $\frac{1}{2}|V\Gamma|$ vertices.

We define the *cut* of the graph Γ, denoted $\operatorname{cut}\Gamma$, as the minimal size of a separating set.

We define the *separation profile* of a bounded degree infinite graph G as the following nondecreasing function from $\mathbb{N}^*$ to $\mathbb{N}^*$:

$$\operatorname{sep}_G(n) = \sup_{\substack{\Gamma \subset G \\ |V\Gamma| \le n}} \operatorname{cut}\Gamma.$$

We shall consider such function endowed with the partial order defined by $g \preceq h$ if and only if there exists $D > 0$ such that $g(n) \le Dh(Dn) + D$ for any $n \ge D$. We denote by $\asymp$ and $\prec$ the associated equivalence relation and strict partial order, respectively.

As noticed in [1], the factor $\frac{1}{2}$ does not play an important role in the previous definition. Replacing it by any $\beta \in (0, 1)$ would give an equivalent profile.

The separation profile is a coarse-geometric monotone invariant (see Proposition 3.1). To our knowledge, the only such invariants that were previously defined are volume growth and asymptotic dimension; see Gromov [8]. The separation profile is a much finer invariant and has been generalized by Hume, Mackay and Tessera [10] into a spectrum of profiles called Poincaré profiles. For a survey on this topic, we refer to the first part of the thesis of the second author [11].

It is proved in [1] (see also Hume and Mackay [9]) that if a hyperbolic group has $\mathrm{sep}_G(n) \prec \log(n)$ then $\mathrm{sep}_G(n)$ is bounded and G is virtually free.

In this paper we investigate the smallest possible nonvirtually free case, namely $\mathrm{sep}_G(n) \preceq \log(n)$.

Theorem A *Let G be a hyperbolic group with $\mathrm{sep}_G(n) \preceq \log(n)$. Then G is Fuchsian or splits over finite or virtually cyclic subgroups.*

This theorem is proved in Section 2, but let us give here a sketch of proof. The first step consists in showing that the spheres of G have bounded separating sets. This is done by projecting the separating set of some suitable annulus. Then, we make these sphere separating sets converge in ∂G. This implies the existence of local cut points in ∂G, and the conclusion follows from Bowditch [3].

Corollary 1.2 *Let G be a hyperbolic group without 2-torsion. If $\mathrm{sep}_G(n) \preceq \log(n)$ then G can be inductively built from Fuchsian groups and free groups by amalgamations and HNN extensions over finite or virtually cyclic groups.*

Proof We can apply Theorem A to G. Either G is Fuchsian and we are done, or G splits over virtually cyclic groups. The edge groups are virtually cyclic, hence quasiconvex in G. This implies that the vertex groups of this splitting are quasiconvex and hence hyperbolic. By the monotonicity of the separation profile, the separation profile of the vertex groups H is $\mathrm{sep}_H(n) \preceq \mathrm{sep}_G(n) \asymp \log(n)$. Therefore, we can successively apply Theorem A to split G over virtually cyclic subgroups. Using the strong accessibility by Louder and Touikan [13] this process terminates. □

A group with conformal dimension at least one always has a separation profile bounded below by log, from [9]. Using a recent result of Carrasco and Mackay [5] giving a characterization of hyperbolic groups with conformal dimension one, we get the following corollary.

Corollary 1.3 *Let G be a one-ended hyperbolic group with no 2-torsion. If the (Ahlfors regular) conformal dimension of G is strictly greater than 1, then its separation profile is strictly greater than* log.

In this generality, to our knowledge this improves the previously known lower bounds. We do not know if this is sharp.

Lower bounds on separation profiles can be obtained from Poincaré inequalities in the boundary at infinity of hyperbolic groups, see Hume, Mackay and Tessera [10, Theorem 13]. Finding general Poincaré inequalities is an important challenge and this corollary can be seen as a step in this direction.

The following theorem shows that the converse of Theorem A (and the subsequent corollaries) is false.

Theorem B *Let S be the surface amalgam obtained by gluing two closed orientable hyperbolic surfaces along a closed filling curve in each. Then*, $\mathrm{sep}_{\pi_1 S}(n) \succ \log(n)$.

From Carrasco and Mackay [5], such a group has conformal dimension equal to 1.

From [10], a hyperbolic group with conformal dimension one always have a separation profile bounded above by any n^ϵ, with $\epsilon > 0$. To our knowledge, this is this is the first example of such a group whose separation profile is not logarithmic. This implies in particular that the conformal dimension is not attained [10, Theorem 11].

We believe that when the curves are not filling, the separation profile is actually log.

Question 1.4 Let S be a simple surface amalgam obtained by gluing two closed hyperbolic orientable surfaces along simple curves. Do we have $\mathrm{sep}_{\pi_1 S} \asymp \log$?

From Hume, Mackay and Tessera [10] study of relations between conformal dimension and separation profiles, we as well can formulate the following question:

Question 1.5 If a hyperbolic group has a separation profile bounded above by n^ϵ for every positive ϵ, does it imply that it has conformal dimension one?

Acknowledgements Lazarovich is supported by the Israel Science Foundation (grant 1562/19), and by the German–Israeli Foundation for Scientific Research and Development. Le Coz is supported by the Israel Science Foundation (grant 2919/19), the FWO and the F.R.S.–FNRS under the Excellence of Science (EOS) program (project ID 40007542).

The authors would like to thank John Mackay and Ilya Gekhtman for interesting discussions. The authors would also like to thank the referee for spotting a mistake in the first version of the paper, and for valuable corrections.

2 Proof of Theorem A

Let G be a one-ended hyperbolic group. By abuse of notation let us denote by G also the Cayley graph of G with respect to some fixed finite generating set, and assume it is δ-hyperbolic. We denote by o the identity element of G. For every $R > 0$, B_R denotes the ball, and S_R the sphere, of radius R centred at o. We denote by $A_{[R_1,R_2]}$ the annulus $B_{R_2} - B_{R_1-1}$.

Definition 2.1 For each $R > 0$, let $\pi_R : G - B_R \to S_R$ be a projection defined by $\pi_R(y) = [o, y] \cap S_R$, where $[o, y]$ is some choice of geodesic joining o and y.

For any $\alpha > 0$, we call an *α-step path* any family of vertices $v_1, \dots, v_k$ such that $d(v_i, v_{i+1}) \leq \alpha$ for any $i = 1, \dots, k-1$.

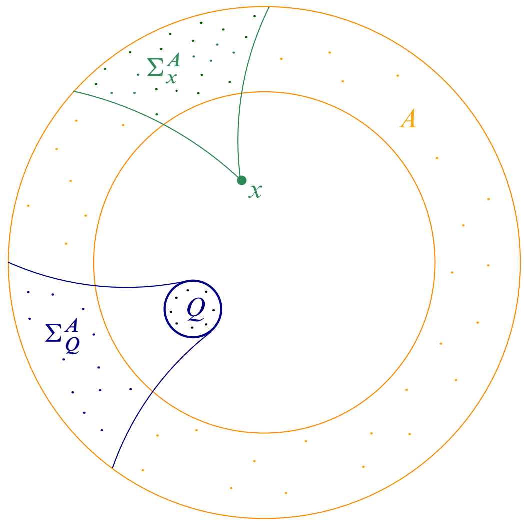


Figure 1: Shadows and sectors.

Fact 2.2 • For all $R>0$, if γ, γ' are two geodesics from o to points $x, y \in G - B_R$ then $d(\gamma(R), \gamma'(R)) \le d(x,y) + 2\delta$. In particular, $d(\pi_R(x), \pi_R(y)) \le d(x,y) + 2\delta$ for all $x, y \in G - B_R$.

- If γ, γ' are two geodesic rays from o that represent the same point at infinity then $d(\gamma(R), \gamma'(R)) \le 2\delta$.
- Since G is one-ended, there is a constant δ_1 such that the δ_1-neighbourhood of any sphere in G is connected.

Proof The first two assertions are a straightforward consequence of the δ-slimness of geodesic triangles in G. Let us prove the third assertion: Let $z_1, z_2 \in S_R$. From [2, Lemma 3.1], there is some c such that there exist infinite geodesic rays ζ_1, ζ_2 from o in G such that $d(z_i, \zeta_i(R)) \le c$ for $i = 1, 2$. Since G is one-ended, ∂G is path connected [4; 14], and so there is a continuous path from ζ_1 to ζ_2. By extending π_R to ∂G, we can "project" any continuous path in ∂G to a $(2\delta+1)$-step path in S_R. In particular the vertices $\zeta_1(R)$ and $\zeta_2(R)$ in S_R can be joined by a $(2\delta+1)$-step path in S_R, and hence the vertices z_1, z_2 in S_R can be joined by a $(c+2\delta+1)$-step path. If we set $\delta_1 = c + 2\delta + 1$, then it follows that the δ_1-neighbourhood of any sphere of G is connected. □

From now on, we will assume that δ_1 stands for the constant of Fact 2.2.

Definition 2.3 The *shadow* Σ_x of a point $x \in G$ is that set of all points $y \in G$ such that a geodesic from o to y passes through x. For every subset $Q \subseteq G$ denote its *shadow* by $\Sigma_Q = \bigcup_{x \in Q} \Sigma_x$. For a point x and $\tau \ge 0$ denote by $\Sigma_{x,\tau} = \Sigma_{B(x,\tau)}$ its τ-*shadow* where $B(x,\tau)$ is the ball of radius τ around x in G. Similarly, for $Q \subseteq G$ and $\tau \ge 0$, denote by $\Sigma_{Q,\tau} = \Sigma_{(Q)_\tau}$ its τ-*shadow*. Finally, for a subset $A \subset X$ we will denote by Σ_x^A the intersection $\Sigma_x \cap A$, and similarly $\Sigma_Q^A, \Sigma_{x,\tau}^A, \Sigma_{Q,\tau}^A$ (see Figure 1).

We will need the following strengthening of Fact 2.2.

Fact 2.4 For every τ there exists τ' such that for every $x \in G$ and $r \geq d(o,x)$, the set $\Sigma_{x,\tau} \cap S_r$ is contained in a single connected component of $\Sigma_{x,\tau'} \cap (S_r)_{\delta_1}$.

Proof For any fixed $o' \in G$, geodesic ray $\xi : [0,\infty) \to G$ starting from o', $R \gg \delta$ and $k > 2\delta$, define $V(o',\xi,R,k)$ to be the set in ∂G of all (equivalence classes of) geodesic rays ζ from o' such that $d(\zeta(R),\xi(R)) \leq k$. The sets $\{V(o',\xi,R,k)\}_{R\gg 0}$ form a neighbourhood basis for the ideal point corresponding to ξ in ∂G. By [4; 14], the Gromov boundary ∂G is locally path connected. Therefore, for every geodesic ray ξ from o' and R there exists $L = L(o',\xi,R,k)$ such that $V(o',\xi,R+L,k+4\delta)$ is in a path component of $V(o',\xi,R,k)$. Since ∂G is compact, there exists $L' = L'(o',R,k)$ such that $V(o',\xi,R+L',k+4\delta)$ is in a component of $V(o',\xi,R,k)$ for every geodesic ray ξ starting from o'. Note that by G-equivariance, L' does not depend on o'. Our next goal is to show that L' does not depend on R as well:

For a fixed $k_0 \geq 4\delta+1$ and $R_0 \gg \tau,\delta$, let $\lambda = L'(R_0,k_0-2\delta)$, then by δ-slimness,

$$V(o,\xi,R+\lambda,k_0) \subseteq V\big(\xi(R-R_0),\xi|_{[R-R_0,\infty)},R_0+\lambda,k_0+2\delta\big) \quad \text{for all } R \gg R_0.$$

By the above,

$$V\big(\xi(R-R_0),\xi|_{[R-R_0,\infty)},R_0+\lambda,k_0+2\delta\big)$$

is in a path component of $V(\xi(R-R_0),\xi,R_0,k_0-2\delta)$ which by δ-slimness is contained in $V(o,\xi,R,k_0)$. Therefore, we have shown that there exists λ, such that $V(o,\xi,R+\lambda,k_0)$ is in a path component of $V(o,\xi,R,k_0)$ for all geodesic rays $\xi \in \partial G$ and $R \gg \delta$.

Observing the trivial inclusion $V(o,\xi,R,k_0) \subseteq V(o,\xi,R+\lambda,k_0+2\lambda)$, we get that $V(o,\xi,R+\lambda,k_0)$ is in a path component of $V(o,\xi,R+\lambda,k_0+2\lambda)$ for all sufficiently large R and geodesic ray ξ. Or equivalently, $V(o,\xi,R,k_0)$ is in a path component of $V(o,\xi,R,k_0+2\lambda)$ for all sufficiently large R and geodesic ray ξ.

We proceed as in the proof of the previous fact. Set $R = d(o,x)$, and let $r \geq R$. By [2, Lemma 3.1], there exists c such that for every two points $z_1,z_2 \in \Sigma_{x,\tau} \cap S_r$ there exist two geodesics rays ζ_1,ζ_2 from o such that $d(z_i,\zeta_i(r)) \leq c$ for $i = 1,2$. By the inequality in Fact 2.2, $d(\pi_R(z_i),\zeta_i(R)) \leq c+2\delta$. Since $z_i \in \Sigma_{x,\tau}$ we get that $d(\zeta_i(R),x) \leq c+2\delta+\tau$ and hence $d(\zeta_1(R),\zeta_2(R)) \leq 2(c+2\delta+\tau)$. If we set $k_0 = 2(c+2\delta+\tau)$ we see that ζ_1,ζ_2 are in $V(o,\zeta_1,R,k_0)$. Thus, by the above, they can be connected by a continuous path in $V(o,\zeta_1,R,k_0+2\lambda)$. Projecting this path using π_r gives rise to a $(2\delta+1)$-step path in S_r between $\zeta_1(r)$ and $\zeta_2(r)$. and hence a δ_1-step path between z_1,z_2, where $\delta_1 = c+2\delta+1$ (as in Fact 2.2). Replacing each δ_1 step by a path of length δ_1 in $(S_r)_{\delta_1}$ we get a path p between z_1,z_2. Using the inequality in Fact 2.2, we see that if we further project the path p to S_R, we see that $\pi_R \circ p$ stays in the ball of radius $\tau' = k_0+2\lambda+\delta_1+2\delta$ around x. Thus, the path p is contained in $\Sigma_{x,\tau'} \cap (S_r)_{\delta_1}$. □

Fact 2.5 There exist constants $\alpha > 1$ and $\tau_0 \geq 0$ such that for all $\tau \geq \tau_0$ there exists K such that:

- For all $0 \leq R$, we have
$$K^{-1}\alpha^R \leq |S_R| \leq K\alpha^R.$$

- For all $0 \le R_1 \le R_2$, we have
$$K^{-1}(\alpha^{R_2}-\alpha^{R_1}) \le |A_{[R_1,R_2]}| \le K(\alpha^{R_2}-\alpha^{R_1}).$$
- For all $R, R' \ge 0$ and $D \subseteq S_R$, we have
$$K^{-1}\alpha^{R'}|D| \le |\Sigma_{D,\tau} \cap S_{R+R'}| \le K\alpha^{R'}|D|.$$
- For all $R \ge 0$, $R_2 \ge R_1 \ge 0$ and $D \subseteq S_R$, we have
$$K^{-1}(\alpha^{R_2}-\alpha^{R_1})|D| \le |\Sigma_{D,\tau} \cap A_{[R+R_1,R+R_2]}| \le K(\alpha^{R_2}-\alpha^{R_1})|D|.$$

Proof The first two items are immediate consequences of estimates of cardinality of balls given by Coornaert [6, théorème 7.2].

The third item is an immediate consequence of properties of the Patterson–Sullivan measure on ∂G; see Coornaert [6, Proposition 6.1]. The properties that are needed are detailed in Gouëzel, Mathéus and Maucourant [7, inequality 2.9 from the proof of Lemma 2.13, and the fact that the covering number of shadows is finite (item (1) on page 1216)].

The fourth item is obtained by summing up the inequalities given by the previous item for $R' \in [R_1, R_2]$. □

Definition 2.6 We say that G has *bounded sphere separation* if for every $\epsilon > 0$ there exists a number M such that for all R there exists a set $P_R \subseteq S_R$ such that $|P_R| \le M$, and each component of $(S_R)_{\delta_1} - (P_R)_{\delta_1}$ has size at most $\epsilon|(S_R)_{\delta_1}|$, where δ_1 is the smallest integer satisfying that the δ_1-neighbourhood of any sphere in G is connected.

Remark 2.7 According to Fact 2.2, the constant δ_1 in the definition above exists.

We are now able to state our key lemma.

Lemma 2.8 *If G is hyperbolic and $\mathrm{sep}_G(n) \preceq \log(n)$, then G has bounded sphere separation.*

Proof Let $\epsilon > 0$. Let τ_0, α be as in Fact 2.5, set $\tau = \tau_0 + 2\delta_1$ and let K be the constant of Fact 2.5. Let τ_0' and τ' be the constants of Fact 2.4 corresponding to τ_0 and τ, respectively. Without loss of generality by enlarging either τ_0' or τ', we may assume that $\tau' = \tau_0' + 2\delta$.

Let $A = A(R)$ be the annulus $A_{[2R,3R]}$. For $\beta \in (0,1)$, to be determined later, let $C \subset A(R)$ be such that each connected component of $A(R) - C$ contains at most $\beta|A(R)|$ vertices. By Fact 2.5, $|A(R)| \asymp \alpha^{3R}$, and from the assumption that $\mathrm{sep}_G \preceq \log$, we can suppose $|C| \preceq_\beta \log(|A(R)|) \preceq_\beta R$. Concretely, let

$$|C| \le cR \tag{2-1}$$

for some c (which depends on the choice of β).

Let $P_R \subset S_R$ be the set of all x such that $\Sigma_{x,\tau'} \cap (S_r)_{\delta_1} \cap C \neq \varnothing$ for all $r \in [2R+\delta_1, 3R-\delta_1]$. If $x \in P_R$, there are at least $(R-2\delta_1)/2\delta_1$ values of $r \in [2R+\delta_1, 3R-\delta_1]$ for which $(S_r)_{\delta_1}$ are disjoint, and $\Sigma^A_{x,\tau'} \cap C$ is assumed to meet all of them. Therefore we have $|\Sigma^A_{x,\tau'} \cap C| \geq (R-2\delta_1)/2\delta_1$. As long as $R \gg \delta_1$ this implies

$$|\Sigma^A_{x,\tau'} \cap C| \geq R/3\delta_1. \tag{2-2}$$

Since the τ'-shadows corresponding to vertices in S_R that are $2\tau' + 4\delta$ apart are disjoint, it follows from (2-1) and (2-2) that

$$|P_R| \leq 3c\delta_1 |B_{2\tau'+4\delta}| =: M. \tag{2-3}$$

So, there is a uniform bound M (that depends on β) on the size of P_R. It remains to show that upon choosing β small enough we can ensure that $(P_R)_{\delta_1}$ separates $(S_R)_{\delta_1}$ into components of size at most $\epsilon|(S_R)_{\delta_1}|$.

Claim 2.9 *There exists $K' > 0$ such that for every $R \gg_\beta 0$, if $x \in S_R - P_R$ then $\Sigma^A_{x,\tau} - C$ has a subset T_x of size $|T_x| \geq \frac{1}{2K'}|\Sigma^A_{x,\tau}|$ which is contained in a connected component of $A(R) - C$.*

Proof Since $x \notin P_R$, there exists $r \in [2R+\delta_1, 3R-\delta_1]$ such that $\Sigma_{x,\tau'} \cap (S_r)_{\delta_1} \cap C = \varnothing$. By Fact 2.4, $\Sigma_{x,\tau} \cap S_r$ is contained in a path component of $\Sigma_{x,\tau'} \cap (S_r)_{\delta_1}$, hence also in a path component E of $A(R) - C$. Let T_x be the intersection $E \cap \Sigma^A_{x,\tau}$.

It remains to show the lower bound on $|T_x|$. To do so, we will use τ_0-shadows of points $y \in \Sigma_{x,\tau_0} \cap S_{2R}$. Note that Σ_{y,τ_0} is contained in $\Sigma_{x,\tau}$ (since $\tau = \tau_0 + 2\delta$).

Let Q_x be the collection of all points $y \in \Sigma_{x,\tau_0} \cap S_{2R}$ such that $\Sigma^A_{y,\tau_0'} \cap C \neq \varnothing$. For each point $z \in C$ there are at most $|B_{\tau_0'+2\delta}|$ many $y \in \Sigma_{x,\tau_0} \cap S_{2R}$ such that $z \in \Sigma^A_{y,\tau_0'}$. Together with the assumption that $|C| \leq cR$ it follows that $|Q_x| \leq cR|B_{\tau_0'+2\delta}| \asymp_\beta R$. Since $|\Sigma_{x,\tau_0} \cap S_{2R}| \asymp \alpha^R$, the complementary set $Q'_x = (\Sigma_{x,\tau_0} \cap S_{2R}) - Q_x$ satisfies

$$|Q'_x| \geq \tfrac{1}{2}|\Sigma_{x,\tau} \cap S_{2R}|, \tag{2-4}$$

for any large enough $R \gg_\beta 0$.

For points $y \in Q'_x$, by Fact 2.4, Σ^A_{y,τ_0} is contained in a connected component of $\Sigma^A_{y,\tau_0'}$, and so in a component of $\Sigma^A_{x,\tau'} - C$ (since $\tau' = \tau_0' + 2\delta$) and intersects $\Sigma_{x,\tau} \cap S_r$. Thus, by the definition of T_x we have $\Sigma^A_{y,\tau_0} \subseteq T_x$ for all $y \in Q'_x$. Or equivalently, $\Sigma^A_{Q'_x,\tau_0} = \bigcup_{y \in Q'_x} \Sigma^A_{y,\tau_0} \subset T_x$. Thus, $|T_x| \geq |\Sigma^A_{Q'_x,\tau_0}|$.

Similar to Fact 2.5, we have a constant K' such that any subset $Q \subseteq \Sigma_{x,\tau_0} \cap S_{2R}$ satisfies

$$\frac{|Q|}{|\Sigma_{x,\tau} \cap S_{2R}|} \leq K' \frac{|\Sigma^A_{Q,\tau_0}|}{|\Sigma^A_{x,\tau}|}. \tag{2-5}$$

By (2-5) and (2-4) we get $|T_x| \geq \frac{1}{2K'}|\Sigma^A_{x,\tau}|$. □

Let $D' \subset (S_R)_{\delta_1} - (P_R)_{\delta_1}$ be a connected subset. We need to show $|D'| \leq \epsilon |(S_R)_{\delta_1}|$. Let D be the set of elements of S_R that are at distance at most δ_1 from D'. Define $T_D = \bigcup_{x \in D} T_x$, with T_x given by Claim 2.9.

Claim 2.10 *T_D is in a connected component of $A(R) - C$.*

Proof Since D_{δ_1} is connected, it suffices to show that for any $x, x' \in D$ at distance at most $2\delta_1$ from each other then T_x and $T_{x'}$ intersect.

Let then $x, x' \in D$ be such that $d(x, x') \leq 2\delta_1$. Then, $\Sigma_{x,\tau_0} \cap S_{2R} \subseteq \Sigma_{x,\tau} \cap \Sigma_{x',\tau} \cap S_{2R}$ and by Fact 2.5 contains $\succeq \alpha^R$ points. As in the proof of Claim 2.9 we see that if R is large enough, there exists a point $y \in \Sigma_{x,\tau} \cap \Sigma_{x',\tau} \cap S_{2R}$ which is in the complement of both Q_x and $Q_{x'}$. As before, this implies that Σ^A_{y,τ_0} is in both T_x and $T_{x'}$. □

By assumption on C, this implies that we have

(2-6) $$|T_D| \leq \beta |A(R)|.$$

Claim 2.11 *We have*

(2-7) $$\sum_{x \in D} |T_x| \leq |B_{2\tau+4\delta}| |T_D|.$$

Proof Let us show that there exists a map $\phi \colon D \to D$ such that

- $d(x, \phi(x)) \leq 2\tau$ and in particular $|\phi^{-1}(\phi(x))| \leq |B_{2\tau+4\delta}|$,
- $|T_x| \leq |T_{\phi(x)}|$, and
- if $y \neq y' \in \operatorname{Im} \phi$ then $T_y \cap T_{y'} = \varnothing$.

Assuming we have constructed such a map, the claim follows by the following inequality:

(2-8) $$\sum_{x \in D} |T_x| \leq |B_{2\tau+4\delta}| \sum_{y \in \operatorname{Im} \phi} |T_y| \leq |B_{2\tau+4\delta}| |T_D|.$$

To construct the map ϕ, let $x \in D$ be a point maximizing $|T_x|$. Let $Z \subseteq D$ be the collection of all points $x' \in D$ satisfying $T_{x'} \cap T_x \neq \varnothing$. Define ϕ on Z by $\phi(x') = x$. Note that if $d(x, x') > 2\tau + 4\delta$ then $\Sigma_{x,\tau} \cap \Sigma_{x',\tau} = \varnothing$ and hence $T_x \cap T'_x = \varnothing$. It follows that if $x' \in Z$ then $d(x, x') \leq 2\tau + 4\delta$, and by the choice of x, $|T_{x'}| \leq |T_x|$.

Remove all the points in Z from D, and iterate the construction above until ϕ is defined on all D. □

We deduce that for large enough $R \gg_\beta 0$ we have

$$\begin{aligned}
|D'| &\le |B_{\delta_1}||D| \\
&\le |B_{\delta_1}|K\alpha^{-2R}|\Sigma^A_{D,\tau}| && \text{(from Fact 2.5)} \\
&\le |B_{\delta_1}|K\alpha^{-2R}\sum_{x\in D}|\Sigma^A_{x,\tau}| \\
&\le 2K'|B_{\delta_1}|K\alpha^{-2R}\sum_{x\in D}|T_x| && \text{(from Claim 2.9)} \\
&\le 2|B_{\delta_1}||B_{2\tau+4\delta}|KK'\alpha^{-2R}|T_D| && \text{(from (2-7))} \\
&\le 2|B_{\delta_1}||B_{2\tau+4\delta}|KK'\alpha^{-2R}\beta|A(R)| && \text{(from (2-6))} \\
&\le 2|B_{\delta_1}||B_{2\tau+4\delta}|K^2K'\beta|S_R| && \text{(from Fact 2.5)} \\
&\le 2|B_{\delta_1}||B_{2\tau+4\delta}|K^2K'\beta|(S_R)_{\delta_1}|.
\end{aligned}$$

Let $w = 2|B_{\delta_1}||B_{\delta_1}||B_{2\tau+4\delta}|K^2K'$, this is a constant that depends only of G. Thus, we get

$$|D'| \le w\beta|(S_R)_{\delta_1}|. \tag{2-9}$$

For every $\epsilon > 0$, set $\beta = \frac{\epsilon}{w}$. By (2-3) there exists M such that for every R, the set $P_R \subseteq S_R$ that we constructed satisfies $|P_R| \le M$, and by (2-9) each component D' of $(S_R)_{\delta_1} - (P_R)_{\delta_1}$ has size $|D'| \le \epsilon|(S_R)_{\delta_1}|$ for large enough $R \gg_\beta 0$. We have proved the bounded sphere separation property for large enough R. This completes the proof of Lemma 2.8. □

Definition 2.12 Let X be a connected topological space. We say that a subset F *topologically separates* X if $X - F$ is not connected.

Lemma 2.13 *If G has bounded sphere separation, then ∂G has a finite topologically separating set.*

Proof Following Definition 2.6, let δ_1 is the smallest integer satisfying that the δ_1-neighbourhood of any sphere in G is connected. We start with the following claim.

Claim 2.14 *There exist $K > 0$ and $\tau_1 \ge \delta_1$ such that for any $R < R'$, if $(P_{R'})_{\delta_1}$ separates $(S_{R'})_{\delta_1}$ into connected components of size at most $\frac{1}{K}|(S_{R'})_{\delta_1}|$, then $(\pi_R(P_{R'}))_{\tau_1}$ separates $(S_R)_{\delta_1}$ into connected components of size at most $\frac{1}{2}|(S_R)_{\delta_1}|$.*

Proof For the constant τ_0 of Fact 2.5 let $\tau = \tau_0 + 2\delta_1$, and let τ' be the corresponding constant from Fact 2.4. Let $\tau_1 = \tau' + 2\delta_1$. If $x \in S_R - (\pi_R(P_{R'}))_{\tau_1}$ then $P_{R'} \cap \Sigma_{x,\tau'} = \varnothing$. It follows from Fact 2.4 that $\Sigma_{x,\tau} \cap S_{R'}$ is in a component of $\Sigma_{x,\tau'} \cap (S_{R'})_{\delta_1}$ and so in a component of $(S_{R'})_{\delta_1} - (P_{R'})_{\delta_1}$. If we have two points $d(x,y) \le 2\delta_1$ then the sets $\Sigma_{x,\tau}$ and $\Sigma_{y,\tau}$ intersect. This implies that for every connected subset D' of $(S_R)_{\delta_1} - (\pi_R(P_{R'}))_{\tau_1}$, the set $\Sigma_{D,\tau_0} \cap S_{R'}$ is connected in $(S_{R'})_{\delta_1} - (P_{R'})_{\delta_1}$, where D is the set of points in S_R at distance at most δ_1 from D'. The conclusion of claim follows from the fact that sizes of D and D' differ by some constant factor, as in Fact 2.5. □

Claim 2.15 *For each large enough R, we can choose $P_R \subset S_R$ such that*

(1) $(P_R)_{\tau_1}$ *separates* $(S_R)_{\delta_1}$ *into connected components of size at most* $\frac{1}{2}|(S_R)_{\delta_1}|$,

(2) $P_{R_1} = \pi_{R_1}(P_{R_2})$, *for every* $R_1 \leq R_2$.

Proof We can assume without any loss of generality that the projection maps are chosen so that we have $\pi_{R_1}(x) = \pi_{R_1} \circ \pi_{R_2}(x)$ for every $R_1 < R_2$ and $x \in G - B_{R_2}$.

From the assumption of bounded sphere separation, for every large enough $R' > 0$, $P_{R'} \subset S_{R'}$ of bounded size M satisfying that $(P_{R'})_{\delta_1}$ separates $(S_{R'})_{\delta_1}$ into connected components of size at most $\frac{1}{K}|(S_{R'})_{\delta_1}|$, where K is given by Claim 2.14.

From Claim 2.14, for every $R < R'$, the set $P_R := \pi_R(P_{R'})$ satisfies property (1). Now, for every $R_1 < R_2 < R'$, we have $P_{R_1} = \pi_{R_1}(P_{R_2})$ since $\pi_{R_1} = \pi_{R_1} \circ \pi_{R_2}$.

Since the spheres of G are finite, we can proceed to an extraction to obtain a sequence R'_n such that for every $R > 0$ the sequence $(\pi_R(P_{R'_n}))_{n \geq 0}$ is constant (it is only defined when $R'_n \geq R$). Without any loss of generality we can assume that we have $P_R = \pi_R(P_{R'_n})$. We finally get the desired property that $P_{R_1} = \pi_{R_1}(P_{R_2})$ for every $R_1 < R_2$. □

Now the sequence P_R has a limit $P \subseteq \partial G$ as $R \to \infty$. To complete the proof of Lemma 2.13 it remains to show that P topologically separates ∂G.

From above, there exist $\xi, \eta \in \partial G$ such that $\pi_R(\xi)$ and $\pi_R(\eta)$ are in different components of $(S_R)_{\delta_1} - (P_R)_{\delta_1}$ for all large enough R. Assume for contradiction that ξ and η are in the same component of $\partial G - P$. The boundary ∂G is path connected, let γ be a path in $\partial G - P$ connecting ξ and η. There exists $\epsilon > 0$ such that the path γ avoids the ϵ-neighbourhood of P in ∂G. Let R be big enough so that $\pi_R^{-1}(x_{2\delta_1}) \cap \partial G$ is of diameter $\leq \epsilon/2$ for each $x \in S_R$.

Thus, $\pi_R \circ \gamma$ is a $2\delta_1$-step path in S_R which avoids $(P_R)_{2\delta_1}$. Completing it with a collections of geodesic arcs of length at most δ_1, we get a path in $(S_R)_{\delta_1}$ which avoids $(P_R)_{\delta_1}$, and connects $\pi_R(\xi)$ and $\pi_R(\eta)$. A contradiction. This ends the proof of Lemma 2.13. □

Proof of Theorem A By Lemmas 2.8 and 2.13 we see that ∂G is topologically separated by a finite set of points. It therefore has a local cut point. It follows from Bowditch [3] that G splits over a virtually cyclic group or G is a Fuchsian group. □

3 Proof of Theorem B

In this section, we construct a hyperbolic group with superlogarithmic separation profile whose boundary has conformal dimension one. Let us start by giving the following proposition.

Proposition 3.1 [1, Lemma 1.3] *Let G and H be bounded degree infinite graphs such that there exists a coarse embedding $G \to H$. Then,* $\mathrm{sep}_G \preceq \mathrm{sep}_H$.

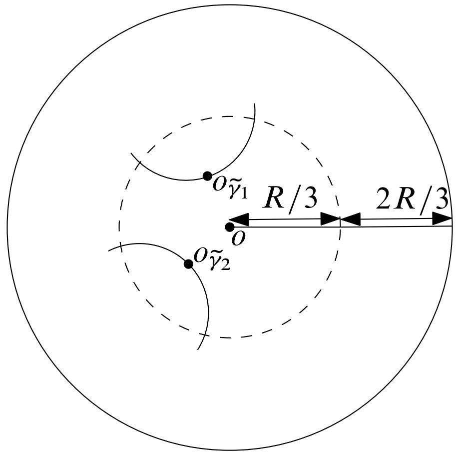


Figure 2: The set X of Proposition 3.2.

Recall that quasi-isometric embeddings are examples of coarse embeddings. This proposition implies that the separation profile is invariant under coarse equivalences and quasi-isometries. In particular we can consider separation profiles more generally for metric spaces that are coarsely equivalent to graphs of bounded degree. This is what we will do in this section for the hyperbolic plane.

Let Σ and Σ' be two closed hyperbolic orientable surfaces, and $\gamma \subset \Sigma,\ \gamma' \subset \Sigma'$ be two closed filling geodesic curves. Recall that a curve on a surface is said to be *filling* when its complementary is homeomorphic to a union of disks. Let $S = (\Sigma \sqcup \Sigma')/\gamma \simeq \gamma'$ be the space obtained by gluing Σ and Σ' along γ and γ'.

The universal cover $\widetilde{S}$ of S consists of copies of hyperbolic planes, that we will call sheets, glued together along the geodesic lines which correspond to the lifts of γ and γ'.

Let F be one of the sheets covering Σ. For a lift $\widetilde{\gamma} \subset F$ of γ let $F_{\widetilde{\gamma}}$ be the adjacent sheet covering Σ' which is glued to F along $\widetilde{\gamma}$.

Let $R > 0$. Let B_R (resp. $B_{R/3}$) be the balls of radius R (resp. $R/3$) in F centred around o. Let us consider

$$X = B_R \cup \bigcup_{\widetilde{\gamma} \cap B_{R/3} \neq \varnothing} B_{F_{\widetilde{\gamma}}}(o_{\widetilde{\gamma}}, R/3) \subset \widetilde{S}$$

where the union ranges over all lifts $\widetilde{\gamma}$ of γ that intersect $B_{R/3}$ and $B_{F_{\widetilde{\gamma}}}(o_{\widetilde{\gamma}}, R/3)$ is the ball of radius $R/3$ in the sheet $F_{\widetilde{\gamma}}$ centred at the point $o_{\widetilde{\gamma}}$ on $\widetilde{\gamma}$ which is closest to o. See Figure 2.

Proposition 3.2 *The set X satisfies* $\operatorname{cut} X \succ R$.

Let us prove how Theorem B can be deduced from this proposition.

Proof of Theorem B The fundamental group $\pi_1 S$ is quasi-isometric to the universal cover $\widetilde{S}$. Thus, we can compute the separation profile of $\widetilde{S}$ instead of that of $\pi_1 S$, and the theorem follows from Proposition 3.2. □

Proof of Proposition 3.2

Claim 3.3 *Most of the volume of X lies in the ball $B_R \subset F$:*

$$\mathrm{vol}(X) \asymp \mathrm{vol}(B_R).$$

Proof The volume of a ball $B(o,r)$ of radius r in the hyperbolic plane is

$$\mathrm{vol}(B(o,r)) = 2\pi(\cosh(r) - 1) \asymp e^r.$$

The number of lifts of a geodesic that intersect a ball $B(o,r)$ is $\preceq e^r$. Thus,

$$\mathrm{vol}(B(o,R)) \preceq \mathrm{vol}(X) \preceq \mathrm{vol}(B(o,R)) + e^{R/3}\,\mathrm{vol}(B(o,R/3)) \preceq \mathrm{vol}(B(o,R)).$$ □

Let C be a (1-thick) cutset of X, that is every connected component of $X - C$ has volume at most $\alpha \operatorname{vol} X$ for some $\alpha < 1$. Up to taking a small enough α, C has to separate the ball B_R. We want to show that C must have volume strictly bigger than $\log \mathrm{vol}(X) \asymp \log \mathrm{vol}(B(o,R)) \asymp R$. Let us assume for a contradiction that we have (up to constants), $\operatorname{vol} C \preceq R$.

Let $\Lambda = \partial C$. The total length of Λ is $O(R)$: indeed, C has volume R and can be chosen to be a union of balls of radius 1 in a given net and so the length of their boundary component has to be $O(R)$.

The components of Λ are either proper arcs or simple closed curves in B_R. Let $\widehat{\Lambda}$ be the collection of geodesics with the same endpoints as the arcs of Λ. Let C' be the union of the components of $X - \widehat{\Lambda}$ that include $C \cap \partial B_R$. See Figure 3.

Lemma 3.4 (i) $\mathrm{vol}(C \triangle C') \preceq R$.

(ii) *Every component E of $B_R - C$ of size $\succ R$ corresponds to a unique component E' of $B_R - C'$ such that $\mathrm{vol}(E \triangle E') \preceq R$, and vice versa.*

Proof (i) The difference between the sets C and C' has boundary in $\Lambda \cup \widehat{\Lambda}$. Then, since the total length of Λ and $\widehat{\Lambda}$ is $O(R)$, it follows from the isoperimetric inequality on the hyperbolic plane, that $\mathrm{vol}(C \triangle C') \preceq \mathrm{length}(\Lambda \cup \widehat{\Lambda}) \preceq R$.

(ii) By the isoperimetric inequality, a component E of $B_R - C$ of size $\succ R$ must intersect ∂B_R. There is a component E' of $B_R - C'$ with $E \cap \partial B_R = E' \cap \partial B_R$. The difference $E \triangle E'$ comprises of sets which are bounded by Λ and $\widehat{\Lambda}$. The volume of this difference can be bounded by $\preceq R$ again by the isoperimetric inequality. □

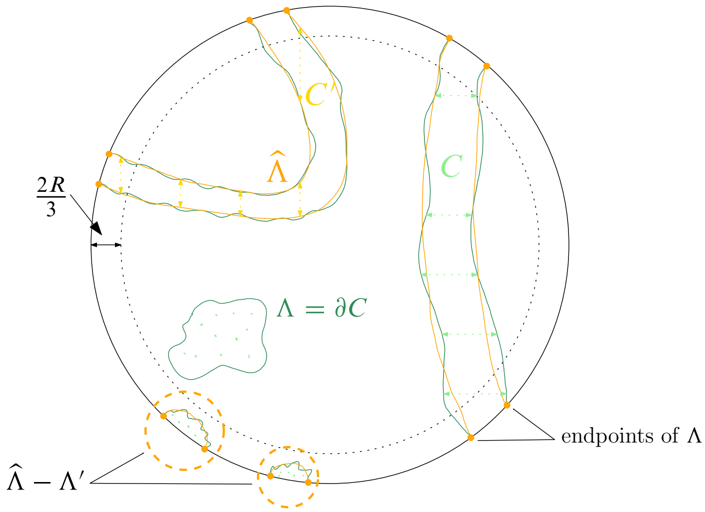


Figure 3: The separating set C of the hyperbolic ball.

Let Λ' be a set of geodesics in $\widehat{\Lambda}$ that meet $B_{R/3} = B_F(o, R/3)$. Any geodesic in Λ' must have a segment joining $\partial B_{R/3}$ and ∂B_R, and thus must have length at least $2R/3$. Since length$(\widehat{\Lambda}) \preceq R$ there are $O(1)$ many geodesics in Λ'. Let k be the number of geodesics in Λ'.

Let $m(x, y, z)$ denote the centre of the geodesic triangle spanned by a triple points $x, y, z \in \mathbb{H}^2$. Let

$$M = \{m(o, x, y) \mid x, y \in \Lambda' \cap \partial B_R\}.$$

There are at most $(2k)^2$ points in M.

Divide $B_{R/3}$ into $3((2k)^2 + 1)$ radial annuli of same width, called *layers*. By the pigeonhole principle, there exist three consecutive layers A_-, A, A_+ that do not contain a point of M.

Claim 3.5 *For R large enough, we have:*

(i) *The intersection $\Lambda' \cap A$ consists of geodesics joining the inner and outer boundaries of A.*

(ii) *If α is a component of $\Lambda' \cap A$, and α is a subarc of $\lambda \in \Lambda'$, then α is at Hausdorff distance at most δ from the arc of intersection of one of the two geodesics connecting o and $\partial\lambda$ with A.*

(iii) *If α_1, α_2 are components of $\Lambda' \cap A$, then either the Hausdorff distance $d_H(\alpha_1, \alpha_2) \leq 3\delta$ or they are at distance $\sqrt{R}$ apart.*[1]

Proof (i) Otherwise, a component α of $\Lambda' \cap A$ is a geodesic arc connecting the outer boundary of A to itself. Let p be the point on α closest to o, let λ be the geodesic of Λ' to which α belongs, and let x, y be the endpoints of λ in $S_F(o, R)$. Then, $m(o, x, y) \in M$ is at distance δ from p, contradicting the assumption that $A_- \cup A \cup A_+$ does not include points of M.

[1]The function $\sqrt{R}$ can be replaced by any function $o(R)$.

(ii) Consider the geodesic triangle consisting of the geodesic λ and the two geodesics connecting its endpoints to o. By assumption, the centre of this geodesic is not in $A_- \cup A \cup A_+$, hence by slimness of triangles in $\mathbb{H}^2$ the segment α is δ-close to one of the sides.

(iii) Let α_1, α_2 be components of $\Lambda' \cap A$. Let λ_1 (resp. λ_2) be the geodesic in Λ containing α_1 (resp. α_2). By (ii), α_1 (resp. α_2) is δ-close to a radial geodesic λ_1' (resp. λ_2') connecting o and one of the endpoints of λ_1 (resp. λ_2). Let $\alpha_1' = \lambda_1' \cap A$ (resp. $\alpha_2' = \lambda_2' \cap A$). It suffices to prove that $d_H(\alpha_1', \alpha_2') \le \delta$, or they are at distance $\sqrt{R} + 2\delta$ apart.

Let $p_1, q_1 \in \alpha_1'$ (resp. $p_2, q_2 \in \alpha_2'$) be the intersection of α_1' (resp. α_2') with the inner and outer boundaries of A, respectively. If $d(q_1, q_2) \le \delta$ then $d_H(\alpha_1', \alpha_2') \le \delta$ by the convexity of the metric. Similarly, if $d(p_1, p_2) \ge \sqrt{R} + 2\delta$ then α_1', α_2' are at least $\sqrt{R} + 2\delta$ apart. Otherwise, $d(p_1, p_2) \le \sqrt{R} + 2\delta$ and $d(q_1, q_2) > \delta$, then the centre of the triangle with sides λ_1', λ_2' is at distance at most $\sqrt{R} + 3\delta$ from α_1'. For R large enough, such a point must be in $A_- \cup A \cup A_+$ in contradiction to the assumption. □

From the claim above it follows that $\Lambda' \cap A$ consists of at most $2k$ geodesic segments connecting the inner and outer boundaries of A and the relation defined by $\alpha_1 \sim \alpha_2$ if $d_H(\alpha_1, \alpha_2) \le 3\delta$ is an equivalence relation. Let $\mathcal{W}$ be a set of representatives of the classes of this relation. We call the elements in $\mathcal{W}$ *walls*. We call the connected components of $A - \mathcal{W}$ *regions*. By the claim, the walls bounding each region are at distance $\sqrt{R}$ apart.

Claim 3.6 *Let D be a region in A, then there exists a (unique) component E of $X - C$ such that* $\mathrm{vol}(D - E) \preceq R$.

Proof Let α_1, α_2 be the walls bounding D. Let α_1', α_2' be the inner most arcs in D which belong to the equivalence classes of α_1, α_2, respectively. Let E' be the connected component of $X - \widehat{\Lambda}$ which includes the section of A between α_1', α_2'. This section is contained in D, and $D - E'$ consists of two regions which are contained in the 3δ-neighbourhood of $\alpha_1 \cup \alpha_2$. Therefore, $\mathrm{vol}(D - E') \preceq R$. The set C' is bounded by geodesics in $\widehat{\Lambda}$. It cannot contain the component E', as otherwise $\mathrm{vol}(C') \succeq \mathrm{vol}(E') \succ R$. Thus E' is a component of $X - C'$. By Lemma 3.4, E' corresponds to a unique component E of $X - C$, and $\mathrm{vol}(D - E) \le \mathrm{vol}(D - E') + \mathrm{vol}(E' - E) \preceq R$. □

Claim 3.7 *No component E of $X - C$ contains more than $\frac{2}{3}$ of the layer A.*

Proof Let E be a component of $X - C$. Assume for contradiction that $\mathrm{vol}(E \cap A) > \frac{2}{3}\mathrm{vol}(A)$. For every $x \in E \cap A$ consider the ray $x^* = \mathbb{R}_{\ge 1} x \cap B_R = \{tx \in B_R \mid t \ge 1\}$. Let $E_1 = \{x \in E \mid x^* \cap C = \varnothing\}$. Since $\mathrm{vol}(C)$ consists of $O(R)$ balls of radius 1, the set $E - E_1$ consists of at most $O(R)$ 1-neighbourhoods of arcs of length $O(R)$. Whence, $\mathrm{vol}(E - E_1) \preceq R^2$. Consider the set $E_1^* = \bigcup_{x \in E_1} x^*$. Thus $\mathrm{vol}(E_1) > \frac{1}{2}\mathrm{vol}(A)$, and therefore also $\mathrm{vol}(E_1^*) > \frac{1}{2}\mathrm{vol}(B_R)$. The set $E_1^* \subseteq E$, thus $\mathrm{vol}(E) > \frac{1}{2}\mathrm{vol}(B_R)$. We get a contradiction to the assumption that the volume of components of $X - C$ are at most $\frac{1}{2}\mathrm{vol}(B_R)$. □

By the previous two claims there are two regions D_+, D_- of $A - \mathcal{W}$ which correspond to two different components E_+, E_- of $X - C$. We may assume that D_1 and D_2 are adjacent, and are separated by a wall α. Let α_m be the middle third subarc of α.

Claim 3.8 *There is $k \asymp R$, and disjoint lifts $\widetilde{\gamma}_1, \dots, \widetilde{\gamma}_k$ of γ such that $\widetilde{\gamma}_i \cap \alpha_m \neq \varnothing$.*

Proof Consider the union $\Gamma = \bigcup \widetilde{\gamma}$ of all the lifts $\widetilde{\gamma}$ of γ to the universal cover F of Σ. Since γ is filling in Σ, the connected components of $F - \Gamma$ are one of finitely many types of convex hyperbolic nonideal polygons. Let d be the maximal diameter of these polygons. There exists an angle θ such that every geodesic line intersecting one of the polygons, forms an angle θ with at least one of its sides. Let $\mu > 0$ be such that if two geodesic lines l_1, l_2 in the hyperbolic plane intersect a third geodesic line l at points of distance $\geq \mu$ and at angles $\geq \theta$, then l_1, l_2 do not meet.

Let $\ell = \mathrm{length}(\alpha_m) \asymp R$. Every segment of length $2d$ on α_m intersects a lift $\widetilde{\gamma}$ of γ in an angle $\geq \theta$. Thus, the geodesic segment α_m intersects at least $k = \ell/(\mu + 2d)$ lifts $\widetilde{\gamma}_1, \dots, \widetilde{\gamma}_k$ of γ in an angle $\geq \theta$. Note that $k \asymp \ell \asymp R$. By the choice of μ, $\widetilde{\gamma}_1, \dots, \widetilde{\gamma}_k$ are disjoint. □

Let $\widetilde{\gamma}_1, \dots, \widetilde{\gamma}_k$ be the disjoint lifts as in Claim 3.8. The geodesic segments $\widetilde{\gamma}_i \cap D_\pm$ have length at least $\succeq \sqrt{R}$ by Claim 3.5 and by the choice of α_m. The circles around the point $\widetilde{\gamma}_i \cap \alpha_m$ in $F_{\widetilde{\gamma}_i}$ form $\Theta(\sqrt{R})$ disjoint paths connecting points in D_+ to points in D_-. Considering these paths for all $\widetilde{\gamma}_i$, we get $\Theta(R^{3/2})$ disjoint paths connecting D_+ to D_-. By Claim 3.6, $\mathrm{vol}(D_\pm - E_\pm) \preceq R$, and so we have at least $\succeq R^{3/2} - O(R)$ disjoint paths connecting E_+ to E_-. Since E_+ and E_- are different components of $X - C$, each of these paths meets C. We get $\mathrm{vol}(C) \succeq R^{3/2}$ which contradicts $\mathrm{vol}(C) \preceq R$. This ends the proof of Proposition 3.2. □

References

[1] **I Benjamini**, **O Schramm**, **Á Timár**, *On the separation profile of infinite graphs*, Groups Geom. Dyn. 6 (2012) 639–658 MR Zbl

[2] **M Bestvina**, **G Mess**, *The boundary of negatively curved groups*, J. Amer. Math. Soc. 4 (1991) 469–481 MR Zbl

[3] **B H Bowditch**, *Cut points and canonical splittings of hyperbolic groups*, Acta Math. 180 (1998) 145–186 MR Zbl

[4] **B H Bowditch**, *Connectedness properties of limit sets*, Trans. Amer. Math. Soc. 351 (1999) 3673–3686 MR Zbl

[5] **M Carrasco**, **J M Mackay**, *Conformal dimension of hyperbolic groups that split over elementary subgroups*, Invent. Math. 227 (2022) 795–854 MR Zbl

[6] **M Coornaert**, *Mesures de Patterson–Sullivan sur le bord d'un espace hyperbolique au sens de Gromov*, Pacific J. Math. 159 (1993) 241–270 MR Zbl

[7] **S Gouëzel**, **F Mathéus**, **F Maucourant**, *Entropy and drift in word hyperbolic groups*, Invent. Math. 211 (2018) 1201–1255 MR Zbl

[8] **M Gromov**, *Geometric group theory, II: Asymptotic invariants of infinite groups*, Lond. Math. Soc. Lect. Note Ser. 182, Cambridge Univ. Press (1993) MR Zbl

[9] **D Hume**, **J M Mackay**, *Poorly connected groups*, Proc. Amer. Math. Soc. 148 (2020) 4653–4664 MR Zbl

[10] **D Hume**, **J M Mackay**, **R Tessera**, *Poincaré profiles of groups and spaces*, Rev. Mat. Iberoam. 36 (2020) 1835–1886 MR Zbl

[11] **C Le Coz**, *Separation and Poincaré profiles*, PhD thesis, Université Paris-Saclay (2020) Available at `http://www.theses.fr/2020UPASM014`

[12] **R J Lipton**, **R E Tarjan**, *A separator theorem for planar graphs*, SIAM J. Appl. Math. 36 (1979) 177–189 MR Zbl

[13] **L Louder**, **N Touikan**, *Strong accessibility for finitely presented groups*, Geom. Topol. 21 (2017) 1805–1835 MR Zbl

[14] **G A Swarup**, *On the cut point conjecture*, Electron. Res. Announc. Amer. Math. Soc. 2 (1996) 98–100 MR Zbl

NL: *Department of Mathematics, Technion – Israel Institute of Technology*
Haifa, Israel

CLC: *Department of Mathematics: Algebra and Geometry (WE01), Ghent University*
Ghent, Belgium

Current address: *Oppida*
Montigny-le-Bretonneux, France

`lazarovich@technion.ac.il`, `corentinlecoz@outlook.com`

`https://lazarovich.net.technion.ac.il/`

Received: 19 November 2021 Revised: 28 February 2023

Algebraic & Geometric Topology **25**:1 (2025) 55–106
DOI: 10.2140/agt.2025.25.55
Published: 5 March 2025

Topology and geometry of flagness and beltness of simple handlebodies

ZHI LÜ
LISU WU

We consider a class of right-angled Coxeter orbifolds, called simple handlebodies, which are a generalization of right-angled Coxeter simple polytopes. We generalize the notions of flag and belt in the setting of simple polytopes into the setting of simple handlebodies, and prove the following two topological properties characterized in terms of combinatorics: a simple handlebody is orbifold-aspherical if and only if it is flag; and the orbifold fundamental group of a simple handlebody contains a rank-two free abelian subgroup if and only if this simple handlebody contains an $\square$-belt. Furthermore, together with some results of geometry, it is shown that the existence of some curvatures on manifold double over a simple handlebody can be also characterized in terms of combinatorics.

57R18

1 Introduction

A polytope is called *simple* if its each codimension-k face is the intersection of exact k *facets* (i.e. codimension-one faces). Simple polytopes give rise to many interesting and beautiful connections among topology, geometry, combinatorics and so on.

The story originated from Pogorelov and Andreev. Pogorelov's theorem implies that a 3-dimensional simple polytope (except for tetrahedra) can be embedded into hyperbolic space $\mathbb{H}^3$ with right dihedral angles if and only if the polytope satisfies certain combinatorial conditions (i.e. containing no prismatic 3-circuit and prismatic 4-circuit); see Pogorelov [52]. Furthermore, Andreev's theorem gives a complete characterization of 3-dimensional compact hyperbolic (simple) polytopes having nonobtuse dihedral angles in term of pure combinatorial conditions; see Andreev [2] and Roeder, Hubbard and Dunbar [53].

The theorems of Pogorelov and Andreev play an important role in Thurston's hyperbolization theorem for Haken 3-manifolds [58]. An orbifold version of the hyperbolization theorem was given by Otal [49]. The hyperbolization theorem implies that the main obstructions to hyperbolic structure on closed Haken 3-manifolds (or 3-orbifolds) are asphericity (i.e. π_k trivial for each $k \geq 2$) and atoroidalness (i.e. π_1 contains no subgroup $\mathbb{Z} \oplus \mathbb{Z}$) which is related to the combinatorics of simple polytopes.

The combinatorial characterizations of asphericity and atoroidalness for cubical complexes were given by Gromov [32] and have played fundamental roles in geometric group theory.

(a) A piecewise Euclidean cubical complex is locally CAT(0) (hence, aspherical) if and only if the link of its each vertex is flag (i.e. contains no Δ).

(b) A piecewise Euclidean cubical complex is locally CAT(-1) (hence, atoroidal) if and only if the link of its each vertex is flag and contains no $\square$.

It follows that a right-angled Coxeter group is always CAT(0), and it is atoroidal if and only if it satisfies no $\square$ condition. These results have been strengthened by Moussong to conclude that every Coxeter group is CAT(0), and it is CAT(-1) if and only if it does not contain a copy of $\mathbb{Z} \oplus \mathbb{Z}$; see Moussong [46] or Davis [18, Corollary 12.6.3].

It's worth noting that each n-dimensional simple polytope admits a natural right-angled Coxeter orbifold structure, that is locally modeled on the quotient $\mathbb{R}^n/(\mathbb{Z}_2)^k$ of the $(\mathbb{Z}_2)^k$-reflective action on $\mathbb{R}^n$. A combinatorial characterization of orbifold-asphericity for right-angled Coxeter simple polytopes is derived from a theorem of Davis, Januszkiewicz and Scott [24]. That is, a right-angled Coxeter simple polytope is orbifold-aspherical if and only if the dual of its boundary is a flag complex. On the other hand, the orbifold fundamental group of a right-angled Coxeter simple polytope is a right-angled Coxeter group. Thus, a combinatorial characterization of atoroidalness for right-angled Coxeter simple polytopes is implied in Gromov's result. Right-angled Coxeter simple polytopes (and more general Coxeter simple polytopes) have played an important role in geometric group theory and in constructing high-dimensional hyperbolic manifolds (e.g. see Davis [18], Everitt, Ratcliffe and Tschantz [26], Garrison and Scott [30], and Gromov [32]).

Beyond that, simple polytopes have played an important role in theory of toric varieties, toric geometry, toric topology, etc (e.g. see Barreto, López de Medrano and Verjovsky [5], Buchstaber, Erokhovets, Masuda, Panov and Pak [8; 9], Danilov [15; 16], Davis and Januszkiewicz [22; 23], Fulton [29], Kuroki, Masuda and Yu [38], Lü and Masuda [42], Notbohm [48], and Wu and Yu [61]). There are also various relative works with other topics and viewpoints of simple polytopes (e.g. see Bahri, Bendersky, Cohen and Gitler [3], Bosio and Meersseman [6], Cao and Lü [10], Chen, Lü and Yu [11], Choi and Park [14], Gitler and López de Medrano [31], and Lü and Tan [43]).

In this paper, we consider the combinatorial characterizations of orbifold-asphericity and atoroidalness for a *simple handlebody*, which can be obtained from a right-angled Coxeter simple n-polytope by gluing some specific disjoint facets. Specifically, each simple handlebody Q with dimension $n \geq 3$ satisfies the following conditions.

(a) As an orbifold, the underlying space $|Q|$ is an n-dimensional handlebody of genus $\mathfrak{g} \geq 0$ that is a tubular neighborhood of the wedge sum of $\mathfrak{g}$ circles in $\mathbb{R}^n$ (of course, an n-dimensional handlebody of genus 0 is exactly an n-ball).

(b) The *nerve* of Q, denoted by $\mathcal{N}(Q)$, is a triangulation of the boundary $\partial|Q|$, where $\mathcal{N}(Q)$ is the abstract simplicial complex with a vertex for each *facet* (i.e. codimension-one face) of Q and a $(k-1)$-simplex for each nonempty k-fold intersection.

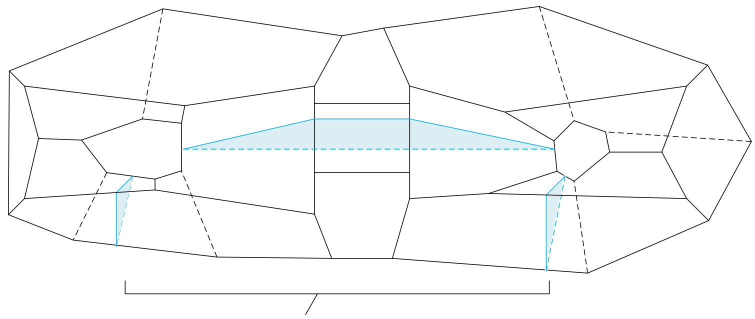


Figure 1: A simple 3-handlebody of genus 2.

(c) Each facet in Q is a simple polytope.

(d) Q can be cut into a simple polytope P_Q along some codimension-one B-belts (called *cutting belts*; for the notion of B-belts, see Definition 3.1).

Conditions (c) and (d) are restrictive conditions of the proof, and will be automatically omitted for $n = 3$; see Lemma 7.14. An example of simple 3-handlebody is shown in Figure 1.

We shall carry out our work from the following aspects:

(I) We generalize the notions of belt and flag in the setting of simple polytopes into the setting of simple handlebodies (see Definitions 3.1 and 3.3). Indeed, there is quite a difference because the underlying space of a simple handlebody is not contractible. As we shall see, the flagness of a simple handlebody Q cannot be defined by the flagness of its nerve $\mathcal{N}(Q)$ in general. Actually, its definition is given in such a different way that Q contains no Δ^k-belt for any $k \geq 2$. Here a Δ^k-belt of Q is an essential embedding suborbifold given by $(\mathbb{Z}_2)^{k+1}$-torus action on S^k.

(II) To understand the implicit structure of a simple handlebody, we introduce the notions of "right-angled Coxeter cells" and "right-angled Coxeter cellular complexes". Then we see that for a right-angled Coxeter cellular complex X, its orbifold fundamental group $\pi_1^{\mathrm{orb}}(X)$ is isomorphic to the orbifold fundamental group of its 2-skeleton (see Proposition 2.9). For a simple handlebody Q, we can give an explicit right-angled Coxeter cellular decomposition of Q, so that we can obtain an explicit presentation of $\pi_1^{\mathrm{orb}}(Q)$, which is just an iterative HNN-extension over some right-angled Coxeter group. Indeed, generally $\pi_1^{\mathrm{orb}}(Q)$ will not be the right-angled Coxeter group of Q, given by only reflections on facets of Q, and it actually contains torsion-free generators.

(III) We will use a "basic construction" of Davis [18, Chapter 5], which plays an important role on our work. This basic construction tells us that each simple n-handlebody Q can be finitely covered by a closed n-manifold M_Q with an action of some 2-torus group G, which is called a *manifold double* over Q. By the theory of orbifold covering,

$$\pi_k^{\mathrm{orb}}(Q) \cong \pi_k(M_Q), \quad k \geq 2.$$

Thus, Q is orbifold-aspherical if and only if M_Q is aspherical. It follows from [24, Theorem 2.2.5] that a simple polytope P is orbifold-aspherical if and only if P is flag; so in general, a simple n-handlebody Q may not be orbifold-aspherical although $|Q|$ is aspherical. In addition, using the basic construction, we can also use $\pi_1^{\mathrm{orb}}(Q)$ and P_Q to construct the orbifold universal cover $\widetilde{Q}$ of Q.

(IV) Based upon (I), (II) and (III), together with the Cartan–Hadamard theorem and the work of Gromov [32, Section 4.2] on nonpositive curvature, we obtain a combinatorial characterization of orbifold-asphericity of simple handlebodies (see Theorem A). Making use of Tits' theorem of Coxeter groups [18, Theorem 3.4.2] and the normal form theorem of HNN-extensions [44, Theorem 2.1, page 182], we also obtain a combinatorial characterization of atoroidalness of simple handlebodies (see Theorem B).

Now let us state our main results as follows.

Theorem A *Let Q be a simple handlebody of dimension $n \geq 3$. Then Q is orbifold-aspherical if and only if it is flag, that is, Q contains no Δ^k-belt for any $k \geq 2$.*

Remark 1.1 Theorem A is a "combinatorial sphere theorem" of simple handlebodies, which can be viewed as a generalization of Davis, Januszkiewicz and Scott [24, Theorem 2.2.5] for simple polytopes. Theorem A tells us that if Q is not flag, then there must exist a Δ^k-belt for some $k \geq 2$ in Q, so that the pullback of the embedding $\Delta^k \hookrightarrow Q$ via the projection $M_Q \to Q$ gives an (equivariant) embedding $S^k \hookrightarrow M_Q$ which represents a nontrivial element in $\pi_k(M_Q)$, as shown in the following diagram:

$$\begin{tikzcd} S^k \arrow[r, hook] \arrow[d] & M_Q \arrow[d] \\ \Delta^k \arrow[r, hook] & Q \end{tikzcd}$$

Our other main result characterizes the rank two free abelian subgroup $\mathbb{Z} \oplus \mathbb{Z}$ in $\pi_1^{\mathrm{orb}}(Q)$ in terms of combinatorics of Q.

Theorem B *Let Q be a simple handlebody of dimension $n \geq 3$. Then there is a rank two free abelian subgroup $\mathbb{Z} \oplus \mathbb{Z}$ in $\pi_1^{\mathrm{orb}}(Q)$ if and only if Q contains a $\square$-belt.*

Remark 1.2 Theorem B is a "combinatorial flat torus theorem" of simple handlebodies. Similar to the pullback way in Remark 1.1, the existence of a $\square$-belt in Q actually means that there exists an essential embedding of 2-dimensional torus T^2 in M_Q, which is an obstacle of the existence of hyperbolic structure or negative curvature on M_Q. For the flat torus theorem of nonpositively curved spaces, one can refer to Bridson and Haefliger [7, Charter 7, Part II] or Lawson and Yau [39].

Next, as further consequences of our two main results, we discuss the topology and geometry of covering spaces over a simple handlebody. Together with some important results in geometry from Davis [18, Proposition I.6.8], Kapovich [37], Kuroki, Masuda and Yu [38, Theorem 1.2], Otal [49, Chapter 7], and

Wu and Yu [61, Proposition 4.9], for a simple handlebody Q, we obtain some relations between the existence of some curvatures on M_Q and the combinatorics of Q.

Corollary 1.3 *Let Q be a simple handlebody of dimension $n \geq 2$, M_Q be the smooth manifold double over Q, and $\widetilde{Q}$ be the orbifold universal cover of Q. Then*:

(i) *The following statements are equivalent*:

  (1) *M_Q is nonpositively curved*;

  (2) *$\widetilde{Q}$ is CAT*(0);

  (3) *Q is flag* (*this is equivalent to Q being orbifold-aspherical*);

  (4) *M_Q is aspherical.*

(ii) *If M_Q admits a strictly negative curvature then Q is flag and contains no $\square$-belt. In particular*, *if Q is a simple polytope P*, *then M_P admits a strictly negative curvature if and only if P is flag and contains no $\square$-belt.*

In the 3*-dimensional case*,

(iii) *M_Q is hyperbolic if and only if it is flag and contains no $\square$-belt. Moreover*, *Q admits a right-angled hyperbolic structure if and only if it is flag and contains no $\square$-belt. In this case*, $\widetilde{Q} \approx \mathbb{H}^3$.

(iv) *When Q is a simple* 3*-polytope*, *M_Q admits a positive scalar curvature if and only if every* 2*-dimensional belt in Q is Δ^2*, *or Q is just a tetrahedron.*

Remark 1.4 The proof of Corollary 1.3 will mainly be finished in Section 7.

- Corollary 1.3(i)–(ii) are based on Gromov's results; see [32]. A metric space is said to be of curvature k if it is locally a CAT(k) space. A comparison theorem in [7, Theorem 1A.6, page 173] tells us that a smooth Riemannian manifold has curvature $\leq k$ if and only if it has sectional curvature $\leq k$. Hence, under the condition that M_Q admits a smooth Riemannian metric, the curvature in the statements of Corollary 1.3(i)–(ii) can be replaced by sectional curvature. See [38, Theorem 1.2] for simple polytopes.

- There are examples of closed orientable 3-manifolds that are aspherical but do not support a Riemannian metric with nonpositive sectional curvature (see Leeb [40]).

- Corollary 1.3(iii) is the hyperbolization theorem on simple 3-handlebodies. The hyperbolization theorem on general right-angled Coxeter 3-orbifolds was considered by Otal [49]. An *irreducible and atoroidal* 3-manifold Q with corners, defined by Otal [49, page 168], implies essentially that all involved Δ^2 and $\square$ suborbifolds in Q are not belts. This is actually equivalent to saying that Q is flag and contains no $\square$-belt. Here our statement is more combinatorial.

- Corollary 1.3(iv) is also a restatement of a result of [61]. A vc(k) in [61] is equivalent to the simple 3-polytope in Corollary 1.3(iv).

- All 2-dimensional right-angled Coxeter orbifolds can be classified by their orbifold Euler numbers; see Thurston [58].

• With a bit of additional argument, the "simple" condition in 3-dimensional case can be generalized to the case of a right-angled Coxeter 3-handlebody whose nerve is an ideal triangulation of its boundary, where the concept of ideal triangulation can be found in Fomin, Shapiro and Thurston [27, Section 2]. In this case, there may exist bad 3-handlebodies, that is, as right-angled Coxeter orbifolds, they cannot be covered by 3-manifolds. So these bad orbifolds cannot admit any hyperbolic metric. See Lemma 7.16. Although so, we can obtain that a right-angled Coxeter 3-handlebody with ideal nerve is hyperbolic if and only if it is very good, flag and contains no $\square$-belt; see Corollary 7.18. More generally, if the nerve of Q is not a triangulation or an ideal triangulation of $\partial|Q|$, then only the flag condition and no $\square$-belt condition can prevent the (right-angled) hyperbolicity of Q. An example is given in Example 7.20.

Structure of the paper

In Section 2, we review the notions of (right-angled Coxeter) orbifolds and manifolds with corners. We introduce the right-angled Coxeter cellular decomposition of right-angled Coxeter orbifolds, and discuss their orbifold fundamental groups. In addition, we also introduce the theory of fundamental domain. In Section 3 we generalize the notions of B-belts and flags from simple polytopes to simple handlebodies. In Section 4, we give a right-angled Coxeter cellular decomposition of a simple n-handlebody Q, so that we can explicitly give a presentation of orbifold fundamental group $\pi_1^{\mathrm{orb}}(Q)$. We further show that this presentation of $\pi_1^{\mathrm{orb}}(Q)$ is an iterative HNN-extension of some right-angled Coxeter group. Moreover, the orbifold universal cover of Q is constructed by using $\pi_1^{\mathrm{orb}}(Q)$ and the simple polytope P_Q associated to Q. In Section 5, we review the work of Gromov, and compute the homology groups of the manifold double and universal cover of a simple handlebody Q by Davis' method, which are useful in the proof of Theorem A. Then we prove Theorem A. In Section 6, we show that the existence of a rank-two free abelian subgroup in the orbifold fundamental group of a simple handlebody Q is characterized by a $\square$-belt in Q (Theorem B). In Section 7, applying Theorems A and B and some results of geometry, we discuss the existence of some curvatures on a smooth manifold double over a simple handlebody Q in terms of the combinatorics of Q.

Acknowledgements

Lü was partially supported by NSFC 11971112. Wu was supported by NSFC 12201359 and the Natural Foundation of Shandong Province (ZR2022QA028). The authors thank the referee very much for valuable comments and suggestions, which made our paper more readable.

2 Preliminaries

2.1 Orbifold

As a generalization of manifolds, an n-dimensional *orbifold* $\mathbb{O}$ is a singular space which is locally modeled on the quotient of a finite group acting on an open subset of $\mathbb{R}^n$. For any $p \in \mathbb{O}$, there is an *orbifold chart*

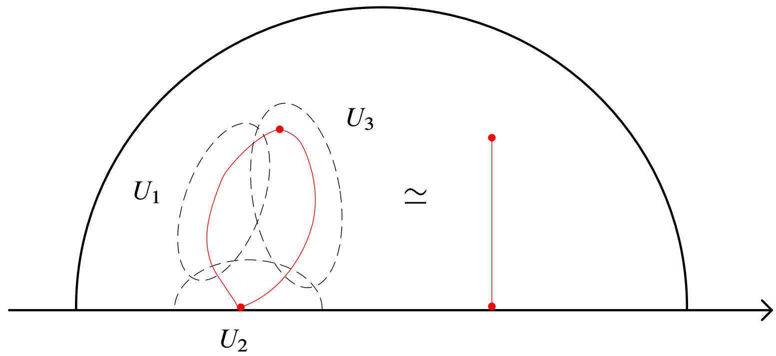


Figure 2: An orbifold loop.

(U, G, ψ) satisfying that U is an n-ball centered at origin O and $\psi^{-1}(p) = O$, where $\psi: U \to U/G$ is the projection map. In particular, the origin O is fixed by G. We called G the *local group* at p.

Definition 2.1 (Thurston [58, Definition 13.2.2]) A *covering orbifold* of an orbifold $\mathbb{O}$ is an orbifold $\widetilde{\mathbb{O}}$ with a projection $\pi: \widetilde{\mathbb{O}} \to \mathbb{O}$, satisfying that:

- Every $x \in \mathbb{O}$ has a neighborhood V which is identified with an open subset U of $\mathbb{R}^n$ modulo a finite group G_x, such that each component V_i of $\pi^{-1}(V)$ is homeomorphic to U/Γ_i, where $\Gamma_i < G_x$ is some subgroup;
- $\pi|_{V_i}: V_i \to V$ corresponds to the natural projection $U/\Gamma_i \to U/G_x$.

An orbifold is *good* (resp. *very good*) if it can be covered (resp. finitely covered) by a manifold. Otherwise it is *bad*. Any orbifold $\mathbb{O}$ has an universal cover $\widetilde{\mathbb{O}}$; see [58, Proposition 13.2.4].

In general, the *orbifold fundamental group* of an orbifold is defined as the deck transformation group of its universal cover; see [58, Definition 13.2.5]. Another equivalent definition uses the notion of based orbifold loops, that is, the orbifold fundamental group is defined by the homotopy classes of based orbifold loops. For more details, see [12, Section 3].

Example 2.2 Let D^2 be the unit disk in $\mathbb{R}^2$. A transformation r on D^2 via $r(x, y) = (x, -y)$ gives a reflective $\mathbb{Z}_2$-action on D^2. The orbit space $D^2/\mathbb{Z}_2$ has a natural orbifold structure. Any $(x, 0) \in D^2/\mathbb{Z}_2$ is a singular point with local group $\mathbb{Z}_2$. Since D^2 is contractible, $\pi_1^{\mathrm{orb}}(D^2/\mathbb{Z}_2) \cong \mathbb{Z}_2$ is generated by the transformation r.

In the viewpoint of orbifold loops, any path between $(x_1, 0)$ and (x_2, y_2) with $y_2 > 0$ can be viewed as a nontrivial orbifold loop. It is clear that $D^2/\mathbb{Z}_2 \cong D^1 \times D^1/\mathbb{Z}_2 \simeq D^1/\mathbb{Z}_2$. Hence, $\pi_1^{\mathrm{orb}}(D^2/\mathbb{Z}_2) \cong \mathbb{Z}_2$ is generated by a based orbifold loop $D^1/\mathbb{Z}_2$; see Figure 2.

Example 2.3 [1] If a discrete group G acts properly discontinuously on a manifold M, then the orbit space M/G canonically inherits an orbifold structure. Here M/G is called the *quotient orbifold* by G acting on M.

Let $p: M \to \mathcal{O}$ be a regular orbifold cover over a good orbifold $\mathcal{O}$, where M is a manifold. Then by orbifold covering theory [12], the orbifold homotopy group of $\mathcal{O}$ is isomorphic to the homotopy group of M,

$$\pi_k^{\mathrm{orb}}(\mathcal{O}) \cong \pi_k(M), \quad k \geq 2.$$

Thus a good orbifold is orbifold-aspherical if and only if its covering manifold is aspherical.

See [1; 12; 13; 19; 54] for more details of orbifold homotopy theory.

2.2 Right-angled Coxeter orbifolds, manifolds with corners and their manifold covers

Following [21; 22], a *right-angled Coxeter n-orbifold Q* is a special n-orbifold locally modeled on the quotient $\mathbb{R}^n/(\mathbb{Z}_2)^n$ of the standard $(\mathbb{Z}_2)^n$-action on $\mathbb{R}^n$ by reflections across the coordinate hyperplanes. A *stratum* of codimension k is the closure of a component of the subspace of $|Q|$ consisting of all points with local group $(\mathbb{Z}_2)^k$, where $|Q|$ denotes the underlying space of Q. It is easy to see that $\mathbb{R}^n/(\mathbb{Z}_2)^n$ possesses the following properties:

- Topologically and combinatorially, $\mathbb{R}^n/(\mathbb{Z}_2)^n$ is the standard simplicial cone
$$\mathscr{C}^n = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_i \geq 0,\ 1 \leq i \leq n\}$$
in $\mathbb{R}^n$.
- The local group at $x = (x_1, \dots, x_n) \in \mathbb{R}^n/(\mathbb{Z}_2)^n$ is the subgroup $(\mathbb{Z}_2)^{c(x)}$, where $c(x)$ is the number of those coordinates $x_i = 0$ in x, called the *codimension* of x.
- For $0 \leq k \leq n$, $(\mathbb{Z}_2)^k$ as a local group determines $\binom{n}{k}$ strata of codimension k, each of which is isomorphic to $\mathbb{R}^{n-k}/(\mathbb{Z}_2)^{n-k}$.

Davis [17, Section 6] (or see [18, Chapter 10, page 180]) defined *n-manifolds with corners*, each of which is a Hausdorff space X together with a maximal atlas of local charts onto open subsets of the standard simplicial cone $\mathscr{C}^n$ such that the overlap maps are homeomorphisms of preserving codimension, where for any chart $\varphi: U \to \mathscr{C}^n$, the codimension of any $x \in U$ is defined as $c(\varphi(x))$, also denoted by $c(x)$, and it is independent of the chart. An *open face* of codimension k is a component of $\{x \in X \mid c(x) = k\}$. A *face* is the closure of such a component.

A right-angled Coxeter orbifold Q naturally inherits the structure of a manifold with corners. On the other hand, since the topological and combinatorial structure of $\mathscr{C}^n$ is compatible with that of right-angled Coxeter orbifold on $\mathbb{R}^n/(\mathbb{Z}_2)^n$, an n-manifold with corners naturally admits a right-angled Coxeter orbifold structure. Furthermore, all strata in a right-angled Coxeter orbifold Q bijectively correspond to all faces in Q as a manifold with corners. A stratum or face of codimension one is called a *facet*.

In this paper we are mainly concerned with a special class of right-angled Coxeter orbifolds, i.e. simple handlebodies, as defined in the beginning of this paper. Let Q be a simple n-handlebody with facet set $\mathscr{F}(Q) = \{F_1, \dots, F_m\}$. For some $n \leq k \leq m$, the map

$$\lambda: \mathscr{F}(Q) \to (\mathbb{Z}_2)^k$$

is called a *characteristic map* if for each l-face f^l in Q (so there are exactly $n-l$ facets, say $F_{i_1}, \dots, F_{i_{n-l}}$, whose intersection is f^l since Q is simple), $\lambda(F_{i_1}), \dots, \lambda(F_{i_{n-l}})$ are independent in $(\mathbb{Z}_2)^k$. Clearly, each l-face f^l in Q determines a subgroup G_{f^l} generated by $\lambda(F_{i_1}), \dots, \lambda(F_{i_{n-l}})$ via λ. Note that each $x \in \partial|Q|$ always lies in the relative interior of a unique face f. Then there is a *manifold cover* of Q defined as

$$\mathcal{U}(Q, (\mathbb{Z}_2)^k) = Q \times (\mathbb{Z}_2)^k/\sim, \tag{2-1}$$

where

$$(x, g) \sim (y, h) \iff \begin{cases} x = y \text{ and } g = h & \text{if } x \in \mathrm{Int}(|Q|), \\ x = y \text{ and } gh^{-1} \in G_f & \text{if } x \in f \subset \partial|Q|. \end{cases}$$

Essentially this is a special case of the "basic construction" of Davis [18, Chapter 5]. It follows from [18, Proposition 10.1.10] that $\mathcal{U}(Q, (\mathbb{Z}_2)^k)$ is an n-dimensional closed manifold and naturally admits an action of $(\mathbb{Z}_2)^k$ with quotient orbifold Q. So a simple handlebody is a very good orbifold.

Lemma 2.4 *A simple handlebody Q is the quotient orbifold of $(\mathbb{Z}_2)^k$ acting on $\mathcal{U}(Q, (\mathbb{Z}_2)^k)$.*

If $k = n$, then $\mathcal{U}(Q, (\mathbb{Z}_2)^n)$ is called *small manifold cover* over Q which is a generalization of small covers over simple polytopes (see [22]), but it may not exist even if Q is a simple polytope (see also [22, Nonexamples 1.22]). However, if $k = m$, we can take $\lambda(F_i) = e_i$ for each facet F_i of Q where $\{e_1, \dots, e_m\}$ is the standard basis of $(\mathbb{Z}_2)^m$, such that there always exists such $\mathcal{U}(Q, (\mathbb{Z}_2)^m)$, called the *manifold double* (see [21, Proposition 2.4]) over Q, which is a generalization of real moment angled manifolds over simple polytopes (see [9]). In this case, for simplicity, we use M_Q to replace $\mathcal{U}(Q, (\mathbb{Z}_2)^m)$.

2.3 The right-angled Coxeter cellular decomposition

Now let us introduce the right-angled Coxeter orbifold cellular decomposition for right-angled Coxeter orbifolds, which will play an important role on the calculation of the orbifold fundamental groups of right-angled Coxeter orbifolds. The more general notion of cellular decomposition of certain orbifolds was considered as q-cellular complexes (or, q-CW complexes) in [4; 51].

Let $r_i : \mathbb{R}^n \to \mathbb{R}^n$ be the i^{th} standard reflection defined by

$$r_i(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) = (x_1, \dots, x_{i-1}, -x_i, x_{i+1}, \dots, x_n).$$

All standard reflections in $\mathbb{R}^n$ induce a standard $(\mathbb{Z}_2)^n$-action on the closed unit n-ball B^n with a right-angled corner $B^n/(\mathbb{Z}_2)^n$ as its orbit space. Of course, $\mathrm{Int}\, B^n$ is $(\mathbb{Z}_2)^n$-equivariantly homeomorphic to $\mathbb{R}^n$.

Definition 2.5 (right-angled Coxeter cells) Let Γ be a group generated by some standard reflections in $\mathbb{R}^n$. Then the quotient B^n/Γ is called a *right-angled Coxeter n-ball*, and the quotient $\mathrm{Int}\, B^n/\Gamma$ is called an *open right-angled Coxeter n-ball*. Note that if Γ is not a trivial group, then the right-angled Coxeter n-ball B^n/Γ is an n-orbifold with boundary $\partial B^n/\Gamma$.

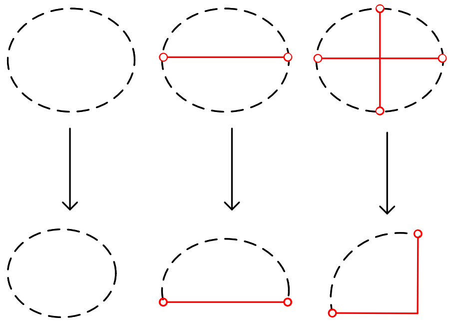


Figure 3: Right-angled Coxeter 2-cells.

If e^n is Γ-equivariantly homeomorphic to Int B^n, then the quotient e^n/Γ is called a *right-angled Coxeter n-cell*, and its closure is call a *closed right-angled Coxeter n-cell*.

For example, a right-angled Coxeter 1-cell is either a connected open interval or a semiopen and semiclosed interval whose closed endpoint gives a local group $\mathbb{Z}_2$. A right-angled Coxeter 2-cell has three possible types, with local group being the trivial group, $\mathbb{Z}_2$ or $(\mathbb{Z}_2)^2$, as shown in Figure 3.

An n-dimensional *right-angled Coxeter cellular complex* (or *Coxeter CW complex*) X can be constructed in the same way as CW complex (see [36, page 5]). A key point is that the attaching map of every n-dimensional right-angled Coxeter cell e^n_α/Γ to the $(n-1)$-skeleton X^{n-1},

$$\phi_\alpha\colon \partial\bar{e}^n_\alpha/\Gamma \to X^{n-1}, \tag{2-2}$$

is required to *preserve the local group of each point in $\partial\bar{e}^n_\alpha/\Gamma$*.

Here the attaching maps $\{\phi_\alpha\}$ of right-angled Coxeter cells with nontrivial local groups have a much stronger restriction than those in CW complexes. Actually, ϕ_α preserving local groups implies that singular points and nonsingular points of each embedding right-angled Coxeter n-cell are still singular and nonsingular, respectively, in X.

Remark 2.6 (right-angled Coxeter cubical cellular complex) Recall that a cubical cellular complex is a CW complex X whose cells are cubes, with the property that for two cubes c and c' of X, $c \cap c'$ is a common face of c and c'; in other words, cubes are glued in X via combinatorial isometries of their faces. Similarly, a *right-angled Coxeter cubical cellular complex* can be defined in the same way whose cells are all right-angled Coxeter cubical cells, that is, the orbits of standard reflections on an n-cube $[-1,1]^n$. For example, the standard cubical cellular decomposition of a simple polytope P (i.e. the cone of the barycentric subdivision of $\mathcal{N}(P)$) gives a right-angled Coxeter cubical cellular complex structure of P. Of course, right-angled Coxeter cubical cellular complexes form a special class of right-angled Coxeter cellular complexes.

Proposition 2.7 *Each simple handlebody has a finite right-angled Coxeter cellular complex structure.*

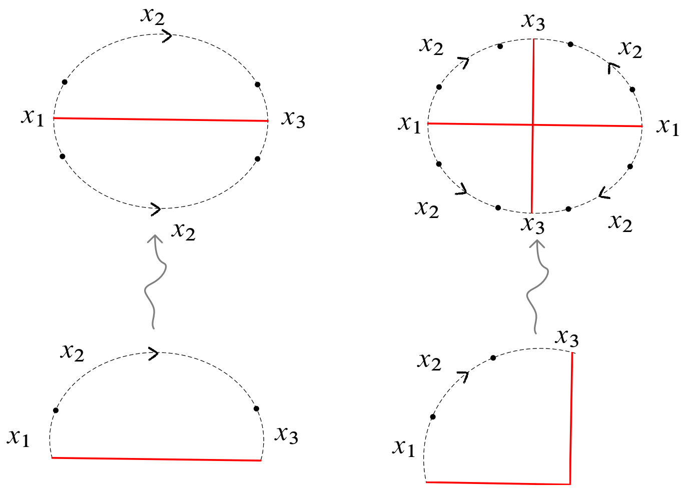


Figure 4: Relations determined by right-angled Coxeter 2-cells in the case $n = 3$.

Proof Let Q be a simple n-handlebody with the associated simple polytope P_Q. Then the standard cubical subdivision of P_Q induces a right-angled Coxeter cellular decomposition of Q. More details will be shown in Section 4. □

Remark 2.8 It should be pointed out that each simple handlebody still has a right-angled Coxeter cubical cellular complex structure. This can be seen in Section 7.1.

In general, a right-angled Coxeter cellular complex is naturally an orbispace. Here its orbifold fundamental group is defined by the homotopy classes of based orbifold loops. For more details, see [12, Section 3]. Although a right-angled Coxeter cell with nontrivial local group is not contractible in the sense of orbifolds, all attaching maps $\{\phi_\alpha\}$ preserving local groups ensure that the orbifold fundamental group of a right-angled Coxeter cellular complex is isomorphic to the orbifold fundamental group of its 2-skeleton.

Proposition 2.9 *Let X be a right-angled Coxeter cellular complex. Then*

$$\pi_1^{\mathrm{orb}}(X^2) \cong \pi_1^{\mathrm{orb}}(X),$$

where X^2 is the 2-skeleton of X.

Proof The argument can be proved in a similar way as shown by Hatcher [36, Proposition 1.26]. The only thing to note is that the local group information of each right-angled Coxeter n-cell can be inherited by the boundary orbifold of its closure in X^{n-1}. □

Remark 2.10 We can easily read out the generators and relations of $\pi_1^{\mathrm{orb}}(X) \cong \pi_1^{\mathrm{orb}}(X^2)$ from the 2-skeleton of a right-angled Coxeter cellular complex X. Let us look at a right-angled Coxeter 2-cell with nontrivial local group in X. Assume that the boundary of a right-angled Coxeter 2-cell with nontrivial local group consists of $x_1, x_2, \ldots, x_n$, where each x_i is a closed oriented orbifold loop in X, and only one endpoint of x_1 and x_n has nontrivial local group. Regard these closed orbifold loops as generators. Then $x_1^2 = x_n^2 = 1$ and the right-angled Coxeter 2-cells with local group $\mathbb{Z}_2$ give a relation $x_1 x_2 \cdots x_n \cdot x_{n-1}^{-1} \cdots x_2^{-1} = 1$, while the right-angled Coxeter 2-cells with local group $\mathbb{Z}_2^2$ give a relation $(x_1 x_2 \cdots x_n \cdot x_{n-1}^{-1} \cdots x_2^{-1})^2 = 1$. This can intuitively be seen from Figure 4 when $n = 3$.

Example 2.11 Let P be a simple polytope with facet set $\mathcal{F}(P)$. Regard P as a right-angled Coxeter orbifold. The standard cubical subdivision of P is a right-angled Coxeter cellular decomposition of P. Calculating the orbifold fundamental group of P by the 2-skeleton of its right-angled Coxeter cellular decomposition, $\pi_1^{\text{orb}}(P)$ can be represented by the right-angled Coxeter group W_P of P:

$$\pi_1^{\text{orb}}(P) \cong W_P = \langle s_F, F \in \mathcal{F}(P) \mid s_F^2 = 1 \text{ for all } F; (s_F s_{F'})^2 = 1 \text{ for } F \cap F' \neq \varnothing \rangle.$$

2.4 Group action and fundamental domain ([18, page 64] or [59, pages 159–161])

Suppose that a discrete group G acts properly on a connected topological space X. A closed subset $D \subset X$ is a *fundamental domain* for the G-action on X if each G-orbit intersects with D and if for each point x in the interior of D, $G(x) \cap D = \{x\}$. In other words, $\{gD \mid g \in G\}$ forms a locally finite cover for X, such that no two of $\{gD \mid g \in G\}$ have common interior points. Such $\{gD \mid g \in G\}$ is called a *decomposition* for X so

$$X = \bigcup_{g \in G} gD$$

and each gD is called a *chamber* of G on X.

Throughout the following, the fundamental domain of G acting on X will be taken as a simple polytope D. Then each $g \in G$ gives a self-homeomorphism of X

$$\phi_g : X \to X$$

by mapping chamber hD to $g \cdot hD$ for any $h \in G$. If two chambers gD and hD have a nonempty intersection which includes some facets of gD and hD, then there is a homeomorphism $\phi_{hg^{-1}}$ that maps gD to hD. Hence, for two facets F and F' from gD and hD, respectively, that are glued together in X, naturally we can assign hg^{-1} and gh^{-1} to F and F', respectively. This means that the action of G on X gives a *characteristic map* on the facets set of D,

$$\lambda : \mathcal{F}(D) \to G.$$

For each facet F of D, $\lambda(F) \in G$ is called a *coloring* on F. Each $\lambda(F) \in G$ naturally determines a self-homeomorphism $\phi_{\lambda(F)} \in \text{Homeo}(X)$, which is called an *adjacency transformation* on X with respect to F. Such $\phi_{\lambda(F)}$ maps each chamber into an adjacent chamber such that the facet F is contained in the intersection of those two chambers. Each adjacency transformation has an inverse adjacency transformation corresponding to a facet F' of D. Of course, $F = F'$ is allowed. In this case, we call F a *mirror* of X associated with G, and the corresponding adjacency transformation is called a *reflection* of X with respect to F.

Remark 2.12 It should be pointed out that two adjacency transformations determined by different facets of D are viewed as being different, although they may correspond to the same self-homeomorphism of X. The inverse adjacency transformation of an adjacency transformation determined by a facet F is exactly determined by another facet F' which is identified with F in X.

All inverse adjacency transformations give an equivalence relation $\sim$ on $\mathscr{F}(D)\times G$, where $(F,g)\sim(F',h)$ if and only if

$$\lambda(F)\cdot g = h, \quad \lambda(F')\cdot h = g. \tag{2-3}$$

In other words, if two chambers gD and hD are attached together by identifying a facet F of gD with a facet F' of hD in X, then $\lambda(F)\cdot\lambda(F') = 1$, which gives a *pair relation* for G. When F is a mirror, the pair relation is $\lambda(F)\cdot\lambda(F) = 1$.

Remark 2.13 The equivalence relation $\sim$ on $\mathscr{F}(D)\times G$ gives an equivalence relation $\sim'$ on $\mathscr{F}(D)$ via the projection $\mathscr{F}(D)\times G \to \mathscr{F}(D)$ as follows: for $F, F' \in \mathscr{F}(D)$,

$$F \sim' F' \iff (F,g)\sim(F',h) \text{ for } g,h \in G \text{ satisfying relation (2-3).}$$

Thus, we can obtain a quotient orbifold $D/{\sim'}$ by attaching some facets on the boundary of D via the equivalence relation $\sim'$ on $\mathscr{F}(D)$.

On the contrary, giving a simple polytope D and a characteristic map satisfying relation (2-3), we can construct a space X with G-action by

$$X = D \times G/{\sim}, \tag{2-4}$$

where the equivalence relation is defined in relation (2-3).

The construction of X gives a natural polyhedral cellular decomposition of X, denoted by $\mathscr{P}(X)$. The dual complex of $\mathscr{P}(X)$ is denoted by $\mathscr{C}(X)$. If each codimension-k face of D in X intersects with exactly 2^k chambers, then each cell of $\mathscr{C}(X)$ is a cube, which is exactly one induced by the standard cubical decomposition of the simple polytope D. Furthermore, if $\mathscr{C}(X)$ is a cubical complex, then the link of each vertex in $\mathscr{C}(X)$ is a simplicial complex which is exactly the boundary complex of the dual of D. The 1-skeleton of $\mathscr{C}(X)$ is exactly the Cayley graph of G with generator set consisting of adjacency transformations determined by all facets of D. Therefore, one has that:

Lemma 2.14 [59, page 160] *The group G is generated by all adjacency transformations.*

To simplify notation, let $\lambda(F_i) = s_i$ or s_{F_i} for each $F_i \in \mathscr{F}(D)$. Then for each $g \in G$, ϕ_g can be decomposed into the composition of some adjacency transformations,

$$g = s_{i_1}s_{i_2}\cdots s_{i_k}.$$

The relations with $s_{i_1}s_{i_2}\cdots s_{i_k} = 1$, except pair relations, are called *Poincaré relations*.

Lemma 2.15 [59, page 161] *The Poincaré relations together with the pair relations form a set of relations of the group G.*

For each codimension-2 face of D, there is a Poincaré relation with form $s_k s_{k-1} \cdots s_1 = 1$ (alternatively, $s_1' \cdots s_k' = 1$, where $s_i' = (s_i)^{-1}$ for each i).

Define a group G_D with generators consisting of all adjacency transformations determined by $\mathcal{F}(D)$ and relations formed by all pair relations and Poincaré relations determined by all codimension 2 faces in D:

(2-5) $$G_D = \langle s_i \text{ for } F_i \in \mathcal{F}(D) \mid s_i s_j = 1 \text{ for } F_i \sim' F_j;\, s_{i_1} s_{i_2} \cdots s_{i_k} = 1 \text{ for each codim-2 face in } D \rangle$$

For the sake of preciseness, suppose again that each codimension-k face of D in X intersects with exactly 2^k chambers. Then the cubical subdivision of D induces a right-angled Coxeter cellular decomposition for the quotient orbifold X/G. It is not difficult to see that $D/{\sim'}$ is isomorphic to X/G as orbifolds. According to Proposition 2.9, G_D is isomorphic to the orbifold fundamental group of the quotient space X/G. Therefore, we have the following lemma.

Lemma 2.16 *The orbifold fundamental group of $D/{\sim'} \cong X/G$ is isomorphic to G_D.*

There is a natural quotient map $\lambda_*\colon G_D \to G$, and the image of λ_* on each adjacency transformation s_F is the coloring on corresponding facet F. Then the fundamental group of X is isomorphic to the kernel of λ_*.

Proposition 2.17 *Let G be a discrete group which acts properly discontinuously on a manifold X. Suppose X is decomposed into $X = \bigcup_{g \in G} gD = D \times G/{\sim}$, where D is a simple polytope and each codimension-k face of D in X intersects with exactly 2^k chambers. Let G_D be the group defined as in expression* (2-5), *and λ_* be the quotient map from G_D to G induced by the characteristic map $\lambda\colon \mathcal{F}(D) \to G$. Then there is a short exact group sequence*

$$1 \to \pi_1(X) \to G_D \xrightarrow{\lambda_*} G \to 1$$

which is induced by the orbifold covering $\pi\colon X \to X/G$.

Proof We refer to Chen [12, pages 40–49]. Here it is only necessary to show that $G_D \cong \pi_1^{\mathrm{orb}}(X/G)$, which is exactly Lemma 2.16. □

Given a simple convex polytope D and a discrete group G, assume that there exists a characteristic map $\lambda\colon \mathcal{F}(D) \to G$ such that $X = D \times G/{\sim}$ is a G-manifold, where $(F, g) \sim (F', h)$ for any $F, F' \in \mathcal{F}(D)$ and $g, h \in G$ if and only if relation (2-3) holds. Then we have the following result.

Corollary 2.18 *Under the assumption of Proposition 2.17, X is simply connected if and only if $G \cong G_D$.*

Example 2.19 Let $\square$ be a square with faces F_1, F_2, F_3 and F_4 colored by e_1, e_2, e_1 and e_2 respectively, where $e_1 = (-1, 1)$ and $e_2 = (1, -1)$ are generators of $(\mathbb{Z}_2)^2$. Then $X = \square \times (\mathbb{Z}_2)^2/\sim \cong T^2$ is a small cover over $\square$ (see [22]), and

$$G_\square = \langle s_1, s_2, s_3, s_4 \mid s_i^2 = 1; (s_1s_2)^2 = (s_2s_3)^2 = (s_3s_4)^2 = (s_4s_1)^2 = 1\rangle \cong (\mathbb{Z}_2 * \mathbb{Z}_2) \oplus (\mathbb{Z}_2 * \mathbb{Z}_2)$$

is the right-angled Coxeter group determined by $\square$. Then $\pi_1(X) \cong \ker \lambda_* = \mathbb{Z}^2$ is a normal subgroup of $G_\square$ generated by Poincaré relations s_1s_3 and s_2s_4.

2.5 Right-angled Coxeter group and HNN-extension

In this subsection, we refer to [18, Chapter 3] and [44, Chapter 4].

Let $w = s_1s_2\cdots s_m$ be a word in a right-angled Coxeter group $W = \langle S \mid R\rangle$. An *elementary operation* on w is one of the following two types of operations:

(i) **Length-reducing** Delete a subword of ss.

(ii) **Braid (commutation)** Replace a subword of the form st with ts if $(st)^2 = 1$ in the relations set R of W.

A word is *reduced* if it cannot be shorten by a sequence of elementary operations.

Theorem 2.20 (Tits [18, Theorem 3.4.2]) *Two reduced words x and y are the same in a right-angled Coxeter group if and only if one of both x and y can be transformed into the other one by a sequence of elementary operations of type (ii).*

Definition 2.21 (Higman–Neumann–Neumann extension [44, page 179]) Let G be a group with presentation $G = \langle S \mid R\rangle$, and let $\phi\colon A \to B$ be an isomorphism between two subgroups of G. Let t be a new symbol out of S. Then the HNN-extension of G relative to ϕ is defined as

$$G*_\phi = \langle S, t \mid R, t^{-1}gt = \phi(g), g \in A\rangle.$$

Let $\omega = g_0t^{\epsilon_1}g_1t^{\epsilon_2}\cdots g_{n-1}t^{\epsilon_n}g_n$ $(n \geq 0)$ be an expression in $G*_\phi$, where each g_i is an element in G (probably g_i may be taken as the unit element 1 in G), and ϵ_i is either number 1 or -1. Then ω is said to be *t-reduced* if there is no consecutive subword $t^{-1}g_it$ or tg_jt^{-1} with $g_i \in A$ and $g_j \in B$.

A *normal form* of an element in $G*_\phi$ is a word $\omega = g_0t^{\epsilon_1}g_1t^{\epsilon_2}\cdots g_{n-1}t^{\epsilon_n}g_n$ $(n \geq 0)$ where

(i) g_0 is an arbitrary element of G;

(ii) if $\epsilon_i = -1$, then g_i is a representative of a coset of A in G;

(iii) if $\epsilon_i = +1$, then g_i is a representative of a coset of B in G;

(iv) there is no consecutive subword $t^\epsilon 1t^{-\epsilon}$.

Theorem 2.22 (the normal form theorem for HNN-extensions [44, Theorem 2.1, page 182]) *Let $G*_\phi = \langle G, t \mid t^{-1}gt = \phi(g), g \in A\rangle$ be an HNN-extension. Then there are two equivalent statements*:

(I) *The group G is embedded in $G*_\phi$ by the map $g \mapsto g$. If $\omega = g_0t^{\epsilon_1}g_1\cdots t^{\epsilon_n}g_n = 1$ in $G*_\phi$, then ω is not reduced.*

(II) *Every element ω of $G*_\phi$ has a unique representation $\omega = g_0t^{\epsilon_1}g_1\cdots t^{\epsilon_n}g_n$ which is a normal form.*

A *t-reduction* of $\omega = g_0t^{\epsilon_1}g_1\cdots t^{\epsilon_n}g_n$ is one of the following two operations:

- replace a subword of the form $t^{-1}gt$, where $g \in A$, by $\phi(g)$;
- replace a subword of the form tgt^{-1}, where $g \in B$, by $\phi^{-1}(g)$.

A finite number of t-reductions leads from $\omega = g_0t^{\epsilon_1}g_1\cdots t^{\epsilon_n}g_n$ to a normal form.

3 Flagness and beltness of simple handlebodies

3.1 B-belts and flagness

Assume that Q is a simple n-handlebody with nerve $\mathcal{N}(Q)$. Denote by Q^* the *dual* of Q, whose facial structure is given by $\mathcal{N}(Q)$.

Definition 3.1 (B-belts) Let $i : B \hookrightarrow Q$ be an embedding closed simple k-suborbifold whose underlying space is a k-ball. We say that $i(B)$ is a *B-belt* of Q if

- i preserves codimensions, i.e. i maps each codimension-d face f of B to a codimension-d face F_f of Q;
- the intersection $\bigcap f_\alpha = \varnothing$ for some facets f_α in B if and only if either $\bigcap F_{f_\alpha} = \varnothing$ or $\bigcup F_{f_\alpha}$ cannot deformatively retract onto B in $|Q|$.

Remark 3.2 The orbifold embedding $i : B \hookrightarrow Q$ preserving codimension is equivalent to that i restricting on the local group of each point in B induces an identity. The statement that $\bigcup F_{f_\alpha}$ cannot deformatively retract onto B in $|Q|$ is equivalent to that there is at least a hole in the area surrounded by $\{F_{f_\alpha}\}$ and B.

A simple polytope P itself is a P-belt. For a B-belt in a simple polytope P, the intersection $\bigcap f_\alpha = \varnothing$ for some facets f_α in B if and only if $\bigcap F_{f_\alpha} = \varnothing$. And each B-belt is *π-injective* in the sense of Lemma 7.1, which is an analogue of π_1-injective surfaces in 3-dimensional manifolds.

A 2-dimensional B-belt in a simple 3-handlebody Q is a k-gon. Traditionally, such a B-belt is also called a *k-belt* of Q. In the case of dimension three, any simple 3-polytope except the tetrahedron has a 2-dimensional B-belt.

Next, we want to generalize the definition of flagness to simple handlebodies in terms of B-belts defined above. Recall that a simplicial complex K with vertex set V is a *flag* complex if every finite subset of V, which is pairwise joined by edges, spans a simplex. Let X be a cubical complex equipped with a piecewise Euclidean structure. Then Gromov's lemma (see [32]) tells us that X is nonpositively curved if and only if the link of each vertex in X is a flag simplicial complex. Furthermore, by the Cartan–Hadamard theorem, a nonpositively curved space is aspherical.

A simple polytope P is *flag* if the boundary complex of its dual is a flag simplicial complex. Let $M \to P$ be a small cover or a real moment angled manifold over P. Then we know from [24, Theorem 2.2.5] that M is aspherical if and only if P is flag. Equivalently, P, as a right-angled Coxeter orbifold, is orbifold-aspherical if and only if it is flag.

Naturally, the flagness of a simple handlebody Q should still be closely related to the orbifold-asphericity of Q. The right-angled Coxeter orbifold structure of Q induces a facial structure of Q which can be carried by the nerve $\mathcal{N}(Q)$ of Q as a manifold with corners. The combinatorial obstruction of orbifold-asphericity of Q contains some quotient orbifolds of S^k for $k \geq 2$ by reflective actions of $(\mathbb{Z}_2)^l$ for $1 \leq l \leq k+1$. Moreover, since Q is a simple handlebody, $S^k/(\mathbb{Z}_2)^{k+1} \cong \Delta^k$ is the unique possible combinatorial obstruction of orbifold-asphericity of Q. Notice that the orbifold-asphericity of Q is determined by both $\mathcal{N}(Q)$ and $|Q|$.

Definition 3.3 A simple handlebody Q is said to be *flag* if it contains no $\triangle^k$-belt for any $k \geq 2$.

Remark 3.4 The notion of B-belt and flagness can be generalized to a right-angled Coxeter orbifold whose underlying space is an arbitrary compact manifold with nonempty boundary. We call such an orbifold a *simple orbifold* if it satisfies conditions (b) and (c) in the definition of simple handlebody. A natural conjecture arises as follows:

Conjecture 3.5 *A simple orbifold Q is orbifold-aspherical if and only if its underlying space $|Q|$ is aspherical as a manifold and Q contains no Δ^k-belt for any $k \geq 2$.*

Davis' results in [18, Theorem 9.1.4] and [20, Theorem 3.5] tell us that the conjecture is true in the cases where $|Q|$ is acyclic or Q has a corner structure defined in [20, Section 3.1]. Our Theorem A also proves the case that Q is a simple handlebody. All of these support the conjecture.

A simple handlebody Q is a simple polytope or there exist finitely many disjoint B-belts of codimension one, called *cutting belts*, such that Q can be cut open into a simple polytope P_Q along those cutting belts. Here the cutting operation is similar to a hierarchy of Haken manifolds (or Haken orbifolds). Note that a simple handlebody is not a Haken orbifold except that it is flag. Refer to [28; 60] for Haken 3-manifolds and generalized Haken manifolds.

Let Q be a simple handlebody. We see that some vertices $F_1, F_2, \ldots, F_k$ of $\mathcal{N}(Q)$ span a simplex $\triangle^{k-1}$ in $\mathcal{N}(Q)$ if and only if the associated vertices span a simplex in $\mathcal{N}(P_Q)$, and they span an *empty simplex* (that is, $\partial\triangle^{k-1} \subset \mathcal{N}(Q)$ but $\triangle^{k-1}$ itself is not in $\mathcal{N}(Q)$) whose interior is contained in the interior of Q^*

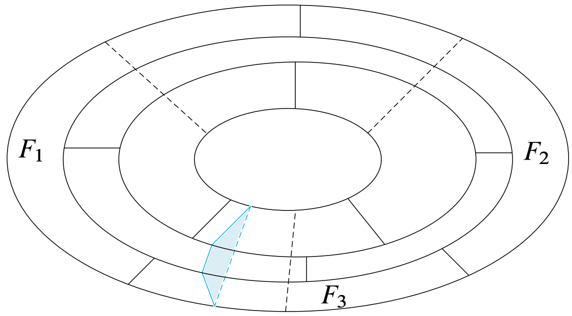


Figure 5: A flag simple 3-handlebody whose nerve is not a flag simplicial complex.

if and only if associated vertices span an empty simplex in $\mathcal{N}(P_Q)$. Specifically, those empty simplices correspond to some $\triangle^k$-belts in Q. Hence, we have the following result.

Lemma 3.6 *A simple handlebody Q is flag if and only if the associated simple polytope P_Q is flag* (*in other words*, $\mathcal{N}(P_Q)$ *is a flag simplicial complex*).

Remark 3.7 Notice that a flag simple handlebody defined above may contain an empty simplex whose interior cannot be embedded in its dual Q^*, as shown in Figure 5 for three pairwise intersected faces F_1, F_2 and F_3 in a flag simple solid torus. Therefore, the statement that $\mathcal{N}(Q)$ is a flag simplicial complex is not equivalent to that Q is a flag simple handlebody.

3.2 □-belts in a simple handlebody

Definition 3.8 A □-belt in a simple handlebody Q is a B-belt where B is a two-disk with a square boundary.

Remark 3.9 (1) In Gromov's paper [32, Section 4.2], *Siebenmann's no □-condition* for a flag simplicial complex K means no *empty square* in K, where an empty square in K must make sure that neither pair of opposite vertices is connected by an edge, which is a special case in our definition.

(2) A prismatic 3-circuit (see [53]) in a simple 3-polytope P^3 determines a Δ^2-belt in P^3. If there is no prismatic 3-circuit in P^3, then P^3 is a flag polytope or a tetrahedron. Similarly for a prismatic 4-circuit (see [53]) in a flag simple 3-polytope, it determines a □-belt in P in our definition.

Next, we give two lemmas as the preliminary of the proof of Theorem B.

Let Q be a simple handlebody, and $B_\square$ be a □-belt in Q with four ordered edges f_1, f_2, f_3 and f_4, any two of which have a nonempty intersection except for pairs $\{f_1, f_3\}$ and $\{f_2, f_4\}$. Assume that each f_i is contained in a facet F_i of Q. Then we may claim that $\{F_i \mid i = 1, 2, 3, 4\}$ must be different from each other. More precisely, we have the following lemma.

Lemma 3.10 *Let Q be a simple handlebody, and $B_{\square}$ be a $\square$-belt in Q. Then*

- *two adjacent edges of $B_{\square}$ cannot be contained in the same facet of Q;*
- *two disjoint edges of $B_{\square}$ cannot be contained in the same facet of Q.*

Proof Assume that the edges $\{f_1, f_2, f_3, f_4\}$ of $B_{\square}$ are contained in four ordered facets $\{F_1, F_2, F_3, F_4\}$ of Q, respectively. If there are two adjacent edges of $B_{\square}$ contained in the same facet of Q — without loss of generality, suppose that $F_1 = F_2$ — then $f_1 \cap f_2 \neq \varnothing$ implies that F_1 has a self-intersection, which is equivalent to there being a 1-simplex which bounds a single vertex in $\mathcal{N}(Q)$. This contradicts that Q is simple.

Similarly, if there are two disjoint edges of $B_{\square}$ contained in the same facet of Q, then one can assume that $F_1 = F_3$. This happens only for the case where the genus of Q is more than zero since $B_{\square}$ is a $\square$-belt in Q. Thus there are some holes between F_1 and $B_{\square}$. However, F_2 is contractible, so this induces that $F_2 \cap F_1$ is disconnected. In other words, there are two 1-simplices which bound the same two vertices in $\mathcal{N}(Q)$. This is also impossible since Q is simple. $\square$

Lemma 3.10 tells us that in a simple handlebody Q, a $\square$-belt can be presented as four different vertices $\{F_1, F_2, F_3, F_4\}$ in $\mathcal{N}(Q)$, which satisfies the following two conditions:

(I) $\{F_1, F_2, F_3, F_4\}$ bounds a square with its interior located in the interior of Q^* and with its edges contained in the 1-skeleton of $\mathcal{N}(Q)$.

(II) The full subcomplex spanned by $\{F_1, F_2, F_3, F_4\}$ in $\mathcal{N}(Q)$ is either a square or a nonsquare subcomplex (containing two 2-simplices gluing along an edge). Here the latter "a nonsquare subcomplex" may happen only when the genus of Q is more than zero.

Example 3.11 (squares in the dual of a simple handlebody) Let Q be a simple handlebody, and Q^* be its dual. There are some possible cases of squares and nonsquares in Q^*, listed in Figure 6, where all vertices and edges are considered in $\mathcal{N}(Q)$. Diagrams (a) and (b) are not squares in Q^*, while (c) and (d) are. Notice that (d) is not an empty square in $\mathcal{N}(Q)$, which is different from the case of Siebenmann's no $\square$-condition, as stated in Remark 3.9(1).

Lemma 3.12 *Let $B_{\square}$ be a $\square$-belt in a simple n-handlebody Q, and B be a cutting belt of Q. Then either $B_{\square}$ and B can be separated in Q, or B intersects transversely with only a pair of disjoint edges of $B_{\square}$.*

Proof Assume that the four ordered edges f_1, f_2, f_3 and f_4 of $B_{\square}$ are contained in four facets F_1, F_2, F_3 and F_4 of Q, respectively. Since $B_{\square}$ and B are contractible, we see that B and $B_{\square}$ can be separated if and only if their boundaries can be separated.

First we assume that ∂B and $\partial B_{\square}$ intersect transversely, meaning that $\partial B \cap \partial B_{\square}$ is a set of isolated points cyclically ordered on the boundary of $B_{\square}$, which is denoted by $\mathcal{V}$. Then $\mathcal{V}$ contains at least two points if $\mathcal{V}$ is nonempty.

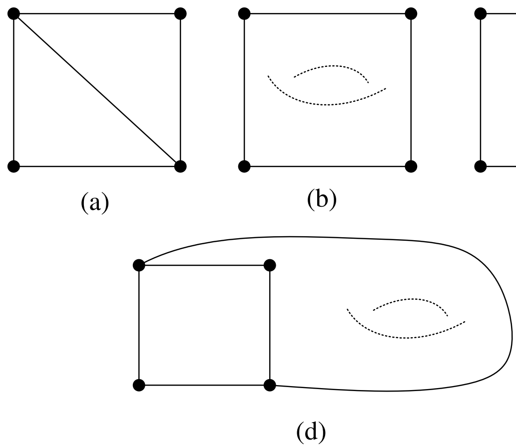


Figure 6: Squares and nonsquares.

Let v and v' be two adjacent points in $\mathcal{V}$. Then there are the following cases:

(i) v and v' are located in the same edge of $B_\square$;

(ii) v and v' are located in two adjacent edges of $B_\square$;

(iii) v and v' are located in two disjoint edges of $B_\square$.

In the case (i), without loss of generality, suppose that $v, v' \in \mathrm{int}(f_1)$. Now if v and v' are contained in the same connected component of $F_1 \cap B$ (without a loss of generality, assume that B is regarded as B_1 of (a) in Figure 7), then we can deform the interior of f_1 such that $f_1 \cap \partial B = \varnothing$ will not contain v and v'. If v and v' are contained in two connected components of $F_1 \cap B$, without loss of generality, assume that B is regarded as B_2 of (a) in Figure 7. Since B is a B-belt, there is a hole surrounded by f_1 and B. This case is allowed (also see (b) and (c) in Figure 7).

In the case (ii), without loss of generality, assume that B intersects with f_1 and f_2. Now if $B \cap F_1 \cap F_2 \neq \varnothing$ (regard B as B_3 of (a) in Figure 7), then we can move vertex $f_1 \cap f_2$ in $F_1 \cap F_2$ such that $\partial B_\square \cap \partial B$ does not contain v and v'.

Repeating this operation, we can assume that any two adjacent points v and v' in $\mathcal{V}$ cannot remove. This means that $B \cap F_1 \cap F_2 = \varnothing$ in the case (ii), so we may regard B as B_4 of (a) in Figure 7. Then by the definition of B-belt, there is a hole in the area surrounded by B, f_1 and f_2 (see (d) in Figure 7). If $|\mathcal{V}| = 2$, then $B_\square$ will not be contractible. This is a contradiction. If $|\mathcal{V}| > 2$, let v'' be a point after v' by the cyclic order of all isolated points in $\mathcal{V}$. If v' and v'' belong to the same edge f of $B_\square$, then there must be a hole surrounded by f and B. If v' and v'' belong to two adjacent edges f' and f'' of $B_\square$, then there is also a hole surrounded by f', f'' and B. If v' and v'' belong to two disjoint edges f' and f'' of $B_\square$, then there is still a hole surrounded by f, f'' and B, where f is the edge containing v. Whichever of all possible cases above happens implies that $\partial B_\square$ is not contractible in $|Q|$, but this is impossible.

The case (iii) is allowed; see B_5 of (a) in Figure 7. So the conclusion holds. $\square$

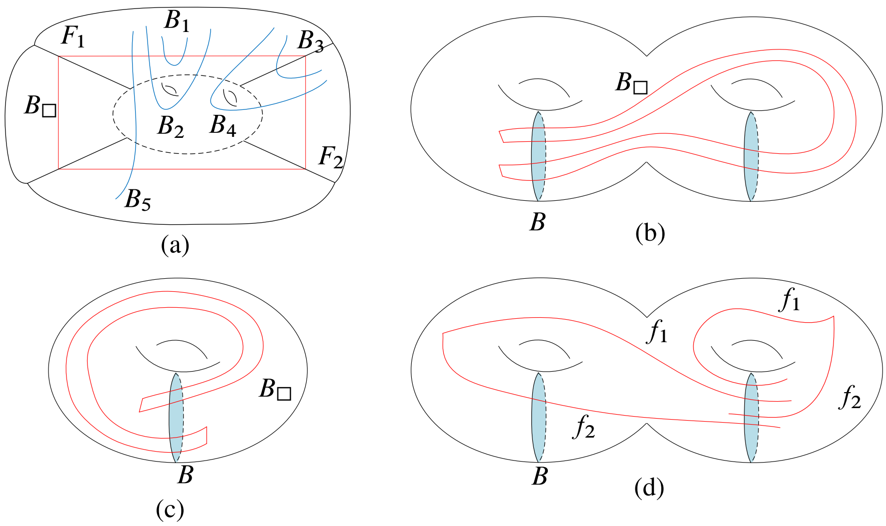


Figure 7: $\square$-belt and cutting belt.

We see that if there are some cutting belts that intersect with $B_\square$, then one can do some deformations such that those cutting belts either do not intersect with $B_\square$ or intersect transversely with only a pair of disjoint edges of $B_\square$.

4 The orbifold fundamental groups of simple handlebodies

4.1 The right-angled Coxeter cellular decomposition of simple handlebodies

Let Q be a simple n-handlebody of genus $\mathfrak{g}$ with facet set $\mathcal{F}(Q) = \{F_1, \dots, F_m\}$. Then we can cut Q into a simple polytope P_Q along $\mathfrak{g}$ cutting belts $B_1, \dots, B_{\mathfrak{g}}$, each of which intersects transversely with some facets of Q and is a simple $(n-1)$-polytope. Two copies of B_i in P_Q, denoted by B_i^+ and B_i^-, are two disjoint facets of P_Q. Since they share the common belt B_i in Q, by $B_i^+ \sim B_i^-$ we denote this share between them. The number of facets of P_Q around B_i^+ is the same as the number of facets of P_Q around B_i^-. In addition, each facet F of P_Q around B_i^+ also uniquely corresponds to a facet F' of P_Q around B_i^- such that F and F' share a common facet in Q, so by $F \cap B_i^+ \sim F' \cap B_i^-$ we mean this share between F and F' via the belt B_i of Q.

Let $\mathcal{F}(P_Q)$ denote the set of all facets in P_Q and $\mathcal{F}_B$ denote the set of those facets in P_Q produced by cutting belts of Q, so $\mathcal{F}_B$ contains $2\mathfrak{g}$ facets of P_Q, appearing in pairs.

P_Q is viewed as a right-angled Coxeter orbifold with boundary consisting of all facets in $\mathcal{F}_B$. By attaching all pairs $B^+ \sim B^-$ in $\mathcal{F}_B$ and all corresponding pairs (F, F') with $F \cap B^+ \sim F' \cap B^-$ together, we can recover Q from P_Q. Thus Q can be regarded as a quotient $P_Q/{\sim}$, and we denote the quotient map by

$$q\colon P_Q \to Q. \tag{4-1}$$

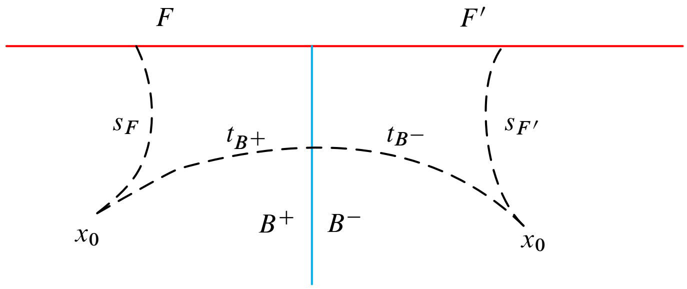


Figure 8: The right-angled Coxeter 2-cell nearby B-belt.

There is a canonical right-angled Coxeter cubical cellular decomposition $\mathscr{C}(P_Q)$ of P_Q, whose cells consist of

- all cubes in the standard cubical decomposition of P_Q;
- all cubes in the standard cubical decomposition of all boundary components of P_Q in $\mathscr{F}_B$.

Moreover, $\mathscr{C}(P_Q)$ induces a right-angled Coxeter cellular decomposition on Q by attaching some cubical cells of the copies of B-belts. Let c be a k-cube in $\mathscr{C}(P_Q)$ and $B \in \mathscr{F}_B$.

- If $c \cap B = \varnothing$, then we may take c as a right-angled Coxeter cubical cell for Q. Such a c corresponds to a codimension-k face in P_Q which is determined by k facets in $\mathscr{F}(P_Q) - \mathscr{F}_B$, so c is of the form $e^k/(\mathbb{Z}_2)^k$.
- If c is a k-cube in $\mathscr{C}(B^+) \subset \mathscr{C}(P_Q)$, then there is also another k-cube $c' \in \mathscr{C}(B^-) \subset \mathscr{C}(P_Q)$. Both c and c' are codimension-one faces of two $(k+1)$-cubes in $\mathscr{C}(P_Q)$. Gluing those two $(k+1)$-cubes by identifying c with c', we obtain a right-angled Coxeter cubical cell with form $e^{k+1}/(\mathbb{Z}_2)^k$.

Finally, we obtain a right-angled Coxeter cellular decomposition of Q, denoted by $\mathscr{C}(Q)$, whose cells are right-angled Coxeter cubes. Of particular note is that $\mathscr{C}(Q)$ is not cubical. This is because there exists the cubical cell glued by two cells c and c' in $\mathscr{C}(P_Q)$ as above, which has a self-intersection, namely the cone point x_0, as shown in Figure 8. The cone point is the only 0-cell in $\mathscr{C}(Q)$, which will be chosen as the basepoint when we calculate the orbifold fundamental group of Q.

4.2 The orbifold fundamental groups of simple handlebodies

Following the above notations, by Proposition 2.9, we can directly write out a presentation of orbifold fundamental group of Q.

Proposition 4.1 *Let Q be a simple handlebody of genus $\mathfrak{g}$, and P_Q be the associated simple polytope with copies of cutting belts $\mathscr{F}_B$. Then $\pi_1^{\mathrm{orb}}(Q)$ has a presentation with generators s_F indexed by $F \in \mathscr{F}(P_Q)$, satisfying the relations*

(1) $s_F^2 = 1$ *for* $F \in \mathscr{F}(P_Q) - \mathscr{F}_B$,

(2) $t_{B^+}t_{B^-} = 1$ *for two* B^+ *and* B^- *with* $B^+ \sim B^-$ *in* $\mathscr{F}_B$,

(3) $(s_F s_{F'})^2 = 1$ *for* $F, F' \in \mathcal{F}(P_Q) - \mathcal{F}_B$ *with* $F \cap F' \neq \varnothing$,

(4) $s_F t_{B^+} = t_{B^+} s_{F'}$ *for* $B^+ \sim B^-$ *in* $\mathcal{F}_B$ *and* $F, F' \in \mathcal{F}(P_Q) - \mathcal{F}_B$ *with* $F \cap B^+ \sim F' \cap B^-$,

where the basepoint of $\pi_1^{\mathrm{orb}}(Q)$ *is the cone point* x_0 *in the interior of* Q.

On the other hand, we show here that $\pi_1^{\mathrm{orb}}(Q)$ is actually an iterative HNN-extension on $W(P_Q, \mathcal{F}_B)$, where $W(P_Q, \mathcal{F}_B)$ is a right-angled Coxeter group determined by facial structure of P_Q by ignoring the facets in $\mathcal{F}_B$:

$$W(P_Q, \mathcal{F}_B) = \langle s_F,\, F \in \mathcal{F}(P_Q) - \mathcal{F}_B \mid s_F^2 = 1 \text{ for all } F;\, (s_F s_{F'})^2 = 1 \text{ for } F \cap F' \neq \varnothing \rangle,$$

which can be regarded as the orbifold fundamental group of P_Q as a right-angled Coxeter orbifold with boundary consisting of the disjoint union of all facets in $\mathcal{F}_B$.

Let B be a cutting belt in Q, and $B^+, B^- \in \mathcal{F}_B$ be two copies of B. Set

$$\mathcal{F}^{B^+} = \{F \in \mathcal{F}(P_Q) - \mathcal{F}_B \mid F \cap B^+ \neq \varnothing\},$$
$$\mathcal{F}^{B^-} = \{F \in \mathcal{F}(P_Q) - \mathcal{F}_B \mid F \cap B^- \neq \varnothing\}.$$

The associated right-angled Coxeter groups W_{B^+} and W_{B^-} are isomorphic since B^+ and B^- are combinatorially equivalent as simple polytopes.

Lemma 4.2 *The maps* $i_{B^+}\colon W_{B^+} \to W(P_Q, \mathcal{F}_B)$ *and* $i_{B^-}\colon W_{B^-} \to W(P_Q, \mathcal{F}_B)$ *induced by inclusions* $B^+ \hookrightarrow P_Q$ *and* $B^- \hookrightarrow P_Q$ *are monomorphisms.*

Proof According to the definition of B-belt, i_{B^+} and i_{B^-} are obviously well defined. There are two group homomorphisms $j_{B^+}\colon W(P_Q, \mathcal{F}_B) \to W_{B^+}$ and $j_{B^-}\colon W(P_Q, \mathcal{F}_B) \to W_{B^-}$, defined by reducing modulo the normal subgroups generated by facets being not in $\mathcal{F}^{B^+}$ and $\mathcal{F}^{B^-}$, such that $j_{B^+} \circ i_{B^+} = \mathrm{id}_{W_{B^+}}$ and $j_{B^-} \circ i_{B^-} = \mathrm{id}_{W_{B^-}}$. The result follows from this. □

Thus, W_{B^+} and W_{B^-} can also be regarded as two isomorphic subgroups of $W(P_Q, \mathcal{F}_B)$ generated by $\{s_F,\, F \in \mathcal{F}^{B^+}\}$ and $\{s_{F'},\, F' \in \mathcal{F}^{B^-}\}$, respectively. Define $\phi_B\colon W_{B^-} \to W_{B^+}$ by $\phi_B(s_{F'}) = s_F$ with $F' \cap B^- \sim F \cap B^+$. Then ϕ_B is a well-defined isomorphism. Furthermore, attaching two facets on P_Q corresponding to the cutting belt B is equivalent to doing one HNN-extension on its orbifold fundamental group $\pi_1^{\mathrm{orb}}(P_Q) = W(P_Q, \mathcal{F}_B)$, giving new elements t_{B^+} and t_{B^-} with certain conditions in $\pi_1^{\mathrm{orb}}(Q)$. By doing an induction on the genus of Q and repeating the use of HNN-extension, the orbifold fundamental group of Q is isomorphic to doing $\mathfrak{g}$ HNN-extensions on the right-angled Coxeter group $W(P_Q, \mathcal{F}_B)$,

$$(Q_{\mathfrak{g}}, B_{\mathfrak{g}}) \longrightarrow \cdots \longrightarrow (Q_1, B_1) \longrightarrow Q_0 = P_Q$$

$$\xrightarrow{\text{cutting}}$$
$$\xleftarrow[\text{HNN-extension}]{}$$

$$G_{\mathfrak{g}} = \pi_1^{\mathrm{orb}}(Q) \longleftarrow \cdots \longleftarrow G_1 \longleftarrow G_0 = W(P_Q, \mathcal{F}_B)$$

where each Q_k is the simple handlebody of genus k obtained from Q_{k+1} by cutting open along the $(k+1)^{\mathrm{st}}$ belt B_{k+1}, which is a right-angled Coxeter orbifold with boundary consisting of double copies of $\{B_{k+1},\dots,B_{\mathfrak{g}}\}$, and each G_k is the orbifold fundamental group of Q_k which is obtained from an HNN-extension on G_{k-1}.

Proposition 4.3 *Let Q be a simple handlebody of genus $\mathfrak{g}$ with cutting belts $B_1,\dots,B_{\mathfrak{g}}$. Then*

$$\pi_1^{\mathrm{orb}}(Q)\cong(\cdots((W(P_Q,\mathcal{F}_B)*_{\phi_{B_1}})*_{\phi_{B_2}})\cdots)*_{\phi_{B_{\mathfrak{g}}}}.$$

Notice that the expression $(\cdots((W(P_Q,\mathcal{F}_B)*_{\phi_{B_1}})*_{\phi_{B_2}})\cdots)*_{\phi_{B_{\mathfrak{g}}}}$ in Proposition 4.3 is independent of orders of ϕ_{B_i}. In addition, the presentation of $\pi_1^{\mathrm{orb}}(Q)$ in Proposition 4.1 can be simplified by deleting all generators t_{B^-} and relations $t_{B^+}t_{B^-}=1$, replaced by only all t_B. It should be pointed out that the right-angled Coxeter group W_Q determined by the facial structure of Q is not a subgroup of $\pi_1^{\mathrm{orb}}(Q)$ in general. Actually, W_Q is the quotient group of $\pi_1^{\mathrm{orb}}(Q)$ with respect to the normal group generated by all t_B.

Remark 4.4 In [25, Theorem 4.7.2], Davis, Januszkiewicz and Scott give a similar form. However, all generators in their paper lifted into the universal space as homeomorphisms onto itself are involutions, i.e. $t_B^2=1$. Here, with a little difference, we require that the lifted action of t_B is free. In particular, the last relation in Proposition 4.1 belongs to a kind of *Baumslag–Solitar relations*, which are related to the HNN-extension. In other words, pasting pairs of facets corresponding to cutting belts of the polytope P_Q can be viewed as a topological explanation for the HNN-extension of their orbifold fundamental groups. More precisely, for a cutting belt B, there are two copies B^+ and B^- in P_Q, and the composite map

$$W_B\cong W_{B^+}\xrightarrow{i_{B^+}}W(P_Q,\mathcal{F}_B)\xrightarrow{i_1}G_1\xrightarrow{i_2}\cdots\xrightarrow{i_g}G_{\mathfrak{g}}=\pi_1^{\mathrm{orb}}(Q)$$

embeds W_B into $\pi_1^{\mathrm{orb}}(Q)$, where i_k is defined by $i_k(h)=h\in G_k$ for $h\in G_{k-1}$. Both W_{B^+} and W_{B^-} are linked in $\pi_1^{\mathrm{orb}}(Q)$ by an isomorphism and the injectivity of i_k is followed by the normal form theorem of HNN-extension (see Theorem 2.22).

4.3 The orbifold universal covers of simple handlebodies

Let Q be a simple handlebody with cutting belts $\{B_1,\dots,B_{\mathfrak{g}}\}$, and P_Q be the simple polytope given by cutting Q with the quotient map $q\colon P_Q\to Q$. Let $\pi_1^{\mathrm{orb}}(Q)$ be the orbifold fundamental group with the presentation in Proposition 4.1. Define a *characteristic map* on the facet set of P_Q,

$$\lambda\colon\mathcal{F}(P_Q)\to\pi_1^{\mathrm{orb}}(Q),$$

given by $\lambda(F)=s_F$ for $F\in\mathcal{F}(P_Q)-\mathcal{F}_B$, and $\lambda(B)=t_B$ for $B\in\mathcal{F}_B$. Then we construct the space

$$\widetilde{Q}=P_Q\times\pi_1^{\mathrm{orb}}(Q)/\sim,\tag{4-2}$$

where $(x,g)\sim(y,h)$ if and only if either

- $x=y\in F\in\mathcal{F}(P_Q)-\mathcal{F}_B$ and $gs_F=h$, or
- $(x,y)\in(B,B')$, $B,B'\in\mathcal{F}_B$, $q(x)=q(y)$ and $t_B\cdot g=h$.

The orbit space of the action of $\pi_1^{\mathrm{orb}}(Q)$ on $\widetilde{Q}$ is Q, so the polytope P_Q can be viewed as the fundamental domain of $\pi_1^{\mathrm{orb}}(Q)$ acting on $\widetilde{Q}$. According to Corollary 2.18:

Lemma 4.5 $\widetilde{Q}$ *is the universal orbifold cover of* Q*.*

5 Proof of Theorem A

This section is devoted to giving the proof of Theorem A.

5.1 Nonpositively curved cubical complex

A geodesic metric space X is *nonpositively curved* if it is a locally CAT(0) space. The Cartan–Hadamard theorem implies that nonpositively curved spaces are aspherical; see [7; 18; 32].

Definition 5.1 (the links in a cubical complex, see [7, Section 7.15] or [18, page 508]) Let K be a cubical complex. For each vertex $v \in K$, its (*geometric*) *link*, denoted by $\mathrm{Lk}(v)$, is a simplicial complex defined by all cubes in K that properly contain v with respect to the inclusion. A d-cube c of K that properly contains v determines a $(d-1)$-simplex $s(c)$ in $\mathrm{Lk}(v)$.

Proposition 5.2 (Gromov lemma; see [32] or [18, Corollary I.6.3]) *A piecewise Euclidean cubical complex is nonpositively curved if and only if the link of its each vertex is a flag complex.*

5.2 Homology groups of manifold covers over simple handlebodies

Let M_Q and $\widetilde{Q}$ be the manifold double and orbifold universal cover over a simple handlebody Q with m facets, respectively. In this subsection, we discuss the homology groups of M_Q and $\widetilde{Q}$.

5.2.1 Homology groups of M_Q By Davis' result (see [18, Theorem 8.12]),

$$H_*(M_Q) \cong \bigoplus_{g\in(\mathbb{Z}_2)^m} H_*(|Q|, \mathscr{F}_g) \tag{5-1}$$

where $\mathscr{F}_g = \bigcup_{s_i\in S(g)} F_i \subset \partial|Q|$, $F_i \in \mathscr{F}(Q)$ and $S(g) = \{s_i \mid l(s_i \cdot g) = l(g)-1\}$ for a reduced word g of length $l(g)$ in $(\mathbb{Z}_2)^m$. If Q is a simple handlebody of genus $\mathfrak{g} \geq 0$, then $|Q| \simeq \bigvee_{\mathfrak{g}} S^1$. By the long exact sequence of homology groups of $(|Q|, \mathscr{F}_g)$, if $* \geq 3$, then

$$H_*(|Q|, \mathscr{F}_g) \cong H_{*-1}(\mathscr{F}_g) \cong H_{*-1}(K_g)$$

where $K_g \simeq \mathscr{F}_g$ is the dual simplicial complex of $\mathscr{F}_g$, which is a subcomplex of $\mathcal{N}(Q)$. Hence for $* \geq 3$,

$$H_*(M_Q) \cong \bigoplus_{J\subset\mathscr{F}(Q)} H_{*-1}(K_J)$$

where J is the set of those facets F_i corresponding to all $s_i \in S(g)$.

For $*=1,2$, we have

$$0 \to H_2(|Q|, \mathcal{F}_g) \to H_1(\mathcal{F}_g) \xrightarrow{(i_g)_*} H_1(|Q|) \cong \mathbb{Z}^{\mathfrak{g}} \to H_1(|Q|, \mathcal{F}_g) \to 0$$

where $i_g : \mathcal{F}_g \to |Q|$ is an inclusion. Then

$$H_1(M_Q) \cong \bigoplus_{g \in (\mathbb{Z}_2)^m} \operatorname{coker}(i_g)_* \quad \text{and} \quad H_2(M_Q) \cong \bigoplus_{g \in (\mathbb{Z}_2)^m} \ker(i_g)_*.$$

Remark 5.3 The formula (5-1) is actually Hochster's formula in the setting of simple handlebodies. When Q is a simple polytope P, Hochster's formula [9, Proposition 3.2.11] can also be expressed as

$$H^{l(g)-i-1}(\mathcal{F}_g) \cong \operatorname{Tor}^{-i,l(g)}_{\mathbb{Z}[v_1,\dots,v_m]}(\mathcal{K}(P), \mathbb{Z})$$

where $\mathcal{K}(P)$ is the Stanley–Reisner face ring of P.

5.2.2 Homology groups of $\widetilde{Q}$ Davis' [18, Theorem 8.12] cannot be directly applied to give the homology groups of $\widetilde{Q}$ since $\pi_1^{\mathrm{orb}}(Q)$ is not a Coxeter group when the genus of Q is more than zero. However, we can employ the method of Davis in [18, Chapter 8] to calculate of the homology groups of $\widetilde{Q}$.

Let Q be a simple handlebody with nerve $\mathcal{N}(Q)$, and P_Q be the associated simple polytope. Let $G = \pi_1^{\mathrm{orb}}(Q)$ be the orbifold fundamental group of Q. We have known that G is an iterative HNN-extension on a right-angled Coxeter group $W(P_Q, \mathcal{F}_B)$. Namely,

$$G = \pi_1^{\mathrm{orb}}(Q) \cong (\cdots((W(P_Q, \mathcal{F}_B) *_{\phi_{B_1}}) *_{\phi_{B_2}}) \cdots) *_{\phi_{B_{\mathfrak{g}}}}$$

where $\mathfrak{g}$ is the genus of Q. For any $w \in G$, consider the reduced normal form

$$w = g_0 t_1 g_1 \cdots g_{m-1} t_m g_m$$

where each g_i is reduced in $W(P_Q, \mathcal{F}_B)$, and each t_i is one of $\{t_B^{\pm 1}\}$ which determines an isomorphism of $\{\phi_B^{\pm 1}\}$ on some subgroups of $\pi_1^{\mathrm{orb}}(Q)$. Denote the generator set of G by

$$\mathcal{S} = \{s_F ; F \in \mathcal{F}(P_Q) - \mathcal{F}_B\} \cup \{t_B ; B \in \mathcal{F}_B\}.$$

For any word $w \in G$, put

$$S(w) = \{s \in \mathcal{S} \mid l(ws) < l(w)\},$$

where $l(w)$ is the word length of the reduced normal form of w in G (i.e. the shortest length between 1 and w in the Cayley graph of G associated with the generator set $\mathcal{S}$). For each subset T of $\mathcal{S}$, let P_Q^T be the subcomplex of P_Q defined by

$$P_Q^T = \bigcup_{t \in T} F_t,$$

where $F_{s_F} = F$ for $s_F \in \mathcal{F}(P_Q) - \mathcal{F}_B$ and $F_{t_B} = B'$ for $B \in \mathcal{F}_B$ with $B \sim B'$.

Let $\widetilde{Q} = P_Q \times G/\sim$ be the universal cover of Q defined as in (4-2). Then we have the following conclusion which generalizes Davis' theorem in [18, Theorem 8.12].

Proposition 5.4 *The homology of $\widetilde{Q}$ is isomorphic to the direct sum*

$$H_*(\widetilde{Q}) \cong \bigoplus_{w \in G} H_*(P_Q, P_Q^{S(w)}),$$

where $G = \pi_1^{\mathrm{orb}}(Q)$ has the presentation in Proposition 4.1.

Remark 5.5 It should be emphasized that here P_Q is not used as a mirrored space in the sense of Davis in [18] although it is a simple polytope. Actually, here we just put P_Q and $\pi_1^{\mathrm{orb}}(Q)$ together to construct the orbifold universal cover $\widetilde{Q}$, but $\pi_1^{\mathrm{orb}}(Q)$ is not a Coxeter group except that the genus of Q is zero.

Corollary 5.6 *If there is an empty k-simplex $\triangle^k$ in $\mathcal{N}(P_Q)$, then $H_k(\widetilde{Q}) \neq 0$.*

Proof Assume that the vertex set of a empty k-simplex $\triangle^k$ in $\mathcal{N}(P_Q)$ is

$$T = \{F_1, F_2, \dots, F_{k+1}\}$$

which does not contain the facet in $\mathcal{F}_B$ (in fact, any facet in $\mathcal{F}_B$ is not the vertex of any empty simplex of $\mathcal{N}(P_Q)$; this is guaranteed by the definition of B-belt). Let $w = s_1 s_2 \cdots s_{k+1}$. Regard T as $\{s_1, \dots, s_{k+1}\}$. Then $S(w) = T$. Moreover, $P_Q^{S(w)} = P_Q^T = \bigcup_{i=1}^{k+1} F_i \simeq \partial\triangle^k \simeq S^{k-1}$. Since P_Q is a contractible ball, by the long exact homology group sequence of pair (P_Q, P_Q^T), we have

$$H_k(P_Q, P_Q^T) \cong H_{k-1}(P_Q^T) \cong H_{k-1}(S^{k-1}) \neq 0.$$

Therefore, by Proposition 5.4, $H_k(\widetilde{Q}) \neq 0$. □

5.2.3 Proof of Proposition 5.4 Before we prove Proposition 5.4, we first give some notation (see [18]).

A subset T of $\mathcal{S}$ is called *spherical* if the subgroup generated by T is a finite subgroup of G. Each s_F in a spherical subset T exactly corresponds to a facet $F \in \mathcal{F}(P) - \mathcal{F}_B$, and $F \cap F' \neq \varnothing$ for any s_F and $s_{F'}$ in a spherical set T. Let W_T be the group generated by a spherical subset T. Then $W_T \cong (\mathbb{Z}_2)^{\#T}$, where $\#T$ denotes the number of all elements in T.

If the set T is the union of a spherical set T_S and a t_B for $B \in \mathcal{F}_B$, then

$$W_T = W_{T_S} \cup t_{B'} W_{T_S},$$

where B' is the facet which is identified with B in Q.

Lemma 5.7 *Let G be the orbifold fundamental group of a simple handlebody with generator set $\mathcal{S}$. Then for each $w \in G$, $S(w)$ is either a spherical subset of $\mathcal{S}$ or the union of a t_B and a spherical subset in $\mathcal{S}$.*

Proof Let $w = g_0 t_1 g_1 \cdots g_{m-1} t_m g_m$ be a reduced normal form in G. We might as well assume that this expression of w is a normal form in the opposite direction for each t_B, that is, each g_i is a representative of a coset of $W_{B_{i+1}}$ or $W_{B'_{i+1}}$ in G, for $i = 0, \dots, m-1$.

It is easy to see that for $F \in \mathscr{F}(P) - \mathscr{F}_B$, $s_F \in S(w)$ if and only if $s_F \in S(t_m g_m)$. If there is a $B \in \mathscr{F}_B$ such that $t_B \in S(w)$, then $g_m t_B = t_B g'_m$ where $g'_m = \phi_B(g_m)$, and the last t_m is t_B^{-1}. For another $t_{B'} \neq t_B$, it cannot reduce the length of w. Thus the conclusion holds. □

For a spherical set $T = S(w)$, we define an element in $\mathbb{Z}W_T \subset \mathbb{Z}W(P_Q, \mathscr{F}_B)$ by the formula

$$\beta_T = \sum_{w \in W_T} (-1)^{l(w)} w.$$

Consider a natural cellular decomposition of P_Q given by its facial structure. Let $C_*(P_Q)$ and $C_*(\widetilde{Q})$ denote the cellular chain complexes of P_Q and $\widetilde{Q}$, respectively, and let $H_*(P_Q)$ and $H_*(\widetilde{Q})$ be their respective homology groups. Since G acts cellularly on $\widetilde{Q}$, $C_*(\widetilde{Q})$ is a $\mathbb{Z}(G)$-module.

Let T be a spherical set. Multiplication by β_T defines a homomorphism $\beta_T : C_*(P_Q) \to C_*(W_T P_Q)$.

Lemma 5.8 *$C_*(P_Q^T)$ is contained in the kernel of $\beta_T : C_*(P_Q) \to C_*(W_T P_Q)$.*

Proof Suppose τ is a cell in P_Q^T. If T is a spherical set, then τ lies in some $F \in \mathscr{F}(P_Q) - \mathscr{F}_B$ such that $s_F \in T$. Let $\mathscr{B}$ be a subset of W_T such that $W_T = \mathscr{B} \cup s_F \mathscr{B}$; then we can write β_T as

$$\beta_T = \sum_{w \in W_T} (-1)^{l(w)} w = \sum_{v \in \mathscr{B}} (-1)^{l(v)} (v - v s_F).$$

Since vF is identified with $v s_F F$ in $\widetilde{Q}$, we have that

$$\beta_T \tau = \sum (-1)^{l(v)} (v - v s_F)\tau = \sum (-1)^{l(v)} (v\tau - v\tau) = 0.$$

Thus, $C_*(P_Q^T) \subset \ker \beta_T$. □

Hence, β_T induces a chain map $C_*(P_Q, P_Q^T) \to C_*(W_T P_Q)$, still denoted by β_T.

For each $w \in G$:

- If $T = S(w)$ is a spherical set, we then define a map
$$\rho^w = w\beta_T : C_*(P_Q, P_Q^T) \xrightarrow{\beta_T} C_*(W_T P_Q) \xrightarrow{w} C_*(wW_T P_Q).$$
Hence, we have a map
$$\rho^w_* : H_*(P_Q, P_Q^T) \to H_*(wW_T P_Q).$$
- If $T = S(w) = \{t_B\} \cup T_S$ where T_S is a spherical set, $t_B s = s t_B$ for any $s \in T_S$ implies that $W_{T_S} < W_B$, i.e. $B \cap F_s \neq \varnothing$ for any $s \in T_S$. So B' does not intersect any F_s; hence for $k > 1$ we have
$$H_k(P_Q, P_Q^T) \cong H_{k-1}(P_Q^T) \cong H_{k-1}\Big(P_Q^{T_S} \coprod B'\Big) \cong H_{k-1}(P_Q^{T_S}) \cong H_k(P_Q, P_Q^{T_S})$$
where P_Q and B' are contractible simple polytopes. Now put
$$\rho^w_* : H_k(P_Q, P_Q^T) \cong H_k(P_Q, P_Q^{T_S}) \xrightarrow{\beta_{T_S}} H_*(W_{T_S} P_Q) \xrightarrow{i_*} H_*(W_T P_Q) \xrightarrow{\times w} H_*(wW_T P_Q).$$

Next, order the elements of G,

$$w_1, w_2, \ldots$$

so that $l(w_i) \le l(w_{i+1})$. For each $n \ge 1$, put

$$X_n = \bigcup_{i=1}^{n} w_i P_Q.$$

To simplify notation, set $w = w_n$.

Lemma 5.9 $$X_{n-1} \cap wP = wP^{S(w)}.$$

Proof Notice that X_{n-1} contains a subgraph of Cayley graph of G associated with the generator set $\mathcal{S}$, where the length between each vertex and the unit element is less than or equal to $l(w)$. Then

$$l(ws) = \begin{cases} l(w) - 1 & \text{if } s \in S(w), \\ l(w) + 1 & \text{if } s \in \mathcal{S} - S(w). \end{cases}$$

A chamber $w_i P_Q$ $(i < n)$ in X_{n-1} intersects with wP_Q in the facet wF if and only if either $w_i \cdot s_F = w$ for $F \in \mathcal{F}(P) - \mathcal{F}_B$ or $w_i \cdot t_F^{-1} = w$ for $F \in \mathcal{F}_B$ where t_F is a torsion-free generator in $\mathcal{S}$; in other words, either $l(ws_F) = l(w) - 1$ or $l(wt_F) = l(w) - 1$. Therefore, $X_{n-1} \cap wP_Q = wP_Q^{S(w)}$. □

Finally let us finish the proof of Proposition 5.4.

Proof of Proposition 5.4 We know from Lemma 5.9 that $X_{n-1} \cap wP_Q = wP_Q^{S(w)}$. Hence, the excision theorem gives an isomorphism

$$H_*(X_n, X_{n-1}) \xrightarrow{\cong} H_*(wP_Q, wP_Q^{S(w)}).$$

Consider the exact sequence of the pair (X_n, X_{n-1}),

$$\cdots \to H_*(X_{n-1}) \xrightarrow{j_*} H_*(X_n) \xrightarrow{k_*} H_*(X_n, X_{n-1}) \to \cdots.$$

We claim that the map k_* is a split epimorphism, which is equivalent to the map

$$k_*^w \colon H_*(X_n) \to H_*(P_Q, P_Q^{S(w)})$$

being a split epimorphism, where k_*^w denotes the composition of k_* with the excision isomorphism and left translation by w^{-1}. Consider the map ρ_*^w on $H_*(P_Q, P_Q^{S(w)})$ whose image is contained in $H_*(wW_{S(w)}P_Q)$. For every $v \neq 1$ in $W_{S(w)}$, we have $l(wv) < l(w)$; hence, $wW_{S(w)}P_Q \subset X_n$. Hence the image of ρ_*^w is contained in $H_*(X_n)$. All these can be seen from the commutative diagram

$$\begin{tikzcd}
H_*(X_n, X_{n-1}) \arrow[r, "\cong"] & H_*(wP_Q, wP_Q^{S(w)}) \arrow[d, "\times w^{-1}"] \\
H_*(X_n) \arrow[u, "k_*"] \arrow[r, shift left, "k_*^w"] & H_*(P_Q, P_Q^{S(w)}) \arrow[l, shift left, "\rho_*^w"] \arrow[d, "\beta_*"] \\
H_*(wW_{S(w)}P_Q) \arrow[u, "i_*"] & H_*(W_{S(w)}P_Q) \arrow[l, "\times w"']
\end{tikzcd}$$

where β_* is induced by multiplication by $\beta_{S(w)}$ when $S(w)$ is a spherical set, and is the composition

$$\beta_{T_S}\circ i_*\colon H_k(P_Q,P_Q^T)\cong H_k(P_Q,P_Q^{T_S})\xrightarrow{\beta_{T_S}} H_*(W_{T_S}P_Q)\xrightarrow{i_*} H_*(W_T P_Q)$$

when $S(w)$ is the union of a $\{t_B\}$ and a spherical set T_S.

Since $\widetilde{Q}$ is the universal cover of Q, $H_1(\widetilde{Q})\cong 0$. For $*>1$, it can be see that $k_*^w\circ\rho_*^w$ is the identity on $H_*(P_Q,P_Q^{S(w)})$ by above diagram. Hence there is the splitting short exact sequence

$$0\to H_*(X_{n-1})\xrightarrow{j_*} H_*(X_n)\xrightarrow{k_*^w} H_*(P_Q,P_Q^{S(w)})\to 0.$$

This implies that

$$H_*(X_n)\cong H_*(X_{n-1})\oplus H_*(P_Q,P_Q^{S(w)})$$

where $H_*(X_1)=H_*(P_Q)=0$. Since $\widetilde{Q}$ is the increasing union of the X_n, we have

$$H_*(\widetilde{Q})=\lim_{n\to\infty}H_*(X_n)\cong\bigoplus_{w\in G}H_*(P_Q,P_Q^{S(w)}).\qquad\square$$

5.3 Proof of Theorem A

Let Q be a simple handlebody and $q\colon P_Q\to Q$ be the quotient map by gluing all paired facets in $\mathcal{F}_B$. Then the orbifold universal cover $\pi\colon\widetilde{Q}\to Q$ of Q can be constructed by (4-2).

Let $\mathscr{C}(P_Q)$ be the standard cubical cellular decomposition of P_Q. For each cube $c\in\mathscr{C}(P_Q)$, each component of $\pi^{-1}(c)$ is a cube in $\widetilde{Q}$. Then $\mathscr{C}(P_Q)$ determines a cubical cellular decomposition of $\widetilde{Q}$, denoted by $\mathscr{C}(\widetilde{Q})$, such that the link of each point v in $\mathscr{C}(\widetilde{Q})$ is exactly the nerve $\mathcal{N}(P_Q)$ of P_Q. Hence, if Q is flag, then P_Q is flag, and so is $\mathcal{N}(P_Q)$. By the Gromov lemma, $\widetilde{Q}$ is nonpositively curved. In fact, $\widetilde{Q}$ is a CAT(0) space. Then by the Cartan–Hadamard theorem, $\widetilde{Q}$ is aspherical. Therefore, Q is orbifold-aspherical.

On the contrary, if Q is orbifold-aspherical, then $\widetilde{Q}$ is contractible. Using an idea of Davis in [20, Section 8.2], we shall show that if P_Q is not flag then $\widetilde{Q}$ is not contractible. Indeed, if P_Q is not flag, then $\mathcal{N}(P_Q)$ contains an empty k-simplex for some $k\geq 2$. The dual of this empty k-simplex gives an essential embedding sphere in $\widetilde{Q}$. By Corollary 5.6, the fundamental class of such a sphere is nontrivial in $H_k(\widetilde{Q})$, which contradicts that $\widetilde{Q}$ is contractible. $\square$

6 Proof of Theorem B

The purpose of this section is to characterize the rank two free abelian subgroup $\mathbb{Z}\oplus\mathbb{Z}$ in $\pi_1^{\mathrm{orb}}(Q)$ in terms of a $\square$-belt in Q. Here $\pi_1^{\mathrm{orb}}(Q)$ is an iterative HNN-extension over a right-angled Coxeter group.

Theorem B *Suppose that Q is a simple n-handlebody. Then there is a rank two free abelian subgroup $\mathbb{Z}\oplus\mathbb{Z}$ in $\pi_1^{\mathrm{orb}}(Q)$ if and only if Q contains a $\square$-belt.*

Remark 6.1 The "simple" condition of a handlebody is necessary in above proposition. In fact, it is easy to see that the orbifold fundamental group of a two-dimensional annulus as a right-angled Coxeter orbifold is isomorphic to $\mathbb{Z} \oplus (\mathbb{Z}_2 * \mathbb{Z}_2)$, which contains a rank two free abelian subgroup $\mathbb{Z} \oplus \mathbb{Z}$. Consider a right-angled Coxeter 3-handlebody Q with a π_1-injective annulus-suborbifold B such that B is a π_1-injective suborbifold; it provides a subgroup $\mathbb{Z} \oplus \mathbb{Z}$ in its orbifold fundamental group. Of course, such a Q is not simple. All of these results are the generalization of [7, Lemma 5.22] which is related to the flat torus theorem in [7, Chapter II.7].

Example 6.2 (squares of Example 3.11) We show that each $\square$ in (c) and (d) of Example 3.11 determines a subgroup $\mathbb{Z} \oplus \mathbb{Z}$ in $\pi_1^{\text{orb}}(Q)$, whereas cases (a) and (b) do not.

In (a), the four facets F_1, F_2, F_3 and F_4 correspond to a suborbifold B which is a quadrilateral in Q^*, but it is not a $\square$-belt in Q. In fact,

$$i_*(\pi_1^{\text{orb}}(B)) \cong W_\square / \langle (s_1 s_3)^2 \rangle \cong (\mathbb{Z}_2)^2 \oplus (\mathbb{Z}_2 * \mathbb{Z}_2) < \pi_1^{\text{orb}}(Q)$$

and s_1s_3, s_2s_4 generate a subgroup $\mathbb{Z}_2 \oplus \mathbb{Z}$ in $i_*(\pi_1^{\text{orb}}(B)) < \pi_1^{\text{orb}}(Q)$, where $i_* \colon \pi_1^{\text{orb}}(B) \to \pi_1^{\text{orb}}(Q)$ is induced by the inclusion $i \colon B \hookrightarrow Q$. Thus, there is no subgroup $\mathbb{Z} \oplus \mathbb{Z}$ in $i_*(\pi_1^{\text{orb}}(B))$.

In (b), $\{F_1, F_2, F_3, F_4\}$ does not determine a quadrilateral suborbifold. Without loss of generality, assume that $\{F_1, F_2, F_3, F_4\}$ bounds only one hole of Q^*. Then there are at least 5 generators in $\pi_1^{\text{orb}}(Q)$ associated to five facets in P_Q, denoted by $\{F_1, F_2, F_3, F_4, F_1'\}$ with $F_1 \cap B^+ \sim F_1' \cap B^-$, where B is the cutting belt of Q and cut F_1 into two facets in P_Q. Thus, (b) induces a subgroup of $\pi_1^{\text{orb}}(Q)$ as follows:

$$\begin{aligned} W_b &:= \langle s_1, s_2, s_3, s_4, s_1', t \mid (s_i)^2 = 1, \forall i; (s_1s_2)^2 = (s_2s_3)^2 = (s_3s_4)^2 = (s_4s_1')^2 = 1; s_1' = ts_1t \rangle \\ &= \langle s_1, s_2, s_3, s_4, t \mid (s_i)^2 = 1, \forall i; (s_1s_2)^2 = (s_2s_3)^2 = (s_3s_4)^2 = (s_4ts_1t^{-1})^2 = 1 \rangle \end{aligned}$$

which contains no subgroup $\mathbb{Z} \oplus \mathbb{Z}$.

In (c) or (d), $\{F_1, F_2, F_3, F_4\}$ determines a $\square$-belt $B_\square$ of Q. If $B_\square$ does not intersect with any cutting belt, then $B_\square$ is kept in P_Q, so there is a subgroup $\mathbb{Z} \oplus \mathbb{Z} < W(P_Q, \mathscr{F}_B) < \pi_1^{\text{orb}}(Q)$. If there are some cutting belts $B_1, B_2, \dots, B_k$ intersecting transversely with only a pair of disjoint edges of $B_\square$, without loss of generality, assume that $B_1, B_2, \dots, B_k$ intersect with two disjoint edges f_1 and f_3 of $B_\square$, where some cutting belts may cut f_1 and f_3 many times; see (c) in Figure 7. Then there is also a subgroup $\mathbb{Z} \oplus \mathbb{Z}$ generated by s_1s_3 and $s_2t_1t_2 \cdots t_ks_4t_k^{-1} \cdots t_2^{-1}t_1^{-1}$ where each t_i is one of $\{t_B^{\pm 1}\}$. Also see Figure 9.

6.1 The special case where Q is a simple polytope

When Q is a simple polytope, Theorem B can be followed by Moussong's theorem; see [46] or [18, Corollary 12.6.3].

Theorem 6.3 (Moussong's theorem; see [46] or [18, Corollary 12.6.3]) *If W_P is the right-angled Coxeter group of a simple polytope P, then there exists a $\mathbb{Z} \oplus \mathbb{Z}$ subgroup in W_P if and only if there is a $\square$-belt in P.*

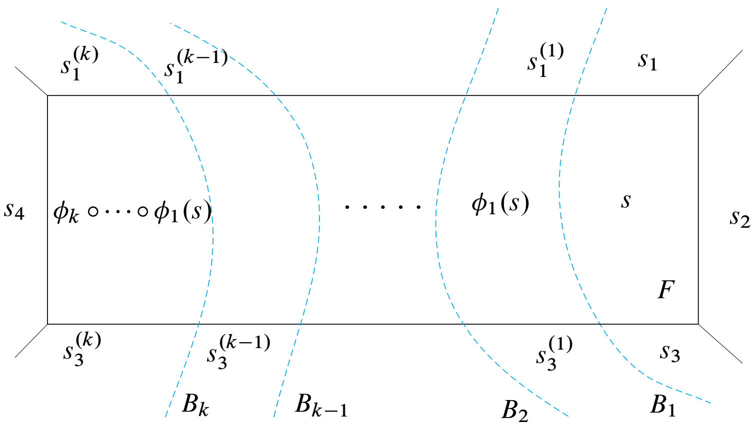


Figure 9: $\square$-belt and $\mathbb{Z}\oplus\mathbb{Z}$.

Next let us deal with the case of a simple handlebody. Let Q be a simple handlebody of genus $\mathfrak{g}>0$, and P_Q be the associated simple polytope obtained by cutting Q open along cutting belts $\{B_i \mid i=1,2,\dots,\mathfrak{g}\}$.

6.2 Proof of the sufficiency of Theorem B

Assume that there is a $\square$-belt $B_\square$ given by $\{F_1,F_2,F_3,F_4\}$ in $\mathcal{N}(Q)$. After cutting Q open along cutting belts B_i, $i=1,2,\dots,\mathfrak{g}$, by Lemma 3.12, there are the following two cases.

- The $B_\square$ is still kept in P_Q. Then $B_\square$ gives a subgroup $\mathbb{Z}\oplus\mathbb{Z}$ in $W(P_Q,\mathcal{F}_B)<\cdots<\pi_1^{\mathrm{orb}}(Q)$, which is generated by s_1s_3 and s_2s_4.
- The $B_\square$ is not kept in P_Q. Then there is only one situation in which some cutting belts B_i intersect transversely with a pair of disjoint edges of $B_\square$, say F_1 and F_3. If $B_\square$ intersects transversely with cutting belts $B_1,B_2,\dots,B_k$ in turn, then s_1s_3 and $s_2t_1\cdots t_k s_4 t_k^{-1}\cdots t_1^{-1}$ generate a subgroup $\mathbb{Z}\oplus\mathbb{Z}$ in $\pi_1^{\mathrm{orb}}(Q)$, as in cases (c) or (b) on Example 3.11. See also Example 6.2. $\square$

6.3 Proof of the necessity of Theorem B

Cutting Q open along a cutting belt B, we get a right-angled Coxeter n-handlebody of genus $\mathfrak{g}-1$, denoted by $Q_{\mathfrak{g}-1}$. Conversely, Q can be recovered from $Q_{\mathfrak{g}-1}$ by gluing its two disjoint boundary facets associated with B, which implies that the orbifold fundamental group of Q is an HNN-extension on $\pi_1^{\mathrm{orb}}(Q_{\mathfrak{g}-1})$. Write $G_{\mathfrak{g}-1}=\pi_1^{\mathrm{orb}}(Q_{\mathfrak{g}-1})$, and let W_{B^+} and W_{B^-} be two isomorphic subgroups of $G_{\mathfrak{g}-1}$ determined by two copies of B. Then we have

$$\pi_1^{\mathrm{orb}}(Q)\cong G_{\mathfrak{g}-1}*_\phi=\langle G_{\mathfrak{g}-1},t \mid t^{-1}at=\phi(a), a\in W_{B^-}\rangle \tag{6-1}$$

where $\phi\colon W_{B^-}\to W_{B^+}$ is an isomorphism by mapping $s'\in W_{B^-}$ into $s\in W_{B^+}$. Generally, $\pi_1^{\mathrm{orb}}(Q)$ is isomorphic to $\mathfrak{g}$ HNN-extensions on the right-angled Coxeter group $W(P_Q,\mathcal{F}_B)$ as we have seen in the proof of Proposition 4.3:

$$\pi_1^{\mathrm{orb}}(Q)\leftarrow G_{\mathfrak{g}-1}\leftarrow\cdots\leftarrow G_1\leftarrow G_0=W(P_Q,\mathcal{F}_B),$$

where each G_k is also an HNN-extension over G_{k-1} for $1 \le k \le \mathfrak{g}-1$, and $G_0 = W(P_Q, \mathscr{F}_B)$ is a right-angled Coxeter group.

According to the normal form theorem of HNN-extensions (see Theorem 2.22), each element x in $\pi_1^{\mathrm{orb}}(Q)$ has a unique iterative normal form. First, write

$$x = g_0 t_{\mathfrak{g}}^{\epsilon_1} g_1 t_{\mathfrak{g}}^{\epsilon_2} \cdots g_{n-1} t_{\mathfrak{g}}^{\epsilon_n} g_n$$

as a normal form for $t_{\mathfrak{g}}$ where $g_i \in G_{\mathfrak{g}-1}$. Next, inductively each g_i is also a normal form in G_k for $1 \le k \le \mathfrak{g}-1$. More generally, x has a unique form

$$x = g_0 t_1 g_1 \cdots g_{m-1} t_m g_m \tag{6-2}$$

where each g_i is reduced in $G_0 = W(P_Q, \mathscr{F}_B)$, and each t_i is one of $\{t_B^{\pm 1}\}$ which determines an isomorphism of $\{\phi_B^{\pm 1}\}$ on some subgroups of $\pi_1^{\mathrm{orb}}(Q)$. This expression of x is a normal form with respect to all possible t_B. The expression in (6-2) is called a *reduced normal form* of x in $\pi_1^{\mathrm{orb}}(Q)$. The number m is called the (*total*) *t-length of x*.

By applying the Tits theorem (see Theorem 2.20) and the normal form theorem of HNN-extensions (see Theorem 2.22), we have the following conclusion.

Lemma 6.4 *Two reduced words x and y are the same in $\pi_1^{\mathrm{orb}}(Q)$ if and only if one of x and y can be transformed into the other one by a sequence of commutations of right-angled Coxeter groups and t-reductions of HNN-extensions.*

Next, we prove two lemmas.

Lemma 6.5 *If there is a subgroup $\mathbb{Z} \oplus \mathbb{Z}$ in $\pi_1^{\mathrm{orb}}(Q)$, then one generator of $\mathbb{Z} \oplus \mathbb{Z}$ can be presented as a cyclically reduced word in $W(P_Q, \mathscr{F}_B)$.*

Proof Assume that there is a subgroup $\mathbb{Z} \oplus \mathbb{Z}$ in $\pi_1^{\mathrm{orb}}(Q)$, which is generated by two reduced normal forms as in (6-2),

$$x = g_0 t_1 g_1 \cdots g_{m-1} t_m g_m, \quad y = h_0 t_1' h_1 \cdots h_{n-1} t_n' h_n.$$

Then $xy = yx$ in $\pi_1^{\mathrm{orb}}(Q)$. By Lemma 6.4, xy and yx have the same reduced normal form as in (6-2). We do t-reductions on

$$xy = g_0 t_1 g_1 \cdots g_{m-1} t_m g_m \cdot h_0 t_1' h_1 \cdots h_{n-1} t_n' h_n,$$
$$yx = h_0 t_1' h_1 \cdots h_{n-1} t_n' h_n \cdot g_0 t_1 g_1 \cdots g_{m-1} t_m g_m.$$

Since x and y are reduced normal forms, xy and yx have the same tails. Without loss of generality, assume that $m \ge n$. Write $\tilde{y} = t_1' h_1 \cdots h_{n-1} t_n' h_n = h_0^{-1} y$. Then x can be written as

$$x = g_0 t_1 g_1 \cdots t_{m-n} g_{m-n} \tilde{y} = g_0 t_1 g_1 \cdots t_{m-n} g_{m-n} \cdot h_0^{-1} y.$$

Since x and y generate $\mathbb{Z} \oplus \mathbb{Z}$, both y and xy^{-1} do so. The word xy^{-1} has a shorter t-length. We further do t-reductions on xy^{-1} to get a normal form, also denoted by x.

We can always continue to do this algorithm, so that we can take either x or y from $W(P_Q, \mathcal{F}_B)$. Suppose $y = h \in W(P_Q, \mathcal{F}_B)$.

Furthermore, we can assume that h is a cyclically reduced word in $W(P_Q, \mathcal{F}_B)$. In fact, if h is not cyclically reduced, without loss of generality, assume that h is of the form $w^{-1}h'w$, where w is an arbitrary word and h' is a cyclically reduced word in $W(P_Q, \mathcal{F}_B)$. Then we replace h by h', such that h' and wxw^{-1} generate a $\mathbb{Z} \oplus \mathbb{Z}$ in $\pi_1^{\mathrm{orb}}(Q)$. □

Lemma 6.6 *Let $x = g_0 \cdot t_1 \cdots t_k$ be a reduced normal form, where $g_0 \in W(P_Q, \mathcal{F}_B)$ and each t_i is one of $\{t_B^{\pm 1}, B \in \mathcal{F}_B\}$, and h be a cyclically reduced word in $W(P_Q, \mathcal{F}_B)$. Then x, h cannot generate a $\mathbb{Z} \oplus \mathbb{Z}$ in $\pi_1^{\mathrm{orb}}(Q)$.*

Proof If x, h generate a $\mathbb{Z} \oplus \mathbb{Z}$ in $\pi_1^{\mathrm{orb}}(Q)$, then

$$x \cdot h = g_0 \cdot t_1 \cdots t_k \cdot h = g_0 h' \cdot t_1 \cdots t_k,$$

where $h' = \phi_1 \circ \cdots \circ \phi_k(h)$ is the image of the composition of some ϕ_i on h.

We first claim that h' is reduced in $W(P_Q, \mathcal{F}_B)$, and the word length of h' and h are equal. In fact, for each i, ϕ_i is an isomorphism from some W_{B^-} to W_{B^+} which maps generators to generators, and all W_{B^+} and W_{B^-} are subgroups of $W(P_Q, \mathcal{F}_B)$.

Next, we claim that $h = h'$. In fact, $xh = hx$ implies that $g_0 h' = h g_0$, that is $h' = g_0^{-1} h g_0$. Let $h = s_1 \cdots s_n$. If there exists a letter s in g_0 such that:

- $l(hs) = l(h) + 1$ and $hs \neq sh$, then $l(h') > l(h)$ which is a contradiction.
- $l(hs) = l(h) - 1$, then there is a letter $s_k = s$ in h such that $s_j s = s s_j \neq 1$ for any $j > k$. So h equal to a new word $s_1 \cdots s_{k-1} s_{k+1} \cdots s_n s$. Moreover, $g_0 h' = h g_0$ implies that there is a letter $s_i = s$ in $h g_0$ can be moved to the head of word $h g_0$.
  - If $i = k$, then s commutes with all letters except s_k in h. Since there are two s canceled in $h g_0$ by the relation $s^2 = 1$, there are two canceled s in $g_0 h'$. Since g_0 is reduced, s commutes with all letters except s in g_0 and there is an s that can be moved to the head of word h'. Now we consider new $\bar{h}' := sh'$ and $\bar{h} := sh$. Then there is no s in sh and s commutes with all letters in sh.
  - Otherwise, $i \neq k$. Then h equals to a word with form $s\bar{h}s$, which contradicts that h is cyclically reduced.

Hence all letters in g_0 commute with h. That is, $g_0 h = h g_0 = g_0 h'$, Thus $h = h' = \phi_1 \circ \cdots \circ \phi_k(h)$.

If $\phi_1 \circ \cdots \circ \phi_k = \mathrm{id}$, then the associated sequence $t_1 \cdots t_k = 1$, which contradicts that x is reduced. If $\phi_1 \circ \cdots \circ \phi_k \neq \mathrm{id}$ has a fixed point in all letters in h, then there is a letter s_0 in h such that $\phi_1 \circ \cdots \circ \phi_k(s_0) = s_0$. Furthermore, $s_0, \phi_k(s_0), \ldots, \phi_1 \circ \cdots \circ \phi_{k-1}(s_0)$ determine a noncontractible facet in Q, which contradicts that Q is simple. More generally, if $\phi_1 \circ \cdots \circ \phi_k \neq \mathrm{id}$ has no fixed point, then there is a generator s_1 as a letter in h, such that $s_2 = \phi_1 \circ \cdots \circ \phi_k(s_1) \neq s_1$. Continue this procedure; one can get a sequence

$s_1, s_2, s_3, \ldots$, such that each s_i is a generator as a letter in h and $s_i = \phi_1 \circ \cdots \circ \phi_k(s_{i-1})$. However, the word length of h is finite; thus there must be two same elements in the sequence. Geometrically, this means that there is a noncontractible facet in Q, which contradicts that Q is simple. □

Now let us give the proof of the necessity of Theorem B in the general case.

Proof of the necessity of Theorem B Suppose that there are two elements x and y in $\pi_1^{\mathrm{orb}}(Q)$ which generate a rank two free abelian subgroup $\mathbb{Z} \oplus \mathbb{Z}$. Our arguments are divided into the following steps.

Step 1 *Simplify two generators x and y of $\mathbb{Z} \oplus \mathbb{Z}$ by doing t-reductions.*

Lemma 6.5 tells us that one of x or y can be chosen as a cyclically reduced word h in $W(P_Q, \mathcal{F}_B)$, say $y = h$. Now if x is also a word in $W(P_Q, \mathcal{F}_B)$ (i.e. the t-length of x is zero), then by Moussong's theorem (see Theorem 6.3), there is a $\square$-belt in P_Q which can appear in Q, as desired.

Next let us consider the case in which the t-length of x is greater than zero. Let

$$x = g_0 t_1 g_1 \cdots g_{m-1} t_m g_m$$

be a reduced normal form in $\pi_1^{\mathrm{orb}}(Q)$. Then $xh = hx$ implies that

$$\begin{aligned} g_m h &= h g_m, & t_m \cdot h &= \phi_m(h) \cdot t_m, \\ g_{m-1} \cdot \phi_m(h) &= \phi_m(h) \cdot g_{m-1}, & t_{m-1} \cdot \phi_m(h) &= \phi_{m-1} \circ \phi_m(h) \cdot t_{m-1}, \\ &\vdots & & \\ g_0 \cdot \phi_1 \circ \cdots \circ \phi_m(h) &= \phi_1 \circ \cdots \circ \phi_m(h) \cdot g_0, & & \end{aligned}$$

where each $\phi_i : W_{B_i} \to W_{B_i'}$ is an isomorphism determined by some $B_i \in \mathcal{F}_B$, each $\phi_i \circ \cdots \circ \phi_m(h)$ is an expression in $W_{B_i'} \cap W_{B_{i-1}}$ for $i = 2, \ldots, m$ and $\phi_1 \circ \cdots \circ \phi_m(h) \in W_{B_1'}$ for $h \in W_{B_m}$. Here two B_i and B_j may correspond to the same $B \in \mathcal{F}_B$.

Step 2 *Find facets F_1 and F_3 around B or B' in P_Q.*

Without loss of generality, $h \in W(P_Q, \mathcal{F}_B)$ is a cyclically reduced word. Since h is a free element in $W_{B_m} \cap W(P_Q, \mathcal{F}_B)$, we can take two generators s_1 and s_3 in h corresponding to two disjoint facets F_1 and F_3 of P_Q such that F_1 and F_3 intersect with B_m. In particular, $s_1 s_3$ is a free element in $W_{B_m} < W(P_Q, \mathcal{F}_B) = G_0 < \cdots < G_{\mathfrak{g}-1} < G_{\mathfrak{g}} = \pi_1^{\mathrm{orb}}(Q)$.

Step 3 *Find the facet F_2 which intersects F_1 and F_3.*

If $g_m \neq 1$, since x is a normal form, then g_m is a representative of a coset of W_{B_m} in $\pi_1^{\mathrm{orb}}(Q)$. Thus there is a generator $s_2 \notin S(W_{B_m})$ in g_m such that $hs_2 = s_2 h$, where $S(W_{B_m})$ is the generator set of W_{B_m}. This generator s_2 determines a facet F_2 in P_Q, as desired.

If $g_m = 1$, then $x \cdot h = g_0 t_1 g_1 \cdots t_m \cdot h = g_0 t_1 g_1 \cdots t_{m-1} g_{m-1} \cdot \phi_m(h) \cdot t_m$. A similar argument shows that either there is a $s_2 \notin S(W_{B_{m-1}})$ as desired, or

$$x \cdot h = g_0 t_1 g_1 \cdots t_{m-1} \cdot \phi_m(h) \cdot t_m = g_0 t_1 g_1 \cdots t_{m-2} g_{m-2} \cdot \phi_{m-1} \circ \phi_m(h) \cdot t_{m-1} t_m.$$

We can continuously carry out the above procedure. Finally we can arrive at two possible cases:

- There exists some $g_i \neq 1$ for $i > 0$. Then there must be a letter s_2 in g_i which determines the required F_2.
- x is of the form $x = g_0 \cdot t_1 \cdots t_m$, where $g_0 \in W(P_Q, \mathcal{F}_B)$ and $t_1 \cdots t_m$ is a word formed by letters in $\{t_B^{\pm 1}\}$. By Lemma 6.6, x and h cannot generate a subgroup $\mathbb{Z} \oplus \mathbb{Z}$ in $\pi_1^{\mathrm{orb}}(Q)$, so $x = g_0 \cdot t_1 \cdots t_m$ is impossible.

Thus, we can always find a facet F_2 from a nontrivial g_i in the reduced form (6-2) of x where $i > 0$.

Step 4 *Find a facet F_4 such that F_1, F_2, F_3 and F_4 determine a $\square$-belt in Q.*

We proceed our argument as follows.

(I) If there is only a $g_i \neq 1$ (i.e. $g_j = 1$ for any $j \neq i$) in the expression of x, then $x = t_1 \cdots t_i \cdot g_i \cdot t_{i+1} \cdots t_m$, where i must be more than zero by Lemma 6.6. Now $xh = hx$ implies that $t_1 \cdots t_i \cdot t_{i+1} \cdots t_m = 1$. Actually, if $t_1 \cdots t_i \cdot t_{i+1} \cdots t_m \neq 1$, then $\phi_1 \circ \cdots \circ \phi_m(h) = h$ implies that there is a noncontractible facet in Q, which is impossible (also see the proof of Lemma 6.6). Thus, $t_1 \cdots t_i = (t_{i+1} \cdots t_m)^{-1}$, so $x = t_1 \cdots t_i \cdot g_i \cdot t_{i+1} \cdots t_m = (t_{i+1} \cdots t_m)^{-1} g_i (t_{i+1} \cdots t_m)$. Since x, h generate a $\mathbb{Z} \oplus \mathbb{Z}$, we see that $g_i, \phi_{i+1} \circ \cdots \circ \phi_m(h)$ generate a $\mathbb{Z} \oplus \mathbb{Z}$ in $W(P_Q, \mathcal{F}_B)$. Then by Moussong's theorem (see Theorem 6.3), there is a $\square$-belt in Q.

(II) If there are at least two nontrivial $g_i, g_j \neq 1$ in x where $0 < j < i \leq m$ but $g_k = 1$ for all $k > j$ and $k \neq i$, then one may write $x = \cdots t_j \cdot g_j \cdot t_{j+1} \cdots t_i \cdot g_i \cdot t_{i+1} \cdots t_m$. So we have

$$\begin{aligned} xh &= \cdots t_j \cdot g_j \cdot t_{j+1} \cdots t_i \cdot g_i \cdot t_{i+1} \cdots t_m \cdot h \\ &= \cdots t_j \cdot g_j \cdot t_{j+1} \cdots t_i \cdot g_i \cdot h' t_{i+1} \cdots t_m \\ &= \cdots t_j \cdot g_j \cdot t_{j+1} \cdots t_i \cdot h' g_i \cdot t_{i+1} \cdots t_m \\ &= \cdots t_j \cdot g_j \cdot h'' \cdot t_{j+1} \cdots t_i \cdot g_i \cdot t_{i+1} \cdots t_m \end{aligned} \tag{6-3}$$

where $h' = \phi_{i+1} \circ \cdots \circ \phi_m(h)$ and $h'' = \phi_{j+1} \circ \cdots \circ \phi_m(h)$. Since $xh = hx$, we have that $g_j h'' = h'' g_j$, so we can take a generator s_4 in g_j (not in $S(W_{B_j})$) such that $h'' s_4 = s_4 h''$. Similarly, here s_4 determines a facet F_4 of P_Q such that $F_4 \cap F_1'' \neq \varnothing$ and $F_4 \cap F_3'' \neq \varnothing$ where F_1'' and F_3'' are two facets of P_Q determined by the images of $\phi_{j+1} \circ \cdots \circ \phi_m$ on s_1 and s_3. In particular, $F_2 \neq F_4$ in P_Q. Otherwise, the intersection of $q(F_1)$ and $q(F_2)$ in Q is disconnected where $q \colon Q \to P_Q$ is the quotient map defined in (4-1), which contradicts that Q is simple. Hence, we get a $\square$-belt in Q.

(III) If there are only g_0 and g_i that are nontrivial in x where $i > 0$, then one may write

$$x = g_0 t_1 \cdots t_i \cdot g_i \cdot t_{i+1} \cdots t_m.$$

Without loss of generality, assume that g_0 and g_i are two reduced words in $W(P_Q, \mathscr{F}_B)$. Now if $x = g_0 t_1 \cdots t_i \cdot g_i \cdot t_{i+1} \cdots t_m = t_1 \cdots t_i \cdot g_0' g_i \cdot t_{i+1} \cdots t_m$ where $g_0' = \phi_i^{-1} \circ \cdots \circ \phi_1^{-1}(g_0)$, then by the proof of Lemma 6.6, $xh = hx$ implies that $t_1 \cdots t_i \cdot t_{i+1} \cdots t_m = 1$. As in case (I), $g_0' g_i$ and $h' = \phi_{i+1} \circ \cdots \circ \phi_m(h)$ generate a $\mathbb{Z} \oplus \mathbb{Z}$ in $W(P_Q, \mathscr{F}_B)$. Hence we can find a $\Box$ in Q. If

$$x = g_0 t_1 \cdots t_i \cdot g_i \cdot t_{i+1} \cdots t_m = g_0' t_1 \cdots t_j g_0'' t_{j+1} \cdots t_i \cdot g_0''' g_i \cdot t_{i+1} \cdots t_m$$

where g_0'' cannot cross t_{j+1} and $g_0 = g_0' \cdot \phi_1 \circ \cdots \circ \phi_j(g_0'') \cdot \phi_1 \circ \cdots \circ \phi_i(g_0''')$, then as in case (II), there is a generator s_4 in g_0'' which is not in $S(W_{B'_{j+1}})$. Then s_4 determines a facet F_4 of P_Q such that F_4 intersects with F_1 and F_3 in Q. So there is a $\Box$-belt in Q.

Together with all arguments above, we complete the proof. $\Box$

7 Applications

Throughout the following, we always assume that Q is a genus $\mathfrak{g}$ simple handlebody with m facets and M is the manifold double over Q. Moreover, for the sake of discussion, we always assume that M is a smooth manifold. In this section, we shall show that some B-belts in Q can play a role in the obstruction of the existence of some Riemannian metrics on M. First we can see that every B-belt of Q is a π*-injective* suborbifold in the sense of the following Lemma 7.1.

Lemma 7.1 *Let $i: B \hookrightarrow Q$ be a belt of Q. Then $i_*: \pi_1^{\mathrm{orb}}(B) \to \pi_1^{\mathrm{orb}}(Q)$ is an injection. Moreover, if B is not orbifold-aspherical, then Q is not orbifold-aspherical.*

Proof If B is a cutting belt of Q, then by Lemma 4.2, B is π_1-injective.

If B is a belt of Q which is disjoint with any cutting belts of Q, then B can be embedded into P_Q, so the induced map $\pi_1^{\mathrm{orb}}(B) \to \pi_1^{\mathrm{orb}}(P_Q)$ is an injection. Thus $i_*: \pi_1^{\mathrm{orb}}(B) \to \pi_1^{\mathrm{orb}}(P_Q) \to \pi_1^{\mathrm{orb}}(Q)$ is an injection.

If B intersects with some cutting belts, then we can always do some deformation on B such that it intersects transversely with those cutting belts. Thus, we may assume that B is split into $B_1, B_2, \ldots, B_k$ by cutting belts $E_1, E_2, \ldots, E_{k-1}$ such that for each i, B_i and B_{i+1} exactly intersect with E_i since all E_i are simple polytopes and $|B|$ is a ball. In addition, it is also easy to see that for each i, $\pi_1^{\mathrm{orb}}(B_i)$ is a right-angled Coxeter group. Now

$$\pi_1^{\mathrm{orb}}(B) = \pi_1^{\mathrm{orb}}(B_1) * \pi_1^{\mathrm{orb}}(B_2) * \cdots * \pi_1^{\mathrm{orb}}(B_k) / \langle R(E_i) \mid i = 1, \ldots, k-1 \rangle$$

where $R(E_i)$ is a relation set consisting of all equations $s = t$, each of which is associated with $F_s \sim_{E_i} F_t$ where $F_s \in \mathscr{F}(B_i)$ and $F_t \in \mathscr{F}(B_{i+1})$.

Now for $g \in \pi_1^{\mathrm{orb}}(B_i)$, $i_*(g) = t_{i-1}^{-1} \cdots t_1^{-1} g t_1 \cdots t_{i-1}$. Define $\kappa: \pi_1^{\mathrm{orb}}(Q) \to \pi_1^{\mathrm{orb}}(B)$,

$$\kappa(s) = \begin{cases} s & \text{if } F_s \in \bigcup \mathscr{F}(B_i), \\ 1 & \text{if } F_s \in \mathscr{F}(P_Q) - \bigcup \mathscr{F}(B_i), \end{cases}$$

where all torsion free generators are mapped to 1. Then it is clear that κ is well defined and $\kappa \circ i_* = \mathrm{id}$. Hence, i_* is an injection.

Let $\widetilde{Q}$ and $\widetilde{B}$ be the universal cover of Q and B, respectively. If B is not orbifold-aspherical, then there is an integer $k \geq 2$ such that $\pi_k(\widetilde{B}) \neq 0$ and $\pi_i(\widetilde{B}) = 0$ for any $1 \leq i < k$. By the Hurewicz theorem, $H_k(\widetilde{B}) \cong \pi_k(\widetilde{B}) \neq 0$. Hence by Proposition 5.4 there is a Δ^k-belt in B. So there is a Δ^k-belt of Q. Hence by Theorem A, Q is not orbifold-aspherical. □

Lemma 7.2 *The manifold double of a simple handlebody is orientable.*

Proof Let Q be a simple n-handlebody with m facets and cutting belts $\{B_1, \dots, B_{\mathfrak{g}}\}$, P_Q be the associated simple polytope, and $M = Q \times (\mathbb{Z}_2)^m/\sim$ be the manifold double over Q, as defined in (2-1). It suffices to prove that $H_n(M;\mathbb{Z}) \cong \mathbb{Z}$. We shall use the proof method of Nakayama and Nishimura (see [47, Theorem 1.7]). The combinatorial structure of P_Q defines a natural cellular decomposition of M. We denote by $\{(C_k(M), \partial_k)\}$ the chain complex associated with this cellular decomposition. In particular, $C_n(M)$ and $C_{n-1}(M)$ are the free abelian groups generated by $\{P_Q\} \times (\mathbb{Z}_2)^m = \{(P_Q, g) \mid g \in (\mathbb{Z}_2)^m\}$ and $\mathcal{F}(P_Q) \times (\mathbb{Z}_2)^m/\sim' = \{[F, g] \mid F \in \mathcal{F}(P_Q), g \in (\mathbb{Z}_2)^m\}$, respectively, where the equivalence class of $\mathcal{F}(P_Q) \times (\mathbb{Z}_2)^m$ is defined by the equivalence relation

$$\begin{aligned} (F, g) &\sim' (F, g \cdot e_F) && \text{if } F \in \mathcal{F}(P_Q) - \mathcal{F}_B, \\ (B^+, g) &\sim' (B^-, g) && \text{if } B^+, B^- \in \mathcal{F}_B. \end{aligned}$$

It should be pointed out that actually there is a characteristic map $\lambda\colon \mathcal{F}(Q) \to (\mathbb{Z}_2)^m$ in the construction of $M = Q \times (\mathbb{Z}_2)^m/\sim$ such that $\{\lambda(F) = e_F \mid F \in \mathcal{F}(Q)\}$ is the standard basis $\{e_i \mid i = 1, \dots, m\}$ of $(\mathbb{Z}_2)^m$. For any facet F' in $\mathcal{F}(P_Q) - \mathcal{F}_B$, there must be a facet F in $\mathcal{F}(Q)$ such that $F' = F$ or $F' \subsetneq F$, so F' and F are colored by the same element e_F of $(\mathbb{Z}_2)^m$. For any B in $\mathcal{F}_B$, since $\operatorname{Int} B \subset \operatorname{Int}(Q)$, we adopt the convention that B is colored by the unit element $\mathbf{e}_0$ of $(\mathbb{Z}_2)^m$. In other words, the characteristic map $\lambda\colon \mathcal{F}(Q) \to (\mathbb{Z}_2)^m$ induces a compatible characteristic map $\lambda'\colon \mathcal{F}(P_Q) \to (\mathbb{Z}_2)^m$ such that for any $F' \in \mathcal{F}(P_Q) - \mathcal{F}_B$, $e_{F'} = \lambda'(F') = \lambda(F) = e_F$ where $F \in \mathcal{F}(Q)$ with $F' \subset F$, and for B in $\mathcal{F}_B$, $\lambda'(B) = \mathbf{e}_0$.

We give an orientation on each facet F_i and $B_i^{\pm}$ such that the orientation of B_i^+ is exactly the inverse orientation of B_i^-, so

$$\partial P_Q = \sum_{F \in \mathcal{F}(P_Q)} F = F_1 + \cdots + F_{m'} + B_1^+ + \cdots + B_{\mathfrak{g}}^+ + B_1^- + \cdots + B_{\mathfrak{g}}^- = \sum_{F \in \mathcal{F}(P_Q) - \mathcal{F}_B} F$$

where m' is the number of all facets in $\mathcal{F}(P_Q) - \mathcal{F}_B$.

Let $c_n = \sum_{g \in (\mathbb{Z}_2)^m} n_g (P, g)$ be an n-cycle of $C_n(M)$ where $n_g \in \mathbb{Z}$. Then

$$\partial(c_n) = \left[\sum_{g \in (\mathbb{Z}_2)^m} n_g \sum_{F \in \mathcal{F}(P_Q)} (F, g) \right] = \sum_{[F,g] \in ((\mathcal{F}(P_Q) - \mathcal{F}(B)) \times (\mathbb{Z}_2)^m)/\sim'} (n_g + n_{g e_F})[F, g] = 0,$$

which induces that $n_g = -n_{ge_F}$ for any facet $F \in \mathcal{F}(P_Q) - \mathcal{F}_B$ and $g \in (\mathbb{Z}_2)^m$. Let $l(g)$ denote the word length of g presented by $\{e_F\}$. For any $g \in (\mathbb{Z}_2)^m$, there exists a subset $\mathcal{I}_g = \{F_{i_1}, \dots, F_{i_k}\}$ of $\mathcal{F}(P_Q) - \mathcal{F}_B$ such that $g = \prod_{F \in \mathcal{I}_g} e_F$. Then we see easily that

$$n_g = -n_{ge_{F_{i_1}}} = n_{ge_{F_{i_1}}e_{F_{i_2}}} = \cdots = (-1)^{l(g)} n_{g \prod_{F \in \mathcal{I}_g} e_F} = (-1)^{l(g)} n_{\mathbf{e}_0}$$

so $c_n = n_{\mathbf{e}_0} \sum_{g \in (\mathbb{Z}_2)^m} (-1)^{l(g)} (P, g)$. Then we obtain that $H_n(M; \mathbb{Z}) = \ker \partial_n \cong \mathbb{Z}$ is generated by $\sum_{g \in (\mathbb{Z}_2)^m} (-1)^{l(g)} (P, g)$; hence M is orientable. □

7.1 Nonpositive curvature

Recall that a geodesic metric space X is *nonpositively curved* if it is a locally CAT(0) space. In general, X is said to be of *curvature* $\le k$ (in the sense of Alexandrov) if it is locally a CAT(k) space. See [7, Definition 1.2, page 159].

Theorem 7.3 [7, Theorem 1A.6, page 173] *A smooth Riemannian manifold N is of curvature $\le k$ in the sense of Alexandrov if and only if the sectional curvature of N is $\le k$.*

Proposition 7.4 [50, Corollary 6.2.4] *If (N, g) is a complete Riemannian manifold of nonpositive curvature, then the fundamental group is torsion free.*

Let Q be a simple handlebody of dimension $n \ge 3$ and genus $\mathfrak{g}$, and $M \to Q$ be the smooth manifold double over Q, as defined in (2-1). Let P_Q be the simple polytope obtained from Q by cutting open along $\mathfrak{g}$ disjoint cutting belts $B_1, \dots, B_{\mathfrak{g}}$ in Q. More precisely, P_Q can be obtained in the following way: For each belt B_i, choose a regular neighborhood $N(B_i)$ of B_i that is homeomorphic to $B_i \times [-1, 1]$ as manifolds with corners. Clearly $N(B_i)$ is identified with a simple polytope, and it can also be understood as the disk D^1-bundle of the trivial normal bundle of B_i in Q. Then we get P_Q by removing the interiors of trivial D^1-bundles $B_i \times [-1, 1]$ of all B_i.

In order to use the Gromov lemma as above, we need a cubical cellular structure of the manifold double M over Q. For this, we perform the following procedure:

(1) First we decompose Q into more pieces,

$$Q = P_Q \bigcup_{i=1}^{\mathfrak{g}} N^+(B_i) \cup N^-(B_i),$$

where $N^+(B_i) = B_i \times [0, 1]$ and $N^-(B_i) = B_i \times [-1, 0]$ satisfy $N(B_i) = N^+(B_i) \cup N^-(B_i)$.

(2) Next, the standard cubical decompositions of P_Q and all $N^{\pm}(B_i)$ determine a right-angled Coxeter cubical cellular decomposition of Q, denoted by $\mathscr{C}(Q)$. Specifically, all cone points of P_Q and all $N^{\pm}(B_i)$ will be 0-cells with trivial local group in $\mathscr{C}(Q)$. There are two kinds of $k(>0)$-cubes in the cubical decompositions of P_Q and all $N^{\pm}(B_i)$, each of which either intersects transversely with an $(n-k)$-face $f^{n-k} = F_{i_1} \cap \dots \cap F_{i_k}$ or intersects transversely with an $(n-k)$-face $f^{n-k} = F_{i_1} \cap \dots \cap F_{i_{k-1}} \cap B_i$. The first type of cubes determine right-angled Coxeter cubical

cells of the form $e^k/(\mathbb{Z}_2)^k$ in $\mathscr{C}(Q)$, and the second type of cubes determine right-angled Coxeter cubical cells of the form $e^k/(\mathbb{Z}_2)^{k-1}$. Then, $\mathscr{C}(Q)$ is obtained by attaching each pair associated with B_i of the second type of cubes together. It is clear that $\mathscr{C}(Q)$ is a right-angled Coxeter cubical cellular decomposition of Q.

(3) Finally, by pulling back $\mathscr{C}(Q)$ to M via the covering map $p\colon M \to Q$, one can obtain a cubical cellular decomposition of M, denoted by $\mathscr{C}(M)$, such that each cube in $\mathscr{C}(M)$ is a connected component of $p^{-1}(c)$ for c in $\mathscr{C}(Q)$. In particular, all vertices in $\mathscr{C}(M)$ exactly consist of the liftings of all cone points in $\mathscr{C}(Q)$.

Lemma 7.5 *Let v be a vertex in $\mathscr{C}(M)$. Then $\mathrm{Lk}(v)$ in $\mathscr{C}(M)$ is combinatorially isomorphic to one of the nerves $\mathcal{N}(P_Q)$, $\mathcal{N}(N^+(B_i))$ or $\mathcal{N}(N^-(B_i))$.*

Proof In fact, if $p(v)$ is the cone point of P_Q, then each $k(>0)$-cube adjacent to v gives a $(k-1)$-simplex in $\mathcal{N}(P_Q)$, which corresponds to an $(n-k)$-face of P_Q. Therefore, $\mathrm{Lk}(v) \cong \mathcal{N}(P_Q)$. The same argument can be applied to the case where $p(v)$ is the cone point of $N^+(B_i)$ or $N^-(B_i)$. □

Proposition 7.6 *M is nonpositively curved if and only if Q is flag.*

Proof Let $\mathscr{C}(M)$ be the cubical cellular decomposition of M discussed above. The Gromov lemma (see Proposition 5.2) tells us that M is nonpositively curved if and only if the link of each vertex in the cubical cellular decomposition of M is flag. By Lemma 7.5, the latter of the above statement means that $\mathcal{N}(P_Q)$ and all $\mathcal{N}(N^{\pm}(B_i))$ are flag, so P_Q and all $N^{\pm}(B_i)$ are flag simple polytopes. This is also equivalent to saying that Q is flag. Thus, if a simple handlebody Q is flag, then its manifold double M is nonpositively curved.

Conversely, if M is nonpositively curved, then by the Cartan–Hadamard theorem, M is aspherical. Then Q is orbifold-aspherical. By Theorem A, Q is flag. □

7.2 Strictly negative curvature

Proposition 7.7 (Preissmann [50, Theorem 6.2.6]) *If (N, g) is a compact manifold of negative curvature, then any abelian subgroup of the fundamental group is cyclic.*

Proposition 7.8 (Gromov [32]) *Let $\mathscr{C}$ be a cubical complex. Suppose that the link of each vertex in $\mathscr{C}$ is flag and contains no $\square$. Let us give $\mathscr{C}$ a $(N, -\epsilon)$ geometry, where each cube in $\mathscr{C}$ is isomorphic to the unit cube in the hyperbolic space of curvature $-\epsilon$. If ϵ is sufficiently small then $K(\mathscr{C}) \le -\epsilon$.*

Lemma 7.9 *Let Q be a simple handlebody with m facets, and M be the manifold double over Q. Then $\mathbb{Z} \oplus \mathbb{Z} < \pi_1(M)$ if and only if $\mathbb{Z} \oplus \mathbb{Z} < \pi_1^{\mathrm{orb}}(Q)$.*

Proof If $\mathbb{Z} \oplus \mathbb{Z} < \pi_1(M)$, then the short exact sequence

$$1 \to \pi_1(M) \to \pi_1^{\mathrm{orb}}(Q) \xrightarrow{\lambda} (\mathbb{Z}_2)^m \to 1$$

induces that $\mathbb{Z}\oplus\mathbb{Z} < \pi_1(M) < \pi_1^{\mathrm{orb}}(Q)$. Conversely, if $\mathbb{Z}\oplus\mathbb{Z} < \pi_1^{\mathrm{orb}}(Q)$, then by Theorem B, there is a $\square$-belt in Q and $\mathbb{Z}\oplus\mathbb{Z} < \pi_1^{\mathrm{orb}}(Q)$ is generated by two pairs of disjoints facets in $\square$. Denote the generators of $\mathbb{Z}\oplus\mathbb{Z}$ by $x = s_1s_3$ and $y = s_2t_1t_2\cdots t_ks_4t_k^{-1}\cdots t_2^{-1}t_1^{-1}$ where each t_i is one of $\{t_B^{\pm 1}\}$; then $\lambda(x^2) = \lambda(y^2) = 1$. Hence $x^2, y^2 \in \ker\lambda \cong \pi_1(M)$. So $\mathbb{Z}\oplus\mathbb{Z} < \pi_1(M)$. □

Proposition 7.10 *If M admits a strictly negative curvature, then Q is flag and contains no $\square$-belt. Specially, if Q is a simple polytope, then M admits a strictly negative curvature if and only if Q is flag and contains no $\square$-belt.*

Proof If M admits a strictly negative curvature, then by Proposition 7.6, Q is flag. Furthermore, if there is a $\square$-belt in Q, then by Theorem B, there is a rank-two abelian subgroup in $\pi_1^{\mathrm{orb}}(Q)$. By Lemma 7.9, there is a subgroup $\mathbb{Z}\oplus\mathbb{Z}$ in $\pi_1(M)$. By Preissmann's result (Proposition 7.7), M cannot admit a strictly negative curvature. Hence, Q contains no $\square$-belt.

When Q is a simple polytope, then the standard cubical cellular decomposition of Q induces a cubical cellular decomposition of M, denoted by $\mathscr{C}(M)$, such that each point $v \in \mathscr{C}(M)$ has link $\mathscr{N}(Q)$. By a result of Gromov (Proposition 7.8), if $\mathscr{N}(Q)$ is flag and contains no $\square$-belt, then M admits a strictly negative curvature. □

Remark 7.11 We are inclined to think that M admits a strict negative curvature if a simple handlebody Q is flag and contains no $\square$-belt. But we cannot find a suitable cubical decomposition of M so as to use Gromov's result. Of course, this is related to the weak hyperbolization conjecture: "Let N be a closed aspherical manifold. Then either $\pi_1(N)$ contains $\mathbb{Z}\oplus\mathbb{Z}$ or $\pi_1(N)$ is Gromov-hyperbolic". See [37, Conjecture 20.12].

7.3 Hyperbolic curvature

When Q is a simple 3-polytope, the Pogorelov theorem, which is right-angled case of the Andreev theorem (see [2; 53]), states that Q admits a right-angled hyperbolic structure if and only if it is flag and contains no 4-belt, where "right-angled" means that all dihedral angles are $\pi/2$. This gives a combinatorial equivalent description of the hyperbolicity of simple 3-polytopes as a right-angled Coxeter orbifold. Now Q is also called a (right-angled) hyperbolic polyhedra in $\mathbb{H}^3$.

As a generalization of 3-dimensional hyperbolic polyhedra, a 3-manifold with corners is *hyperbolic* if its interior admits a hyperbolic metric which can be extended to the boundary such that its all faces are totally geodesic (or locally convex). Moreover we say that a 3-manifold with corners is *right-angled hyperbolic* if its all dihedral angles are $\pi/2$.

Notice that a hyperbolic 3-manifold with corners is not right-angled in general. A 3-manifold with corners can be equipped with many different orbifold structures. A hyperbolic structure on a 3-manifold with corners should be compatible with an orbifold structure on it. Hence there may be different hyperbolic structures on a 3-manifold with corners. The hyperbolization of 3-manifolds with corners corresponds to the generalization of the Andreev theorem [2; 53]. This question is still open now. However, the hyperbolic structure of a hyperbolic closed 3-orbifold (or 3-manifold) is unique by the Mostow rigidity theorem (see [45]).

Here we mainly consider the right-angled hyperbolicity of simple 3-manifolds with corners, where a simple 3-manifold with corners is given by forgetting the orbifold structure on a simple 3-orbifold. Then one can obtain the same understanding for right-angled hyperbolicity from the following two geometric objects:

(1) a right-angled hyperbolic simple 3-manifold with corners;

(2) a hyperbolic simple 3-orbifold (as a right-angled Coxeter 3-orbifold).

Thus, a right-angled hyperbolic simple 3-manifold with corners is a hyperbolic simple 3-orbifold, and vice versa.

Proposition 7.12 *Let Q be a 3-dimensional simple handlebody. Then M is hyperbolic if and only if Q is flag and contains no $\square$-belt.*

Proof Together with Perelman's work, Thurston's hyperbolization theorem implies that a closed orientable 3-manifold is hyperbolic if and only if it is aspherical and atoroidal, where if there is no subgroup $\mathbb{Z} \oplus \mathbb{Z}$ in $\pi_1(M)$, then M is atoroidal. By Lemma 7.2, M is always orientable. Together with Theorems A and B and Lemma 7.9, we know that M is aspherical and atoroidal if and only if Q is flag and contains no $\square$-belt. □

By 3-dimensional hyperbolic manifold theory (see [49]), M is hyperbolic if and only if Q is hyperbolic. Hence, we have:

Corollary 7.13 *Q is right-angled hyperbolic as a 3-manifold with corners if and only if Q is flag and contains no $\square$-belt.*

For a higher-dimensional simple handlebody, the fundamental group of a closed hyperbolic manifold is a discrete convex-cocompact subgroup of $\mathrm{Isom}(\mathbb{H}^n)$, and contains no subgroup $\mathbb{Z} \oplus \mathbb{Z}$. Hence, if M is hyperbolic, then Q must be flag and contains no $\square$-belt. By a result in [18, Corollary 6.11.6] there must exist a triangle or quadrilateral face in a simple polytope of dimension greater than 4. So

- a simple handlebody with dimension greater than 4 cannot be hyperbolic.

In the 4-dimensional case, it is not clear whether only 120-cells are hyperbolic simple 4-polytopes; see [30].

7.3.1 3-handlebodies with simplicial nerve A 3-*handlebody with simplicial nerve* is just a simple 3-orbifold whose underlying space is a 3-handlebody. Let Q be a 3-handlebody with simplicial nerve and with genus $\mathfrak{g} > 0$. Next we show that Q can always be cut into a simple polytope along some codimension-one B-belts.

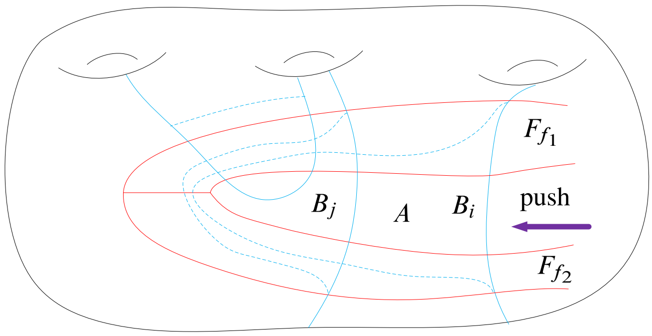


Figure 10: Modifying the boundary of B_i.

Lemma 7.14 *A* 3*-handlebody with simplicial nerve is a simple* 3*-handlebody.*

Proof Let Q be a 3-handlebody of genus $\mathfrak{g} \geq 0$ with simplicial nerve. If $\mathfrak{g} = 0$, then it is easy to check that Q is a simple 3-polytope by the Steinitz theorem and [41, Proposition 3.4]. If $\mathfrak{g} > 0$, let

$$\{(D_i^2, \partial D_i^2) \hookrightarrow (|Q|, \partial|Q|) \mid i = 1, 2, \dots, \mathfrak{g}\}$$

be some disjoint compressing 2-disks in $|Q|$ such that $|Q|$ is cut into a connected 3-ball along those compressing 2-disks. Considering the facial structure determined by the triangulation $\mathcal{N}(Q)$ of $\partial|Q|$, we can always do some slight deformations for the boundaries of compressing 2-disks on faces of Q, so that $\{D_i^2\}$ can be modified into some embedded suborbifolds $\{B_i\}$ of preserving codimension in Q. Each B_i is a polygon.

Given a B_i, by the definition of B-belts, we see that B_i is not a B-belt if and only if there exist two nonadjacent edges f_1 and f_2 in B_i such that

(i) $F_{f_1} \cap F_{f_2} \neq \varnothing$ (probably F_{f_1} and F_{f_2} can even be the same face of Q);

(ii) $F_{f_1} \cup F_{f_2}$ can deformatively retract onto B_i in $|Q|$ (in fact, $F_{f_1} \cup F_{f_2}$ can deformatively retract onto ∂B in $\partial|Q|$);

where F_{f_1} and F_{f_2} are two 2-faces of Q that contain f_1 and f_2 respectively. So, if B_i is not a B-belt, then there is no hole in the area A in $\partial|Q|$ surrounded by F_{f_1}, F_{f_2} and B_i. Then we can modify the boundary of B_i by pushing the retract of $F_{f_1} \cup F_{f_2}$ into F_{f_1} and F_{f_2}, and throwing some edges of B_i away, as shown in Figure 10, so that one can obtain a new B_i' with fewer edges which intersects transversely with $F_{f_1} \cap F_{f_2}$. In particular, if $F_{f_1} = F_{f_2}$, then f_1 and f_2 will become the same edge in B_i', and if $F_{f_1} \neq F_{f_2}$ then f_1 is adjacent to f_2 in B_i'. In addition, if there is also another suborbifold B_j which intersects with the area A in $\partial|Q|$, this means that B_j is not a belt, too. The above "pushing" process will move the boundary of B_j out from the area A and modify B_j into B_j' with fewer edges such that $B_i' \cap B_j' = \varnothing$. Since B_i is a polygon with finite edges, this process can end after a finite number of steps until one has modified B_i into an B-belt which does not intersect with other B_j.

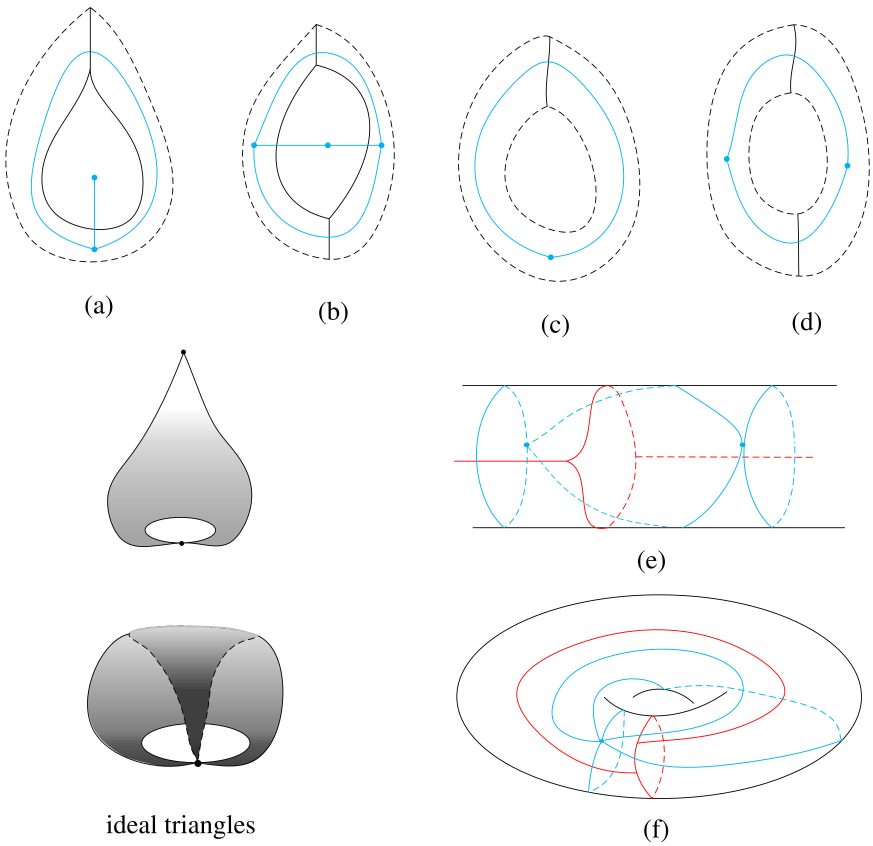


Figure 11: Ideal nerves.

We can perform the same procedure to other non-B-belts in $\{B_j\}_{j\neq i}$. Finally one can obtain a set of disjoint cutting belts such that Q is cut open into a simple 3-polytope along those cutting belts, implying that Q is a simple 3-handlebody. $\square$

Hence, Proposition 7.12 still holds for 3-handlebodies with simplicial nerve.

Proposition 7.15 *A* 3*-handlebody with simplicial nerve* (*the existence of cutting belts is intrinsic*) *is right-angled hyperbolic as a* 3*-manifold with corners if and only if* Q *is flag and contains no* $\square$*-belt.*

7.3.2 3-handlebodies with ideal nerve We say that Q is a 3-*handlebody with ideal nerve* if Q is a right-angled Coxeter 3-orbifold such that its underlying space $|Q|$ is a 3-handlebody and its nerve is an ideal triangulation of the boundary $\partial|Q|$.

Now let Q be a 3-handlebody with ideal nerve. Then, by the definition of ideal triangulation [27, Definition 2.6], the interior of each face of Q is also contractible. On the facial structure of Q, there are three possible cases:

- Some 2-faces of Q are henagons (i.e. 2-faces with only one point of codimension 3 in Q; see (a) in Figure 11) or digons (i.e. 2-faces with only two points of codimension 3 in Q; see (b) in Figure 11).
- There may be some 2-faces with self-intersection (see (c) in Figure 11).
- The intersection of two 2-faces may be not connected (see (d) in Figure 11).

If there is a henagon 2-face of Q, then it gives a self-folded ideal triangle. For example, see the blue part of (a) in Figure 11. In general, if there is a henagon 2-suborbifold in Q, then the nerve of associated faces may give some ideal triangles, such as (e) and (f) in Figure 11. In particular, the nerve of (f) contains only one vertex and two ideal triangles glued along their three edges as shown in Figure 11. All those cases agree with the definition of ideal triangulations in [27, Definition 2.6].

Lemma 7.16 *Let Q be a 3-handlebody with ideal nerve. Then Q is very good if and only if it does not contain a henagon 2-suborbifold.*

Proof By applying a theorem of Morgan or Kato (see [37, Theorem 6.14]), each compact locally reflective 3-orbifold that contains no bad 2-suborbifolds is very good. This means that if there is no henagon 2-suborbifold in Q, then Q is very good. Conversely, if there is a henagon 2-suborbifold in Q, then it is obvious that Q is bad. □

Hence, if there is a henagon 2-suborbifold of Q, then Q cannot be hyperbolic.

Suppose that Q contains no henagon 2-suborbifolds. Then, by Lemma 7.16, Q can be covered finitely by a closed 3-manifold M. In general, Q is not nice in the sense of Davis [18, page 180]; thus there is no natural manifold double defined as in (2-1) for Q.

A digon 2-suborbifold in Q is said to be *essential* if its two vertices are not contained in a unique edge of Q. If there is an essential digon 2-suborbifold in Q, then its nerve $\mathcal{N}(Q)$ will contain two simplices with common vertices. See (d) in Figure 11.

Lemma 7.17 *Let Q be a 3-handlebody with ideal nerve. Assume that there is no henagon suborbifold in Q, and M is a covering manifold over Q. If Q contains an essential digon suborbifold, then M is reducible.*

Proof Assume that two edges of a digon are contained in two 2-faces F_1 and F_2 of Q. Then we consider the double cover of Q, denoted by D_Q, which is obtained by gluing two copies of Q along F_1. At the same time, two copies of F_2 are also glued along $F_1 \cap F_2$, giving an annulus in D_Q. Let M' be a manifold double over D_Q. Then M' can be decomposed into the connected sum of some 3-manifolds, which implies that M' is reducible. Hence, D_Q and Q are reducible. So M is reducible. □

A digon 2-face of Q will give an essential digon 2-suborbifold in Q unless that Q is a trihedron ($S^3/(\mathbb{Z}_2)^3$). Thus in this case Q is reducible as well. Therefore, if there is a henagon 2-suborbifold or an essential digon 2-suborbifold in Q, then Q cannot be hyperbolic.

Next, suppose that Q is not a trihedron and contains no henagon and essential digon 2-suborbifolds. If there are some 2-faces with self-intersection or the intersection of two 2-faces is not connected, then we can always construct some orbifold covers of Q. In fact, we can use some copies of Q to construct a covering space of Q as follows. First we cut open each of copies by using a fixed 2-suborbifold B, and then form a connected handlebody $\widehat{Q}$ by attaching them together along those new facets produced by B. If necessary, we can choose enough copies of Q so as to make sure that this connected handlebody is simple, and is exactly the required covering space of Q. Applying Theorem A gives:

Corollary 7.18 *A 3-handlebody with ideal nerve is hyperbolic if and only if it is not trihedron, tetrahedron and contains no $\triangle^2, \square$-belts and no henagon or essential digon 2-suborbifolds.*

Remark 7.19 Let Q be a 3-handlebody with ideal nerve. We can define henagon 2-suborbifolds and essential digons 2-suborbifolds in Q as 1- and 2-belts of Q, respectively. Then by Lemma 7.16, Q is very good if and only if Q contains no 1-belts. An easy argument gives that a very good Q is flag if and only if it is not trihedron or tetrahedron (i.e. $S^3/(\mathbb{Z}_2)^3$ or $S^3/(\mathbb{Z}_2)^4$) and contains no 2- or 3-belts (i.e. π_1-injective $S^2/(\mathbb{Z}_2)^2$- or $S^2/(\mathbb{Z}_2)^3$-suborbifolds). Furthermore, a very good flag Q is hyperbolic if and only if it contains no 4-belts (i.e. π_1-injective $T^2/(\mathbb{Z}_2)^2$-suborbifolds). Thus, a right-angled Coxeter 3-handlebody with ideal nerve except trihedron and tetrahedron is hyperbolic if and only if it contains no 1-, 2-, 3- or 4-belts.

7.3.3 Example of a nonsimple 3-handlebody

Example 7.20 Let P be the product of a pentagon and $[0, 1]$. Gluing two opposite pentagons of P together such that its diagonal vertices coincide with each other gives a right-angled Coxeter 3-orbifold with its underlying space as a solid torus, denoted by Q. Then Q is a Seifert 3-orbifold. Thus it cannot be hyperbolic. This is because each embedding annulus 2-facet is an obstruction.

7.4 Positive (scalar) curvature

A simple polytope P is *two-neighborly* if any two facets of P has a nonempty intersection. So it is clear that P is two-neighborly if and only if it contains no I-belt. In addition, we also know from [38, Proposition 2.1] that P is two-neighborly if and only if its manifold double M is simply connected. Thus, we have that:

Lemma 7.21 *Let P be a simple polytope. Then the following statements are equivalent.*

- *P is two-neighborly.*
- *P contains no I-belt.*
- *Its manifold double M is simply connected.*

Remark 7.22 Similarly, for a simple handlebody Q, we may define it to be *two-neighborly* if any two vertices in its nerve $\mathcal{N}(Q)$ are connected by an edge. However, if the genus of Q is greater than zero, then we can always find an I-belt between two facets with nonempty intersection such that this I-belt cannot be deformed onto the intersection of two facets in Q. Hence the existence of I-belts cannot be used to detect whether Q is two-neighborly.

Lemma 7.23 *Let N be a triangulable closed n-manifold with $n > 1$. If $\pi_1(N)$ is nontrivial, then the 1-skeleton of any triangulation of N cannot be a complete graph.*

Proof Assume that K is a triangulation of N whose 1-skeleton K^1 is a complete graph. Fixed a vertex x in K, let $N(x)$ be the union of those n-simplices in K which contain x. Then $N(x) \cong D^n$ whose boundary $\partial N(x)$ is a two-neighborly simplicial $(n-1)$-sphere.

Let Δ^n be an arbitrary simplex which does not contain x. Since K^1 is a complete graph, all vertices of Δ^n are contained in $\partial N(x)$. Hence, K^1 is a subcomplex of $N(x)$. So any closed loop in K^1 is contractible in $N(x)$. This means that N is simply connected, giving a contradiction. □

Corollary 7.24 *A (flag) simple handlebody with genus >0 cannot be two-neighborly. In other words, a two-neighborly simple handlebody must be a two-neighborly simple polytope as a manifold with corners.*

Proposition 7.25 (Hopf–Rinow, Myers, [50, Corollary 6.3.2 and Theorem 6.3.3]) *If an n-dimensional closed manifold N admits a complete Riemannian metric of positive sectional curvature, then $\pi_1(N)$ is finite. Specially, $\pi_1(N)$ is 0 or $\mathbb{Z}_2$ if n is even, and N is orientable if n is odd.*

If N admits a complete Riemannian metric of positive Ricci curvature, then $\pi_1(N)$ is finite, too.

For a simple handlebody Q, by the short exact sequence

$$1 \to \pi_1(M) \to \pi_1^{\mathrm{orb}}(Q) \xrightarrow{\lambda} (\mathbb{Z}_2)^m \to 1,$$

if the genus of Q is greater than zero, then any torsion-free generator t determined by a cutting belt is mapped to $1 \in (\mathbb{Z}_2)^m$ via λ. Hence t gives a torsion-free element in $\pi_1(M)$. So we have:

Lemma 7.26 *If the genus of Q is greater than zero, then $\pi_1(M)$ is not finite.*

As a direct consequence of Proposition 7.25 and Lemmas 7.21 and 7.26, we have:

Corollary 7.27 *If M admits a complete Riemannian metric of positive sectional or Ricci curvature, then Q must be a two-neighborly simple polytope, that is, there is no I-belt in Q.*

Conversely, the existence of positive sectional curvature and positive Ricci curvature of a closed (or compact) n-manifold is a very hard question, which is involved in many conjectures and open questions. For example, we know that the real moment angle manifold over $\Delta^2 \times \Delta^2$ is $S^2 \times S^2$. However, it is well known that the existence of positive sectional curvature on $S^2 \times S^2$ is just the Hopf conjecture.

On the other hand, there are some results about the existence of positive scalar curvature. One can refer to some works of Gromov and Lawson [34; 33; 35], Schoen and Yau [55; 56], and Stolz [57]. By Gromov and Lawson [35], a compact manifold of nonpositive sectional curvature cannot carry a metric of positive sectional curvature. So

- if M admits a positive scalar curvature, then Q is not flag.

Moreover, it is reasonable to conjecture that:

Conjecture 7.28 *If a simple polytope Q is two-neighborly, then M admits a positive scalar curvature.*

In the 3-dimensional case, Wu and Yu [61] gave a combinatorial description for the case of real moment-angled manifolds with positive scalar curvature over simple 3-polytopes.

Proposition 7.29 [61, Corollary 4.10] *A real moment-angle manifold (or small cover) over a simple 3-polytope P can admit a Riemannian metric with positive scalar curvature if and only if P is combinatorially equivalent to a polytope obtained from Δ^3 by a sequence of vertex cuts.*

Let P be a simple polytope obtained from Δ^3 by a sequence of vertex cuts. Then except for $P = \Delta^3$, any 2-dimensional belt in P is a Δ^2-belt. Conversely, assume that every 2-dimensional belt is only a Δ^2-belt in a simple polytope P, then it is easy to see that P is a tetrahedron or P can be decomposed into the connected sum of some Δ^3's. This is equivalent to saying that P can be obtained from Δ^3 by a sequence of vertex cuts. Hence, we get an equivalent description of Proposition 7.29 in terms of Δ^2-belts and Δ^3-belts.

Corollary 7.30 *Let P be a simple 3-polytope. Then its manifold double M admits a Riemannian metric with positive scalar curvature if and only if every 2-dimensional belt in P is Δ^2, or P is just a tetrahedron.*

We summarize what we have discussed in Table 1.

M Q	2-neighborly	flag	Pogorelov	description
Sec < 0	not		dim 3	Proposition 7.10
hyperbolic	not		dim 3	dim 3: Proposition 7.12; dim 4: not clear; dim ≥ 5: none
Sec ≤ 0, NPC	not			Proposition 7.6
flat	not		not	[38, Theorem 1.2]
spherical		not	not	[38, Theorem 1.2]
Sec, Ric > 0		not	not	not clear (e.g. Hopf conjecture)
scalar > 0	Conjecture 7.28	not	not	dim 3: [61]; dim > 3: not clear

Table 1: A simple handlebody is called *Pogorelov* if it is flag and contains no $\square$-belt.

References

[1] **A Adem**, **J Leida**, **Y Ruan**, *Orbifolds and stringy topology*, Cambridge Tracts in Math. 171, Cambridge Univ. Press (2007) MR Zbl

[2] **E M Andreev**, *Convex polyhedra of finite volume in Lobačevskiĭ space*, Mat. Sb. 83(125) (1970) 256–260 MR Zbl In Russian; translated in Math. USSR-Sb. 12 (1970) 255–259

[3] **A Bahri**, **M Bendersky**, **F R Cohen**, **S Gitler**, *Decompositions of the polyhedral product functor with applications to moment-angle complexes and related spaces*, Proc. Natl. Acad. Sci. USA 106 (2009) 12241–12244 MR Zbl

[4] **A Bahri**, **D Notbohm**, **S Sarkar**, **J Song**, *On integral cohomology of certain orbifolds*, Int. Math. Res. Not. 2021 (2021) 4140–4168 MR Zbl

[5] **Y Barreto**, **S López de Medrano**, **A Verjovsky**, *Some open book and contact structures on moment-angle manifolds*, Bol. Soc. Mat. Mex. 23 (2017) 423–437 MR Zbl

[6] **F Bosio**, **L Meersseman**, *Real quadrics in $\mathbb{C}^n$, complex manifolds and convex polytopes*, Acta Math. 197 (2006) 53–127 MR Zbl

[7] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR Zbl

[8] **V M Buchstaber**, **N Y Erokhovets**, **M Masuda**, **T E Panov**, **S Pak**, *Cohomological rigidity of manifolds defined by* 3*-dimensional polytopes*, Uspekhi Mat. Nauk 72 (2017) 3–66 MR Zbl In Russian; translated in Russian Math. Surveys 72 (2017) 199–256

[9] **V M Buchstaber**, **T E Panov**, *Toric topology*, Math. Surv. Monogr. 204, Amer. Math. Soc., Providence, RI (2015) MR Zbl

[10] **X Cao**, **Z Lü**, *Möbius transform, moment-angle complexes and Halperin–Carlsson conjecture*, J. Algebraic Combin. 35 (2012) 121–140 MR Zbl

[11] **B Chen**, **Z Lü**, **L Yu**, *Self-dual binary codes from small covers and simple polytopes*, Algebr. Geom. Topol. 18 (2018) 2729–2767 MR Zbl

[12] **W Chen**, *A homotopy theory of orbispaces*, preprint (2001) arXiv math/0102020

[13] **W Chen**, **Y Ruan**, *A new cohomology theory of orbifold*, Comm. Math. Phys. 248 (2004) 1–31 MR Zbl

[14] **S Choi**, **H Park**, *Small covers over wedges of polygons*, J. Math. Soc. Japan 71 (2019) 739–764 MR Zbl

[15] **V I Danilov**, *The geometry of toric varieties*, Uspekhi Mat. Nauk 33 (1978) 85–134 MR Zbl In Russian; translated in Russian Math. Surveys 33 (1978) 97–154

[16] **V I Danilov**, *Birational geometry of three-dimensional toric varieties*, Izv. Akad. Nauk SSSR Ser. Mat. 46 (1982) 971–982 MR In Russian; translated in Math. USSR-Izv. 21 (1983) 269–280

[17] **M W Davis**, *Groups generated by reflections and aspherical manifolds not covered by Euclidean space*, Ann. of Math. 117 (1983) 293–324 MR Zbl

[18] **M W Davis**, *The geometry and topology of Coxeter groups*, Lond. Math. Soc. Monogr. Ser. 32, Princeton Univ. Press (2008) MR Zbl

[19] **M W Davis**, *Lectures on orbifolds and reflection groups*, from "Transformation groups and moduli spaces of curves", Adv. Lect. Math. 16, International, Somerville, MA (2011) 63–93 MR Zbl

[20] **M W Davis**, *Right-angularity, flag complexes, asphericity*, Geom. Dedicata 159 (2012) 239–262 MR Zbl

[21] **M W Davis**, **A L Edmonds**, *Euler characteristics of generalized Haken manifolds*, Algebr. Geom. Topol. 14 (2014) 3701–3716 MR Zbl

[22] **M W Davis**, **T Januszkiewicz**, *Convex polytopes, Coxeter orbifolds and torus actions*, Duke Math. J. 62 (1991) 417–451 MR Zbl

[23] **M W Davis**, **T Januszkiewicz**, *Hyperbolization of polyhedra*, J. Differential Geom. 34 (1991) 347–388 MR Zbl

[24] **M Davis**, **T Januszkiewicz**, **R Scott**, *Nonpositive curvature of blow-ups*, Selecta Math. 4 (1998) 491–547 MR Zbl

[25] **M Davis**, **T Januszkiewicz**, **R Scott**, *Fundamental groups of blow-ups*, Adv. Math. 177 (2003) 115–179 MR Zbl

[26] **B Everitt**, **J G Ratcliffe**, **S T Tschantz**, *Right-angled Coxeter polytopes, hyperbolic six-manifolds, and a problem of Siegel*, Math. Ann. 354 (2012) 871–905 MR Zbl

[27] **S Fomin**, **M Shapiro**, **D Thurston**, *Cluster algebras and triangulated surfaces, I: Cluster complexes*, Acta Math. 201 (2008) 83–146 MR Zbl

[28] **B Foozwell**, **H Rubinstein**, *Introduction to the theory of Haken n-manifolds*, from "Topology and geometry in dimension three", Contemp. Math. 560, Amer. Math. Soc., Providence, RI (2011) 71–84 MR Zbl

[29] **W Fulton**, *Introduction to toric varieties*, Ann. of Math. Stud. 131, Princeton Univ. Press (1993) MR Zbl

[30] **A Garrison**, **R Scott**, *Small covers of the dodecahedron and the* 120*-cell*, Proc. Amer. Math. Soc. 131 (2003) 963–971 MR Zbl

[31] **S Gitler**, **S López de Medrano**, *Intersections of quadrics, moment-angle manifolds and connected sums*, Geom. Topol. 17 (2013) 1497–1534 MR Zbl

[32] **M Gromov**, *Hyperbolic groups*, from "Essays in group theory", Math. Sci. Res. Inst. Publ. 8, Springer (1987) 75–263 MR Zbl

[33] **M Gromov**, **H B Lawson, Jr**, *The classification of simply connected manifolds of positive scalar curvature*, Ann. of Math. 111 (1980) 423–434 MR Zbl

[34] **M Gromov**, **H B Lawson, Jr**, *Spin and scalar curvature in the presence of a fundamental group, I*, Ann. of Math. 111 (1980) 209–230 MR Zbl

[35] **M Gromov**, **H B Lawson, Jr**, *Positive scalar curvature and the Dirac operator on complete Riemannian manifolds*, Inst. Hautes Études Sci. Publ. Math. 58 (1983) 83–196 MR Zbl

[36] **A Hatcher**, *Algebraic topology*, Cambridge Univ. Press (2002) MR Zbl

[37] **M Kapovich**, *Hyperbolic manifolds and discrete groups*, Progr. Math. 183, Birkhäuser, Boston, MA (2001) MR Zbl

[38] **S Kuroki**, **M Masuda**, **L Yu**, *Small covers, infra-solvmanifolds and curvature*, Forum Math. 27 (2015) 2981–3004 MR Zbl

[39] **H B Lawson, Jr**, **S T Yau**, *Compact manifolds of nonpositive curvature*, J. Differential Geom. 7 (1972) 211–228 MR Zbl

[40] **B Leeb**, 3*-manifolds with(out) metrics of nonpositive curvature*, Invent. Math. 122 (1995) 277–289 MR Zbl

[41] **Z Lü**, *Graphs of* 2*-torus actions*, from "Toric topology", Contemp. Math. 460, Amer. Math. Soc., Providence, RI (2008) 261–272 MR Zbl

[42] **Z Lü**, **M Masuda**, *Equivariant classification of* 2*-torus manifolds*, Colloq. Math. 115 (2009) 171–188 MR Zbl

[43] **Z Lü**, **Q Tan**, *Small covers and the equivariant bordism classification of* 2*-torus manifolds*, Int. Math. Res. Not. 2014 (2014) 6756–6797 MR Zbl

[44] **R C Lyndon**, **P E Schupp**, *Combinatorial group theory*, Ergebnisse der Math. 89, Springer (1977) MR Zbl

[45] **G D Mostow**, *Strong rigidity of locally symmetric spaces*, Ann. of Math. Stud. 78, Princeton Univ. Press (1973) MR Zbl

[46] **G Moussong**, *Hyperbolic Coxeter groups*, PhD thesis, Ohio State University (1988) Available at `https://www.proquest.com/docview/303686972`

[47] **H Nakayama**, **Y Nishimura**, *The orientability of small covers and coloring simple polytopes*, Osaka J. Math. 42 (2005) 243–256 MR Zbl

[48] **D Notbohm**, *Colorings of simplicial complexes and vector bundles over Davis–Januszkiewicz spaces*, Math. Z. 266 (2010) 399–405 MR Zbl

[49] **J-P Otal**, *Thurston's hyperbolization of Haken manifolds*, from "Surveys in differential geometry, III", International, Boston, MA (1998) 77–194 MR Zbl

[50] **P Petersen**, *Riemannian geometry*, 3rd edition, Graduate Texts in Math. 171, Springer (2016) MR Zbl

[51] **M Poddar**, **S Sarkar**, *On quasitoric orbifolds*, Osaka J. Math. 47 (2010) 1055–1076 MR Zbl

[52] **A V Pogorelov**, *A regular partition of Lobachevskian space*, Mat. Zametki 1 (1967) 3–8 MR Zbl In Russian; translated in Math. Notes 1 (1967) 3–5

[53] **R K W Roeder**, **J H Hubbard**, **W D Dunbar**, *Andreev's theorem on hyperbolic polyhedra*, Ann. Inst. Fourier (Grenoble) 57 (2007) 825–882 MR Zbl

[54] **I Satake**, *The Gauss–Bonnet theorem for V-manifolds*, J. Math. Soc. Japan 9 (1957) 464–492 MR Zbl

[55] **R Schoen**, **S T Yau**, *Existence of incompressible minimal surfaces and the topology of three-dimensional manifolds with nonnegative scalar curvature*, Ann. of Math. 110 (1979) 127–142 MR Zbl

[56] **R Schoen**, **S T Yau**, *On the structure of manifolds with positive scalar curvature*, Manuscripta Math. 28 (1979) 159–183 MR Zbl

[57] **S Stolz**, *Simply connected manifolds of positive scalar curvature*, Ann. of Math. 136 (1992) 511–540 MR Zbl

[58] **W P Thurston**, *Three-dimensional geometry and topology, I*, Princeton Math. Ser. 35, Princeton Univ. Press (1997) MR Zbl

[59] **È B Vinberg**, **O V Shvartsman**, *Discrete groups of motions of spaces of constant curvature*, from "Geometry, II", Encycl. Math. Sci. 29, Springer (1993) 139–248 MR Zbl

[60] **F Waldhausen**, *On irreducible* 3*-manifolds which are sufficiently large*, Ann. of Math. 87 (1968) 56–88 MR Zbl

[61] **L Wu**, **L Yu**, *Fundamental groups of small covers revisited*, Int. Math. Res. Not. 2021 (2021) 7262–7298 MR Zbl

School of Mathematical Sciences, Fudan University
Shanghai, China

College of Mathematics and Sciences, Shandong University of Science and Technology
Qingdao, China

zlu@fudan.edu.cn, wulisu@sdust.edu.cn

Received: 6 March 2022 Revised: 2 July 2023

Algebraic & Geometric Topology **25**:1 (2025) 107–159
DOI: 10.2140/agt.2025.25.107
Published: 5 March 2025

Property (QT) for 3-manifold groups

SUZHEN HAN
HOANG THANH NGUYEN
WENYUAN YANG

According to Bestvina, Bromberg and Fujiwara, a finitely generated group is said to have property (QT) if it acts isometrically on a finite product of quasitrees so that orbital maps are quasi-isometric embeddings. We prove that the fundamental group $\pi_1(M)$ of a compact, connected, orientable 3-manifold M has property (QT) if and only if no summand in the sphere-disc decomposition of M supports either Sol or Nil geometry. In particular, all compact, orientable, irreducible 3-manifold groups with nontrivial torus decomposition and not supporting Sol geometry have property (QT). In the course of our study, we establish property (QT) for the class of Croke–Kleiner admissible groups and for relatively hyperbolic groups under natural assumptions on the peripheral subgroups.

20F65, 20F67

1 Introduction

1.1 Background and motivation

The study of group actions on quasitrees has recently received a great deal of interest. A *quasitree* means here a possibly locally infinite connected graph that is quasi-isometric to a simplicial tree. Groups acting on (simplicial) trees have been well understood thanks to Bass–Serre theory. On the one hand, quasitrees have the obvious advantage of being more flexible; hence, many groups can act unboundedly on quasitrees but act on any trees with global fixed points. Many hyperbolic groups with Kazhdan's property (T) and mapping class of groups are among many examples that belong to this category (see Manning [38; 39] for other examples). In effect, these are sample applications of a powerful axiomatic construction of quasitrees proposed in the work of Bestvina, Bromberg and Fujiwara [5]. This construction will be fundamental in this paper.

We say that a finitely generated group G has *property* (*QT*) if it acts isometrically on a finite product $X = T_1 \times T_2 \times \cdots \times T_n$ of quasitrees with the L^2-metric such that for any basepoint $o \in X$, the induced orbit map

$$g \in G \mapsto go \in X$$

is a quasi-isometric embedding of G equipped with some (or any) word metric d_G to X. Informally speaking, property (QT) gives an undistorted picture of the ambient group in a reasonably good space.

Here, the direct product structure usually comes from the independence of several negatively curved layers endowed on the group. Such a hierarchy structure has emerged from the study of mapping class groups since Masur and Minsky [40]. In addition, property (QT) is a commensurability invariant as observed by Bestvina, Bromberg and Fujiwara [6] and Button [14], and could be thought of as a stronger property than the finiteness of asymptotic dimension.

Extending their earlier results of [5], Bestvina, Bromberg and Fujiwara [6] recently showed that residually finite hyperbolic groups and mapping class groups have property (QT). It is known that Coxeter groups have property (QT) (see Dranishnikov and Januszkiewicz [23]), and thus every right-angled Artin group has property (QT) (see [6, Induction 2.2]).

In 3-manifold theory, the study of the fundamental groups of 3-manifolds is one of the most central topics. Determining property (QT) of finitely generated 3-manifold groups is the main task of the present paper.

1.2 Property (QT) of 3-manifold groups

Let M be a 3-manifold with finitely generated fundamental group. Since property (QT) is a commensurability invariant, we can assume that M is compact and orientable by considering the Scott core of M and a double cover of M (if M is nonorientable).

In recent years, the theory of special cube complexes — see Haglund and Wise [30] — has led to a deep understanding of 3-manifold groups; see Agol [3] and Wise [53]. By definition, the fundamental group of a compact special cube complex is undistorted in a right-angled Artin group, and then has property (QT) by [23]. However, 3-manifolds without nonpositively curved Riemannian metrics cannot be cubulated by Przytycki and Wise [46] and certain cubulated 3-manifold groups are not virtually compact special (see Hagen and Przytycki [28] and Tidmore [51]). Thus it was left still open to determine the property (QT) for all 3-manifold groups.

By the sphere-disc decomposition, a compact oriented 3-manifold M is a connected sum of prime summands M_i $(1 \le i \le n)$ with incompressible boundary. It is an easy observation that if a group has property (QT) then every nontrivial element is undistorted (see Lemma 2.5), and hence if M_i supports Sol or Nil geometry from the eight Thurston geometries, then $\pi_1(M_i)$ fails to have property (QT). Our first main result is the following characterization of property (QT) for all 3-manifold groups.

Theorem 1.1 *Let M be a connected, compact, orientable 3-manifold. Then $\pi_1(M)$ has property (QT) if and only if no summand in its sphere-disk decomposition supports either Sol or Nil geometry.*

By standard arguments, we are reduced to the case where M is a compact, connected, orientable, irreducible 3-manifold with empty or tori boundary. Theorem 1.1 actually follows from the following theorem.

Theorem 1.2 *Let M be a compact orientable irreducible 3-manifold with empty or tori boundary, and with nontrivial torus decomposition which does not support the Sol geometry. Then $\pi_1(M)$ has property (QT).*

A 3-manifold M as in Theorem 1.2 is called a *graph manifold* if all the pieces in its torus decomposition are Seifert fibered spaces; otherwise M is called a *mixed manifold*. It is well known that the fundamental group of a mixed 3-manifold is hyperbolic relative to a collection of abelian groups and graph manifold groups. As a result, to prove Theorem 1.2, we actually only need to determine the property (QT) of *Croke–Kleiner admissible groups*, and of relatively hyperbolic groups, which will be discussed in detail in the following subsections. These results include but are much more general than the fundamental groups of graph manifolds and mixed manifolds.

1.3 Property (QT) of Croke–Kleiner admissible groups

We first address property (QT) of graph manifolds. Our approach is based on a study of a particular class of graph of groups introduced by Croke and Kleiner [21] which they called *admissible groups* and generalized the fundamental groups of graph manifolds. We say that an admissible group G is a *Croke–Kleiner admissible group* or a *CKA group* if it acts properly discontinuously, cocompactly and by isometries on a complete proper CAT(0) space X. Such an action $G \curvearrowright X$ is called a *CKA action* and the space X is called a *CKA space*. The CKA action is modeled on the JSJ structure of graph manifolds where the Seifert fibration is replaced by the following central extension of a general hyperbolic group H_v:

$$1 \to Z(G_v) \to G_v \to H_v \to 1 \tag{1}$$

where $Z(G_v) = \mathbb{Z}$. It is worth pointing out that CKA groups encompass a much more general class of groups and can be used to produce interesting groups by a "flip" trick from any finite number of hyperbolic groups (see Example 2.14).

The notion of an *omnipotent* group was introduced by Wise [52] and has found many applications in subgroup separability. We refer the reader to Definition 4.6 for its definition and note here that free groups [52], surface groups (see Bajpai [4]), and the more general class of virtually special hyperbolic groups [53] are omnipotent. Nguyen and Yang [43] proved property (QT) for a special class of CKA actions under flip conditions (see Definition 2.18). One of the main contributions of this paper is to remove this assumption and prove the following result in full generality.

Theorem 1.3 *Let $G \curvearrowright X$ be a CKA action where for every vertex group the central extension* (1) *has omnipotent hyperbolic quotient group. Then G has property (QT).*

Remark 1.4 It is a long-standing problem whether every hyperbolic group is residually finite. Wise [52, Remark 3.4] noted that if every hyperbolic group is residually finite, then any hyperbolic group is omnipotent.

Let us comment on the relation between this work and the previous one [43]. As in [43], the ultimate goal is to utilize Bestvina, Bromberg and Fujiwara's projection complex machinery to obtain actions on quasitrees. The common starting point is the class of special paths developed in [43] that record the distances of X. However, the flip assumption (see Definition 2.18) on CKA actions was crucially used there: the fiber lines coincide with boundary lines in adjacent vertex pieces when crossing the boundary plane, roughly speaking. Hence, a straightforward gluing construction works in that case but fails in our general setting. In this paper, we use a completely different projection system to achieve the same purpose with a more delicate analysis.

It is worth mentioning the following fact frequently invoked by many authors: the fundamental group of any graph manifold is quasi-isometric to the fundamental group of some flip manifold as defined by Kapovich and Leeb [34]. This simplification, however, is useless to address property (QT), as such a quasi-isometry does not respect the group actions. Conversely, our direct treatment of any graph manifolds (closed or with nonempty boundary) is new, and we believe it will potentially allow for further applications.

We now explain how we apply Theorem 1.3 to graph manifolds. If M is a graph manifold with nonempty boundary then it always admits a Riemannian metric of nonpositive curvature (see Leeb [35]). In particular, $\pi(M) \curvearrowright \widetilde{M}$ is a CKA action, and thus property (QT) of $\pi_1(M)$ follows immediately from Theorem 1.3. However, closed graph manifolds may not support any Riemannian metric of nonpositive curvature [35], so property (QT) in this case does not follow immediately from Theorem 1.3. We have to make certain modifications on some steps to run the proof of Theorem 1.3 for the fundamental groups of closed graph manifolds (see Section 8.2.1 for details).

1.4 Property (QT) of relatively hyperbolic groups

When M is a mixed 3-manifold, $\pi_1(M)$ is hyperbolic relative to the finite collection $\mathcal{P}$ of fundamental groups of maximal graph manifold components, isolated Seifert components, and isolated JSJ tori (see Bigdely and Wise [8] and Dahmani [22]). Therefore, we need to study property (QT) for relatively hyperbolic groups.

Our main result in this direction is a characterization of property (QT) for residually finite relatively hyperbolic groups, which generalizes the corresponding results of [6] on Gromov-hyperbolic groups.

Theorem 1.5 *Suppose that a finitely generated group H is hyperbolic relative to a finite set of subgroups $\mathbb{P}$. Assume that each $P \in \mathbb{P}$ acts by isometry on finitely many quasitrees T_i $(1 \le i \le n_P)$ such that the induced diagonal action on $\prod_{i=1}^{n_P} T_i$ has property (QT). If H is residually finite, then H has property (QT).*

Remark 1.6 Since maximal parabolic subgroups are undistorted, each $P \in \mathcal{P}$ obviously has property (QT) if G has property (QT). A nonequivariant version of this result was proven by Mackay and Sisto [37].

Remark 1.7 It is well known that mixed 3-manifold groups $G = \pi_1(M)$ are hyperbolic relative to a collection $\mathbb{P}$ of abelian groups and graph manifold groups $P = \pi_1(M_i)$. However, it is still insufficient to derive directly via Theorem 1.5 the property (QT) of G from that of graph manifold groups P asserted in Theorem 1.3, since P may not preserve factors in the finite product of quasitrees. Of course, passing to an appropriate finite-index subgroup $P' < P$ preserves the factors, but it is not clear at all whether P' are peripheral subgroups of a finite-index subgroup G' of G. In order to find such a G', a stronger assumption must be satisfied so that every finite-index subgroup of each P is separable in G. This requires the notion of a *full profinite topology* induced on subgroups (see the precise definition before Theorem 3.5 and a relevant discussion of Reid [47]). See Theorem 3.5 for the precise statement. In the setting of a mixed 3-manifold, Lemma 8.5 verifies that each peripheral subgroup $P \in \mathbb{P}$ of $\pi_1(M)$ satisfies this assumption. Therefore, all mixed 3-manifolds are proven to have property (QT).

We now explain a few algebraic and geometric consequences for groups with property (QT).

Similar to trees, any isometry on quasitrees must be either elliptic or loxodromic [38]. Hence, if a finitely generated group acts properly (in a metric sense) on a finite product of quasitrees, then every nontrivial element is undistorted (Lemma 2.5). Moreover, property (QT) allows one to characterize virtually abelian groups among subexponential growth groups and solvable groups.

Theorem 1.8 *Let G be a finitely generated group.*

(1) *Assume that G has subexponential growth. Then G has property (QT) if and only if G is virtually abelian.*
(2) *Suppose that G is solvable with finite virtual cohomological dimension. Then G has property (QT) if and only if it is virtually abelian.*

By Theorem 1.5, this yields as a consequence that nonuniform lattices in $\mathrm{SU}(n,1)$ and $\mathrm{Sp}(n,1)$ for $n \geq 2$ fail to act properly on finite products of quasitrees.

Corollary 1.9 *A nonuniform lattice in $\mathrm{SU}(n,1)$ for $n \geq 2$ or $\mathrm{Sp}(n,1)$ for $n \geq 1$ does not have property (QT), while any lattice of $\mathrm{SO}(n,1)$ has property (QT) for $n \geq 2$.*

Overview

The paper is structured as follows. In Section 2, we recall the preliminary materials about Croke–Kleiner admissible groups, axiomatic constructions of quasitrees, and we collect a few preliminary observations to prove Theorem 1.8 and to disprove property (QT) for the fundamental groups of 3-manifolds with Sol or Nil geometry. Section 3 contains a proof of Theorem 1.5 and its variant Theorem 3.5. The next four sections aim to prove Theorem 1.3: Section 4 first recalls a cone-off construction of CKA actions from [43] and then outlines the steps executed in Sections 5, 6, and 7 to prove property (QT) for CKA actions. Sections 5 and 6 explain in detail the construction of projection systems of fiber lines and then prove the corresponding distance formula. We finish the proof of Theorem 1.3 in Section 7. In Section 8, we present the applications of the previous results for 3-manifold groups and prove Theorem 1.2 as well as Theorem 1.1.

Acknowledgements

Nguyen was partially supported by Project ICRTM04_2021.07 of the International Centre for Research and Postgraduate Training in Mathematics, Vietnam. Yang was supported by National Key R&D Program of China (SQ2020YFA070059) and National Natural Science Foundation of China (12131009 and 12326601). We are also grateful to the referee for many very helpful comments.

2 Preliminaries

This section reviews concepts property (QT), Croke–Kleiner admissible actions, and the construction of quasitrees. Several observations are made to determine property (QT) of the fundamental groups of 3-manifolds with Sol or Nil geometry. This includes the fact that every element is undistorted in groups with property (QT) and some attempts to characterize by property (QT) the class of virtually abelian groups in solvable/subexponential growth groups.

In the sequel, we use the notion $a \preceq_K b$ if the exists $C = C(K) > 0$ such that $a \leq Cb + C$, and $a \sim_K b$ if $a \preceq_K b$ and $b \preceq_K a$. Also, when we write $a \asymp_K b$ we mean that $a/C \leq b \leq Ca$. If the constant C is universal from context, the subindex $\preceq_K$ shall be omitted.

2.1 Property (QT)

Definition 2.1 We say that a finitely generated group G has *property* (*QT*) if it acts isometrically on a finite product $X = T_1 \times T_2 \times \cdots \times T_n$ of quasitrees with L^2-metric such that for any basepoint $o \in X$, the induced orbit map
$$g \in G \mapsto go \in X$$
is a quasi-isometric embedding of G equipped with some (or any) word metric d_G to X with the product metric d.

Remark 2.2 A group with property (QT) acts properly on a finite product of quasitrees in a metric sense: $d(o, go) \to \infty$ as $d_G(1, g) \to \infty$. We would emphasize that all consequences of the property (QT) in the present paper use merely the existence of a metric proper action.

By definition, a quasitree is assumed to be a graph quasi-isometric to a simplicial tree. This does not lose generality as any geodesic metric space (with an isometric action) is quasi-isometric to a graph (with an equivariant isometric action) by taking the 1-*skeleton of its Rips complex*: the vertex set consists of all points and two points with distance less than 1 are connected by an edge.

The first part of the following lemma allows one to pass to finite-index subgroups in the study of property (QT) of groups, as explained in [6, Section 2.2]. The second part of Lemma 2.3 is an immediate consequence of the definition of property (QT).

Lemma 2.3 (1) *Let $H \leq G$ be a finite-index subgroup of G. Then G has property (QT) if and only if H has property (QT).*

(2) *Let $H \leq G$ be an undistorted subgroup of G. Suppose that G has property (QT), then H also has property (QT).*

Below is a corollary of the de Rham decomposition theorem [26, Theorem 1.1] which will be utilized in the subsequent discussions.

Corollary 2.4 *A finite product $X = T_1 \times T_2 \times \cdots \times T_n$ of quasitrees must have de Rham decomposition*

$$X = \mathbb{R}^k \times T_{k+1} \times \cdots \times T_n$$

if the first k quasitrees $(k \geq 0)$ are all real lines among $\{T_i \mid 1 \leq i \leq n\}$.

A finite product $\prod_{i=1}^n T_i$ of quasitrees has no $\mathbb{R}$*-factor* if no T_i is isometric to $\mathbb{R}$ or a point. In this case, the Euclidean factor $\mathbb{R}^k$ will disappear. In what follows, we present some general results about groups with property (QT).

Lemma 2.5 *Assume that G has property (QT). Then the subgroup generated by an element $g \in G$, is undistorted in G.*

Proof Let $X = \mathbb{R}^k \times T_{k+1} \times \cdots \times T_n$ be the de Rham decomposition of a finite product of quasitrees. By [26, Corollary 1.3], up to passage to finite-index subgroups, G acts by isometries on each factor $\mathbb{R}^k$, and T_i for $k+1 \leq i \leq n$. Let $g \in G$ be an infinite order element. If the image of g is an isometry on the Euclidean space $\mathbb{R}^k$, then it either fixes a point or preserves an axis. If the image of g is an isometry on a quasitree T_i then by [39, Corollary 3.2], it has either a bounded orbit or a quasi-isometrically embedded orbit.

Fix a basepoint $o = (o_k, o_{k+1}, \ldots, o_n) \in X$. If the action of G on X is proper, then by the first paragraph, there must exist an unbounded action of $\langle g \rangle$ on some factor $Y = \mathbb{R}^k$ or $Y = T_i$, so we have $m \leq \lambda |o_k - g^m o_k|_Y + c$ for some $\lambda, c > 0$. Since every isometric orbital map is Lipschitz, we have $|o - g^m o|_X \leq C|1 - g^m|_G$ for some $C > 0$. Noting that $|o - g^m o|_Y \leq |o - g^m o|_X$, we have that the map $m \mapsto g^m$ is a quasi-isometric embedding of $\langle g \rangle \cong \mathbb{Z}$ into G. □

Note that the Sol group embeds quasi-isometrically into a product of two hyperbolic planes (for example, see [19, Section 9]). However, the Sol lattice contains exponentially distorted elements by [41, Lemma 5.2]; as a consequence, we have the following:

Corollary 2.6 *The fundamental group of a 3-manifold with Sol geometry does not have property (QT).*

Corollary 2.7 *The Baumslag–Solitar group* $\mathrm{BS}(1,n)$ *for* $n > 1$ *does not have property (QT).*

2.2 Subexponential growth and solvable groups with property (QT)

The fundamental group of a 3-manifold M with Nil geometry also fails to have property (QT) since it contains quadratically distorted elements (for example, see [41, Proposition 1.2]). Generalizing results about property (QT) of 3-manifolds with Sol or Nil geometry, in the rest of this subsection, we provide a characterization of subexponential growth groups and solvable groups with property (QT) and give the proof of Theorem 1.8.

In the next results, we apply the general conclusions in [17] about the isometric actions on hyperbolic spaces to the actions on quasitrees. By Gromov, unbounded isometric group actions can be classified into the following four types:

(1) *horocyclic* if it has no loxodromic element;
(2) *lineal* if it has a loxodromic element and any two loxodromic elements have the same fixed points in the Gromov boundary;
(3) *focal* if it has a loxodromic element which is not lineal, and any two loxodromic elements have one common fixed point;
(4) *general type* if it has two loxodromic elements without common fixed point.

Proposition 2.8 *Assume that* G *has property (QT). Then there exists a finite-index subgroup* $\dot G$ *of* G *which acts on a Euclidean space* $\mathbb{R}^k$ *with* $k \ge 0$ *and finitely many quasitrees* T_i *for* $1 \le i \le n$ *with lineal or focal or general type action such that the orbital map of* $\dot G$ *into* $\mathbb{R}^k \times \prod_{i=1}^n T_i$ *is a quasi-isometric embedding.*

Moreover, the action on each T_i *can be chosen to be cobounded.*

Proof By Corollary 2.4, the finite product of quasitrees given by property (QT) has the above form of de Rham decomposition. By [26, Corollary 1.3],

$$1 \to \mathrm{Isom}(\mathbb{R}^k) \times \prod_{i=k+1}^{n} \mathrm{Isom}(Y_i) \to \mathrm{Isom}(X) \to F \to 1$$

where F is a subgroup of the permutation group on the indices $\{k+1,\dots,n\}$. Thus, there exists a finite-index subgroup $\dot G$ of G acting on each de Rham factor such that $\dot G \subset \mathrm{Isom}(\mathbb{R}^k) \times \prod_{i=1}^n \mathrm{Isom}(Y_i)$ for $k \ge 0$ and $i \ge k+1$.

First of all, we can assume that the actions of $\dot G$ on $\mathbb{R}^k$ and each T_i is unbounded. Otherwise, we can remove $\mathbb{R}^k$ or T_i with bounded actions from the product without affecting the property (QT).

We now consider the action on T_i for $k+1 \le i \le n$. We then need to verify that the action of $\dot{G}$ on T_i cannot be horocyclic. By way of contradiction, assume that the action of $\dot{G}$ on given T_i is horocyclic.

Note that the proof of [17, Proposition 3.1] shows that the intersection of any orbit of $\dot{G}$ on T_i with any quasigeodesic is bounded. By [39, Corollary 3.2], any isometry on a quasitree T_i has either bounded orbits or a quasigeodesic orbit. Thus, we conclude that any orbit of $\langle h\rangle$ for every $h \in \dot{G}$ on T_i is bounded. We are then going to prove that the action of $\dot{G}$ on T_i has bounded orbits. This is a well-known fact and we present the proof for completeness.

By δ-hyperbolicity of T_i, each $h \in \dot{G}$ (with bounded orbits) has a *quasicenter* $c_h \in T_i$: there exists a constant $D > 0$ depending only on δ such that $|c_h - h^i c_h|_{T_i} \le D$ for $i \in \mathbb{Z}$. Moreover, for any $x \in c_h$ and any $y \in T_i$, the Gromov product $\langle y, hy\rangle_x$ is bounded by a constant C depending only on D. As a consequence, the union Z of quasicenters $\{c_h \mid h \in \dot{G}\}$ has finite diameter. Indeed, note that $\langle y, h_1 y\rangle_x$ and $\langle x, h_2^{-1} x\rangle_y$ are bounded by C for any $x \in c_{h_1}$ and $y \in c_{h_2}$. If there exist two elements, h_1 and h_2, such that the distance $|c_{h_1} - c_{h_2}|_{T_i}$ is sufficiently large relative to C, then the piecewise geodesic path connecting points $(h_1 h_2)^n x$ for $n \in \mathbb{Z}$ would be a sufficiently long local quasigeodesic, so it is a global quasigeodesic. By the previous paragraph, we obtain a contradiction, so the $\dot{G}$-invariant set Z is bounded. Since the action on T_i is assumed to be unbounded, we thus proved that the action on T_i cannot be horocyclic.

At last, it remains to prove the "moreover" statement. By Manning's bottleneck criterion [39], any geodesic is contained in a uniform neighborhood of every path with the same endpoints. Thus, any connected subgraph of a quasitree is uniform quasiconvex and thus is a uniform quasitree. Since G is a finitely generated group, by taking the image of the Cayley graph, we can thus construct a connected subgraph on each quasitree T_i such that the action on the subgraph (quasitree) is cobounded. □

We are able to characterize subexponential groups with property (QT) as follows.

Proposition 2.9 *Let G be a finitely generated group with subexponential growth. Then G has property (QT) if and only if G is virtually abelian.*

Proof We first observe that $\mathbb{R}^k$ in Proposition 2.8 can be replaced by a finite product of real lines. Indeed, consider the action of $\dot{G}$ on Euclidean space $\mathbb{R}^k$. By assumption, $\dot{G}$ is of subexponential growth. It is well known that the growth of any finitely generated group dominates that of quotients, so the image $\Gamma \subset \mathrm{Isom}(\mathbb{R}^k)$ of $\dot{G}$ acting on $\mathbb{R}^k$ has subexponential growth. Since finitely generated linear groups do not have intermediate growth, Γ must be virtually nilpotent. It is well known that virtually nilpotent subgroups in $\mathrm{Isom}(\mathbb{R}^k)$ must be virtually abelian. Thus, Γ contains a finite-index subgroup $\mathbb{Z}^l$ for $1 \le l \le k$. By taking the preimage of $\mathbb{Z}^l$ in $\dot{G}$, we can assume further that $\dot{G}$ acts on $\mathbb{R}^k$ through $\mathbb{Z}^l$. It is clear that $\mathbb{Z}^l$ acts on l real lines $\mathbb{R}_1, \mathbb{R}_2, \dots, \mathbb{R}_l$ such that the product action is geometric. We thus replace $\mathbb{R}^k$ by the product $\prod_{1 \le i \le l} \mathbb{R}_i$ where $\dot{G}$ admits a lineal action on each $\mathbb{R}_i$ by translation.

By Proposition 2.8 the action of $\dot{G}$ on T_i is either lineal or focal or general type. In the latter two cases, $\dot{G}$ contains a free (semi)group by [17, Lemma 3.3], contradicting the subexponential growth of $\dot{G}$. Thus, the action of $\dot{G}$ on each T_i is lineal. By Proposition 2.8, we can assume that T_i is a quasiline.

By [39, Lemma 3.7], a quasiline T admits a $(1, C)$-quasi-isometry ϕ (with a quasi-inverse ψ) to $\mathbb{R}$ for some $C > 0$. A lineal action of G on T is then conjugated to a quasiaction of G on $\mathbb{R}$ sending $g \in G$ to a $(1, C')$-quasi-isometry $\phi g \psi$ on $\mathbb{R}$ for some $C' = C'(C) > 0$. By taking an index at most 2 subgroup, we can assume that every element in G fixes pointwise the two ends of T. Note that a $(1, C')$-quasi-isometry $\phi g \psi$ on $\mathbb{R}$ fixing the two ends of $\mathbb{R}$ is uniformly bounded away from a translation on $\mathbb{R}$. So, for any $x \in \mathbb{R}$, the orbital map $g \mapsto \phi g \psi(x)$ is a quasimorphism $G \to \mathbb{R}$. It is well known that for any amenable group, any quasimorphism must be a homomorphism up to bounded error. We conclude that any $[G, G]$-orbit on T stays in a bounded set.

Therefore, any $[G, G]$-orbit on $\left(\prod_{1\le i\le l} \mathbb{R}_i\right) \times \left(\prod_{1\le i\le n} T_i\right)$ is bounded, so the proper action on X implies that $[\dot{G}, \dot{G}]$ is a finite group. It is well known that if a group has a finite commutator subgroup, then it is virtually abelian [11, Lemma II.7.9]. $\square$

It would be interesting to ask whether Proposition 2.9 holds within the class of solvable groups. In Proposition 2.11 below, we are able to give a positive answer to the previous question when the solvable group has finite virtual cohomological dimension. To this end, we need the following fact.

Lemma 2.10 *Any unbounded isometric action of a meta-abelian group on a quasitree must be lineal.*

Recall that a meta-abelian group is a group whose commutator subgroup is abelian.

Proof Indeed, the abelian group $\Gamma = [G, G]$ (of possibly infinite rank) cannot contain free semigroups, so by [17, Lemma 3.3], the action of Γ on a quasitree T must be bounded or lineal.

Assume first that Γ has a bounded orbit K in T. Since G/Γ is abelian, we have that $g^m h^n K = h^n g^m K$ for any $n, m \in \mathbb{Z}$ and $g, h \in G$, and thus $gh^n K = h^n gK$ has finite Hausdorff distance to $h^n K$ for any $n \in \mathbb{Z}$. Assume that g and h are loxodromic. Then $\{h^n K, n \in \mathbb{Z}\}$ is quasi-isometric to a line. Hence, we obtain that the fixed points of g and h at the Gromov boundary must coincide. This means the action of G on T is lineal.

In the lineal case, Γ preserves some bi-infinite quasigeodesic γ up to finite Hausdorff distance. Since Γ is a normal subgroup in G, we see that every loxodromic element in G also preserves γ up to a finite Hausdorff distance. Thus, the action of G on T is also lineal. $\square$

By Lemma 2.5, a group with property (QT) is *translation proper* in the sense of Conner [18]: the translation length of any nontorsion element is positive. If G is solvable and has finite virtual cohomological dimension, then Conner shows that G is virtually meta-abelian.

Proposition 2.11 *Suppose that a solvable group G has finite virtual cohomological dimension. If G has property (QT) then it is virtually abelian.*

Proof Passing to finite-index subgroups, assume that G is meta-abelian so any quotient of G is meta-abelian. By Lemma 2.10, the action of G on each T_i is lineal.

After possibly passing to an index 2 subgroup, a lineal action of any amenable group G on a quasiline T can be quasiconjugated to an isometric action on $\mathbb{R}$. Indeed, by the proof of Proposition 2.9, conjugating the original action by almost isometries gives a quasiaction of G on $\mathbb{R}$ such that any orbital map induces a quasimorphism of G to $\mathbb{R}$. For amenable groups, any quasimorphism differs from a homomorphism by a uniformly bounded constant. Thus, up to quasiconjugacy, the lineal action of G on T can be promoted to become an isometric action on $\mathbb{R}$.

Consequently, we can quasiconjugate the action of a solvable group G on a finite product of quasitrees to a proper action on a Euclidean space. Thus, G must be virtually abelian. □

Proof of Theorem 1.8 The proof is a combination of Propositions 2.9 and 2.11. □

2.3 CKA groups

Admissible groups, first introduced in [21], are a particular class of graph of groups that includes fundamental groups of 3-dimensional graph manifolds. In this section, we review admissible groups and their properties that will used throughout the paper.

Let $\mathcal{G}$ be a connected graph. We often consider oriented edges from e_- to e_+ and write $e = [e_-, e_+]$. Then $\bar{e} = [e_+, e_-]$ denotes the oriented edge with reversed orientation. Denote by $\mathcal{G}^0$ the set of vertices and by $\mathcal{G}^1$ the set of all oriented edges.

Definition 2.12 A graph of groups $\mathcal{G}$ is *admissible* if the following hold:

1. $\mathcal{G}$ is a finite graph with at least one edge.
2. Each vertex group G_v has center $Z(G_v) \cong \mathbb{Z}$, $H_v := G_v/Z(G_v)$ is a nonelementary hyperbolic group, and every edge subgroup G_e is isomorphic to $\mathbb{Z}^2$.
3. Let e_1 and e_2 be distinct directed edges entering a vertex v, and for $i = 1, 2$, let $K_i \subset G_v$ be the image of the edge homomorphism $G_{e_i} \to G_v$. Then for every $g \in G_v$, gK_1g^{-1} is not commensurable with K_2, and for every $g \in G_v - K_i$, gK_ig^{-1} is not commensurable with K_i.
4. For every edge group G_e, if $\alpha_i : G_e \to G_{v_i}$ is the edge monomorphism, then the subgroup generated by $\alpha_1^{-1}(Z(G_{v_1}))$ and $\alpha_2^{-1}(Z(G_{v_1}))$ has finite index in G_e.

A group G is *admissible* if it is the fundamental group of an admissible graph of groups.

Definition 2.13 We say that an admissible group G is a *Croke–Kleiner admissible group* or *CKA group* if it acts properly discontinuously, cocompactly and by isometries on a complete proper CAT(0) space X. Such an action $G \curvearrowright X$ is called a *CKA action* and the space X is called a *CKA space*.

Example 2.14 (1) Let M be a nongeometric graph manifold that admits a nonpositively curved metric. Lift this metric to the universal cover $\widetilde{M}$ of M, and denote it by d. Then the action $\pi_1(M) \curvearrowright (\widetilde{M}, d)$ is a CKA action.

(2) Let T be the torus complexes constructed in [20]. Then $\pi_1(T) \curvearrowright \widetilde{T}$ is a CKA action.

(3) One may build Croke–Kleiner admissible groups algebraically from any finite number of hyperbolic CAT(0) groups. The following example is for $n = 2$ but the same principle works for any $n \geq 2$. Let H_1 and H_2 be two torsion-free hyperbolic groups that act geometrically on CAT(0) spaces X_1 and X_2 respectively. Then $G_i = H_i \times \langle t_i \rangle$ (with $i = 1, 2$) acts geometrically on the CAT(0) space $Y_i = X_i \times \mathbb{R}$. Any primitive hyperbolic element h_i in H_i gives rise to a totally geodesic torus T_i in the quotient space Y_i/G_i with basis $([h_i], [t_i])$. We rescale Y_i so that the translation length of h_i is equal to that of t_i for each i. Let $f : T_1 \to T_2$ be a *flip* isometry respecting these lengths, that is, an orientation-reversing isometry mapping $[h_1]$ to $[t_2]$ and $[t_1]$ to $[h_2]$. Let M be the space obtained by gluing Y_1 to Y_2 by the isometry f. There is a metric on the space M that gives rise to a locally CAT(0) space (see eg [11, Proposition II.11.6]). By the Cartan–Hadamard theorem, the universal cover $\widetilde{M}$ with the induced length metric from M is a CAT(0) space. Let G be the fundamental group of M. Then the action $G \curvearrowright \widetilde{M}$ is geometric, and G is an example of a Croke–Kleiner admissible group.

Remark 2.15 All graph 3-manifold groups are admissible, but there are closed graph 3-manifold groups that are not CAT(0) groups (see [33]), and thus are not CKA groups. The following is another example. Take two nonvirtually split central extensions of hyperbolic groups by $\mathbb{Z}$ (eg $\widetilde{\mathrm{SL}(2, \mathbb{R})}$ lattices) and amalgamate them over $\mathbb{Z}^2$ to obtain an admissible group. This group cannot act properly on CAT(0) spaces, since central extensions acting on CAT(0) spaces must virtually split as direct products [11, Theorem II.7.1].

A collection of subgroups $\{K_1, \dots, K_n\}$ in a group H is called *almost malnormal* if $\#(gK_i g^{-1} \cap K_j) = \infty$ implies $i = j$ and $g \in K_i$. It is well known that a hyperbolic group is hyperbolic relative to any almost malnormal collection of quasiconvex subgroups [10].

Lemma 2.16 *Let K_e be the image of an edge group G_e into G_v and $\overline{K}_e$ be its projection in H_v under $G_v \to H_v = G_v/Z(G_v)$. Then $\mathbb{P} := \{\overline{K}_e \mid e_- = v, e \in \mathcal{G}^1\}$ is an almost malnormal collection of virtually cyclic subgroups in H_v.*

In particular, H_v is hyperbolic relative to $\mathbb{P}$.

Proof Since $Z(G_v) \subset K_e \cong \mathbb{Z}^2$, we have $\overline{K}_e = K_e/Z(G_v)$ is virtually cyclic. The almost malnormality follows from noncommensurability of K_e in G_v. Indeed, assume that $\overline{K}_e \cap h\overline{K}_{e'}h^{-1}$ contains an infinite order element by the hyperbolicity of H_v. If $g \in G_v$ is sent to h, then $K_e \cap gK_{e'}g^{-1}$ is sent to $\overline{K}_e \cap h\overline{K}_{e'}h^{-1}$. Thus, $K_e \cap gK_{e'}g^{-1}$ contains an abelian group of rank 2. The noncommensurability of K_e in G_v implies that $e = e'$ and $g \in K_e$. This shows that $\mathbb{P}$ is almost malnormal. $\square$

Let $G \curvearrowright X$ be a CKA action where G is the fundamental group of an admissible graph of groups $\mathcal{G}$, and let $G \curvearrowright T$ be the action of G on the associated Bass–Serre tree T of $\mathcal{G}$ (we refer the reader to [21, Section 2.5] for a brief discussion). Let T^0 and T^1 be the vertex and edge sets of T. By CAT(0) geometry:

(1) For every vertex $v \in T^0$, the minimal set $Y_v := \bigcap_{g \in Z(G_v)} \mathrm{Minset}(g)$ of X splits as metric product $\overline{Y}_v \times \mathbb{R}$ where $Z(G_v)$ acts by translation on the $\mathbb{R}$-factor and $H_v = G_v/Z(G_v)$ acts geometrically on the Hadamard space $\overline{Y}_v$. Since H_v is a hyperbolic group, it follows that $\overline{Y}_v$ is a hyperbolic space.

(2) For every edge $e \in T^1$, the minimal set $Y_e := \bigcap_{g \in G_e} \mathrm{Minset}(g)$ of X splits as $\overline{Y}_e \times \mathbb{R}^2 \subset Y_v$ where $\overline{Y}_e$ is a compact Hadamard space and $G_e = \mathbb{Z}^2$ acts cocompactly on the Euclidean plane $\mathbb{R}^2$.

We note that the assignments $v \to Y_v$ and $e \to Y_e$ are G-equivariant with respect to the natural G actions.

We summarize results in [21, Section 3.2] that will be used in this paper.

Lemma 2.17 *Let $G \curvearrowright X$ be a CKA action. Then there exists a constant $D > 0$ such that*

(1) $X = \bigcup_{v \in T^0} N_D(Y_v) = \bigcup_{e \in T^1} N_D(Y_e)$;

(2) *if $\sigma, \sigma' \in T^0 \cup T^1$ and $N_D(Y_\sigma) \cap N_D(Y_{\sigma'}) \neq \varnothing$ then $|\sigma - \sigma'|_T < D$.*

We shall refer to $\widetilde{Y}_v = N_D(Y_v)$ and $\widetilde{Y}_e = N_D(Y_e)$ as vertex and edge spaces for X.

2.3.1 Strips in CKA spaces [21, Section 4.2] We first choose, in a G-equivariant way, a plane $F_e \subset Y_e$ (which we will call *boundary plane*) for each edge $e \in T^1$. For every pair of adjacent edges e_1 and e_2, we choose, again equivariantly, a minimal geodesic from F_{e_1} to F_{e_2}; by the convexity of $Y_v = \overline{Y}_v \times \mathbb{R}$ where $v := e_1 \cap e_2$, this geodesic determines a Euclidean strip $\mathcal{S}_{e_1 e_2} := \gamma_{e_1 e_2} \times \mathbb{R}$ (possibly of width zero) for some geodesic segment $\gamma_{e_1 e_2} \subset \overline{Y}_v$.

Note that $\mathcal{S}_{e_1 e_2} \cap F_{e_i}$ is an axis of $Z(G_v)$. Hence if $e_1, e_2, e \in E$ and $e_i \cap e = v_i \in V$ are distinct vertices, then the angle between the geodesics $\mathcal{S}_{e_1 e} \cap F_e$ and $\mathcal{S}_{e_2 e} \cap F_e$ is bounded away from zero. If $\langle f_1 \rangle = Z(G_{v_1})$ and $\langle f_2 \rangle = Z(G_{v_2})$ then $\langle f_1, f_2 \rangle$ generates a finite-index subgroup of G_e. We remark that the intersection of two strips $\mathcal{S}_{e_1 e}$ and $\mathcal{S}_{e_2 e}$ is a point. Indeed, we have $\mathcal{S}_{e_1 e} \cap \mathcal{S}_{e_2 e} = (\mathcal{S}_{e_1 e} \cap F_e) \cap (\mathcal{S}_{e_2 e} \cap F_e)$. As two lines $\mathcal{S}_{e_1 e} \cap F_e$ and $\mathcal{S}_{e_2 e} \cap F_e$ in the plane F_e are axes of $\langle f_{v_1} \rangle = Z(G_{v_1})$ and $\langle f_{v_1} \rangle = Z(G_{v_2})$, respectively, and $\langle f_1, f_2 \rangle$ generates a finite-index subgroup of G_e, it follows that these two lines are not parallel, and hence their intersection must be a single point.

We note that the intersection of a boundary plane F_e of Y_v with the hyperbolic space $\overline{Y}_v$ is a line. The *boundary lines* $\mathbb{L}_v$ of the hyperbolic space $\overline{Y}_v$ is the collection of lines $\mathbb{L}_v = \{\ell_e := F_e \cap \overline{Y}_v \mid e_- = v\}$.

Definition 2.18 If for each edge $e := [v, w] \in T$, the boundary line $\ell = \overline{Y}_v \cap F_e$ is parallel to the $\mathbb{R}$-line in $Y_w = \overline{Y}_w \times \mathbb{R}$, then the CKA action is called *flip*.

In the sequel, it will be useful to make the following specific choices.

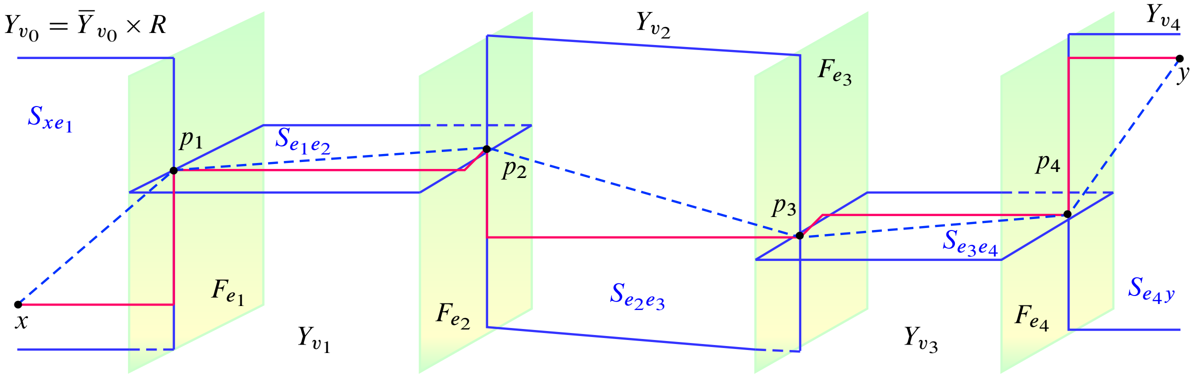


Figure 1: The dotted and blue path from x to y is a special path, and the red path is one L^1-version of it.

Definition 2.19 An *indexed map* $\rho\colon X \to T^0$ is a G-equivariant coarsely Lipschitz map such that $x \in \widetilde{Y}_{\rho(x)}$ for all $x \in X$.

If G acts freely on X, such a map ρ can be constructed as follows. Choose a fundamental set Σ such that Σ contains exactly one point from each orbit. Define $\rho\colon \Sigma \to T^0$ so that $x \in \widetilde{Y}_{\rho(x)}$, and extend ρ equivariantly to the whole space X. By Lemma 2.17.(2), one can show that ρ is a coarsely Lipschitz map: $|\rho(x) - \rho(y)|_T \le L|x-y|_X + L$ for some $L > 0$. See [21, Section 3.3] for more details.

If G acts only geometrically on X, we could replace X with a G-orbit Go for a basepoint o with trivial stabilizer. This does not matter much as we are only interested in the coarse geometry hereafter. By modifying X, we can always assume such a basepoint o exists. Indeed, this can be achieved by attaching a Euclidean cone to a point o such that its nontrivial but finite stabilizer acts freely on its boundary circle. Then we do the modification equivariantly for all translates in Go.

2.3.2 Special paths in CKA spaces Let $G \curvearrowright X$ be a CKA action. We now introduce the class of *special paths* in X.

Definition 2.20 (special paths in X) Let $\rho\colon X \to T^0$ be the indexed map given by Definition 2.19. Let x and y be two points in X. If $\rho(x) = \rho(y)$, a *special path* in X connecting x to y is the geodesic $[x, y]$. Otherwise, let $e_1 \cdots e_n$ be the geodesic edge path connecting $\rho(x)$ to $\rho(y)$ and let $p_i = \mathcal{S}_{e_{i-1}e_i} \cap \mathcal{S}_{e_ie_{i+1}}$ be the intersection point of adjacent strips, where $e_0 := x$ and $e_{n+1} := y$. A *special path* connecting x to y is the concatenation of the geodesics

$$[x, p_1][p_1, p_2] \cdots [p_{n-1}, p_n][p_n, y].$$

Remark 2.21 By definition, except for $[x, p_1]$ and $[p_n, y]$, the special path depends only on the geodesic $e_1 \cdots e_n$ in T, the choice of planes F_e and the indexed map ρ.

Proposition 2.22 [43, Proposition 3.8] *There exists a constant $\mu > 0$ such that every special path γ in X is a (μ, μ)-quasigeodesic.*

Assume that $v_0 = \rho(x), v_{2n} = \rho(y) \in \mathcal{V}$ are such that $d(v_0, v_{2n}) = 2n$ for $n \geq 0$. If γ is a special path between x and y, then we define

$$|x-y|_X^{\mathrm{hor}} := \sum_{i=0}^{2n} |p_i - p_{i+1}|_{Y_{v_i}}^{\mathrm{hor}}, \quad |x-y|_X^{\mathrm{ver}} := \sum_{i=0}^{2n} |p_i - p_{i+1}|_{Y_v}^{\mathrm{ver}} \tag{2}$$

where $p_0 := x$ and $p_{n+1} := y$. By Proposition 2.22, we have

$$|x-y|_X \sim |x-y|_X^{\mathrm{hor}} + |x-y|_X^{\mathrm{ver}}.$$

By definition, the system of special paths is G-invariant, so the symmetric functions $d^h(x,y)$ and $d^v(x,y)$ are G-invariant for any $x, y \in X$.

We partition the vertex set T^0 of the Bass–Serre tree into two disjoint classes of vertices $\mathcal{V}_1$ and $\mathcal{V}_2$ such that if v and v' are in $\mathcal{V}_i$ then $d_T(v, v')$ is even.

Lemma 2.23 [43, Lemma 4.6] *There exists a subgroup $\dot{G}$ of index at most 2 in G preserving $\mathcal{V}_i$ for $i = 1, 2$ such that $G_v \subset \dot{G}$ for any $v \in T^0$.*

2.4 Projection axioms

In this subsection, we briefly recall the work of Bestvina, Bromberg and Fujiwara [5] on constructing a quasitree of spaces.

Definition 2.24 (projection axioms) Let $\mathbb{Y}$ be a collection of geodesic spaces equipped with projection maps

$$\{\pi_Y : \mathbb{Y} - \{Y\} \to \mathcal{P}(Y)\}_{Y \in \mathbb{Y}}$$

where $\mathcal{P}(Y)$ is the power set of Y. Write $d_Y(X, Z) = \mathrm{diam}(\pi_Y(X) \cup \pi_Y(Z))$ for $X \neq Y \neq Z \in \mathbb{Y}$. The pair $(\mathbb{Y}, \{\pi_Y\}_{Y \in \mathbb{Y}})$ satisfies *projection axioms* for a *projection constant* $\xi \geq 0$ if the following hold:

(1) $\mathrm{diam}(\pi_Y(X)) \leq \xi$ when $X \neq Y$.

(2) If X, Y and Z are distinct and $d_Y(X, Z) > \xi$ then $d_X(Y, Z) \leq \xi$.

(3) For $X \neq Z$, the set $\{Y \in \mathbb{Y} \mid d_Y(X, Z) > \xi\}$ is finite.

The following is a useful example to keep in mind throughout the paper. For further details, we refer the reader to the introduction of [5]. In this example, the collection of metric spaces $\mathbb{Y}$ consists of subspaces of a singe metric space; however, we emphasize that this need not be the case in general.

Example 2.25 Let G be a discrete group of isometries of $\mathbb{H}^2$, and $\gamma \in G$ be a loxodromic element with axis γ. Let $\mathbb{Y}$ be the set of all G-translates of γ. Given $Y \in \mathbb{Y}$, let π_Y denote the shortest projection map in $\mathbb{H}^2$. Since all translates of γ are convex, this is a well-defined 1-Lipschitz map. One may check that $(\mathbb{Y}, \{\pi_Y\}_{Y \in \mathbb{Y}})$ satisfies the projection axioms for some constant ξ.

Remark 2.26 Let $(\mathbb{Y}, \{\pi_Y\}_{Y \in \mathbb{Y}})$ satisfy the projection axioms. By [7, Theorem 4.1 and Lemma 4.13], there exists a variant π'_Y of π_Y such that π_Y and π'_Y are uniformly close in Hausdorff distance, and $(\mathbb{Y}, \{\pi'_Y\}_{Y \in \mathbb{Y}})$ satisfies strong projection axioms, ie the axioms are the same as projection axioms except for replacing (2) in Definition 2.24 with the following stronger statement: if X, Y, Z are distinct and $d_Y(X, Z) > \xi$ then $\pi_X(Y) = \pi_X(Z)$ for a projection constant ξ' depending only on ξ.

The following results from [5] will be used in this paper.

- Fix $K > 0$. In [5], a quasitree of spaces $\mathcal{C}_K(\mathbb{Y})$ is constructed for given $(\mathbb{Y}, \{\pi_Y\}_{Y \in \mathbb{Y}})$ which satisfies the projection axioms with constant ξ.
- If $K > 4\xi$ and $\mathbb{Y}$ is a collection of uniform quasilines, then $\mathcal{C}_K(\mathbb{Y})$ is a unbounded quasitree. If $\mathbb{Y}$ admits a group action of G such that $\pi_{gY} = g\pi_Y$ for any $g \in G$ and $Y \in \mathbb{Y}$, then G acts by isometry on $\mathcal{C}_K(\mathbb{Y})$.

Set $[t]_K = t$ if $t \geq K$, otherwise $[t]_K = 0$. Let $x \in X$ and $z \in Z$ for $X, Z \in \mathbb{Y}$. If $X \neq Y \neq Z$, we define $d_Y(x, z) = d_Y(X, Z)$. If $Y = X$ and $Y \neq Z$, then define $d_Y(x, z) = \operatorname{diam}(\pi_Y(x, Z))$. If $X = Y = Z$, let $d_Y(x, z)$ be the distance in Y. The following distance formula from [7] is crucial in what follows.

Proposition 2.27 [7, Theorem 6.3] *Let $(\mathbb{Y}, \{\pi_Y\}_{Y \in \mathbb{Y}})$ satisfy the strong projection axioms with constant ξ. Then for any $x, y \in \mathcal{C}_K(\mathbb{Y})$,*

$$\frac{1}{4} \sum_{Y \in \mathbb{Y}} [d_Y(x, y)]_K \leq |x - y|_{\mathcal{C}_K(\mathbb{Y})} \leq 2 \sum_{Y \in \mathbb{Y}} [d_Y(x, y)]_K + 3K$$

for all $K \geq 4\xi$.

Definition 2.28 (acylindrical action [9; 44]) Let G be a group acting by isometries on a metric space (X, d). The action of G on X is called *acylindrical* if for any $r \geq 0$, there exist constants $R, N \geq 0$ such that for any pair $a, b \in X$ with $|a - b|_X \geq R$ we have

$$\#\{g \in G \mid |ga - a|_X \leq r \text{ and } |gb - b|_X \leq r\} \leq N.$$

By [9], any nontrivial isometry of acylindrical group action on a hyperbolic space is either elliptic or loxodromic. A (λ, c)-quasigeodesic γ for some $\lambda, c > 0$ is referred to as a *quasiaxis* for a loxodromic element g if γ and $g\gamma$ have finite Hausdorff distance depending only on λ and c.

A group is called *nonelementary* if it is neither finite nor virtually cyclic.

Proposition 2.29 [6] *Assume that a nonelementary hyperbolic group H acts acylindrically on a hyperbolic space $\overline{Y}$. For a loxodromic element $g \in H$, consider the set $\mathbb{A}$ of all H-translates of a given (λ, c)-quasiaxis of g for given $\lambda, c > 0$. Then there exists a constant $\theta = \theta(\lambda, c) > 0$ such that for any $\gamma \in \mathbb{A}$, the set*

$$\{h \in G \mid \operatorname{diam}(\pi_\gamma(h\gamma)) \geq \theta\}$$

is a finite union of double $E(g)$-cosets.

In particular, there are only finitely many distinct pairs $(\gamma, \gamma') \in \mathbb{A} \times \mathbb{A}$ satisfying $\operatorname{diam}(\pi_\gamma(\gamma')) > \theta$ up to the action of H.

Lemma 2.30 [54, Lemma 2.14] *Let H be a nonelementary group admitting a cobounded and acylindrical action on a δ-hyperbolic space $(\overline{Y}, d)$. Fix a basepoint o. Then there exists a set $F \subset H$ of three loxodromic elements and $\lambda, c > 0$ with the following property.*

For any $h \in H$ there exists $f \in F$ such that hf is a loxodromic element and the bi-infinite path

$$\gamma = \bigcup_{i \in \mathbb{Z}} (hf)^i([o, ho][ho, hfo])$$

is a (λ, c)-quasigeodesic.

Convention 2.31 When speaking of quasilines in hyperbolic spaces with actions satisfying Lemma 2.30, we always mean (λ, c)-quasigeodesics where $\lambda, c > 0$ depend on F and δ.

3 Property (QT) of relatively hyperbolic groups

In this section, we are going to prove Theorem 1.5. The notion of relatively hyperbolic groups can be formulated from a number of equivalent ways. Here we shall present a quick definition due to Bowditch [10] and recall the relevant facts we shall need without proofs.

Let H be a finitely generated group with a finite collection of subgroups $\mathcal{P}$. Fixing a finite generating set S, we consider the corresponding Cayley graph $\operatorname{Cay}(H, S)$ equipped with path metric d and we denote by $|h|_H = d(1, h)$ the word length.

Denote by $\mathbb{P} = \{hP \mid h \in H, P \in \mathcal{P}\}$ the collection of peripheral cosets. Let $\widehat{H}(\mathbb{P})$ be the *coned-off Cayley graph* obtained from $\operatorname{Cay}(H, S)$ as follows. A *cone point* denoted by $c(P)$ is added for each peripheral coset $P \in \mathbb{P}$ and is joined by *half edges* to each element in P. The union of two half edges at a cone point is called a *peripheral edge*. Denote by $\hat{d}$ the induced path metric after coning-off and $|h|_{\widehat{H}} = \hat{d}(1, h)$.

The pair $(G, \mathcal{P})$ is said to be *relatively hyperbolic* if the coned-off Cayley graph $\widehat{H}(\mathbb{P})$ is hyperbolic and *fine*: any edge is contained in finitely many simple circles with uniformly bounded length.

By [9, Lemma 3.3; 44, Proposition 5.2], the action of H on $\widehat{H}(\mathbb{P})$ is acylindrical.

Let π_P denote the shortest projection in word metric to $P \in \mathbb{P}$ in H and $d_P(x,y)$ the $|\cdot|_H$-diameter of the projections of the points x, y to P. Since $\mathbb{P}$ has the strongly contracting property with bounded intersection property, the projection axioms with a constant $\xi > 0$ hold for $\mathbb{P}$ (see [48]).

3.1 Thick distance formula

A geodesic edge path β in the coned-off Cayley graph $\widehat{H}(\mathbb{P})$ is *K-bounded* for $K > 0$ if the end points of every peripheral edge have d-distance at most K.

By definition, a geodesic $\beta = [x, y]$ can be subdivided into maximal K-bounded nontrivial segments α_i $(0 \le i \le n)$ separated by peripheral edges e_j $(0 \le j \le m)$ where $|(e_j)_- - (e_j)_+|_H > K$. It is possible that $n = 0$: β consists of only peripheral edges.

Define

$$|\beta|_K := \sum_{0 \le i \le n} [\mathrm{Len}(\alpha_i)]_K,$$

which sums up the lengths of K-bounded subpaths of length at least K. It is possible that $n = 0$, so $|\beta|_K = 0$. Define the *K-thick distance*

$$|x-y|_{\widehat{H}}^K = \max\{|\beta|_K\} \tag{3}$$

over all relative geodesics β between x and y. Thus, $|x-y|_{\widehat{H}}^K$ is H_v-invariant.

A relative path without backtracking in $\widehat{H}(\mathbb{P})$ admits nonunique *lifts* in $\mathrm{Cay}(H, S)$ which are obtained by replacing the peripheral edge by a geodesic in $\mathrm{Cay}(H, S)$ with the same endpoints. The distance formula follows from the fact that the lift of a relative quasigeodesic is a quasigeodesic (see [25; 27, Proposition 6.1]). The following formula is made explicit in [48, Theorem 0.1].

Lemma 3.1 *For any sufficiently large $K > 0$ and for any $x, y \in H$,*

$$|x-y|_H \sim_K |x-y|_{\widehat{H}}^K + \sum_{P \in \mathbb{P}} [d_P(x,y)]_K. \tag{4}$$

The following result is proved in [43, Lemma 5.5] under the assumption that H is hyperbolic relative to a set of virtually cyclic subgroups. However, the same proof works for any relatively hyperbolic group.

Lemma 3.2 *For any sufficiently large $K > 0$, there exists an H-finite collection $\mathbb{A}$ of quasilines in $\widehat{H}$ and a constant $N = N(K, \widehat{H}, \mathbb{A}) > 0$ such that for any two vertices $x, y \in \widehat{H}$,*

$$|x-y|_{\widehat{H}}^K \sim_N \sum_{\ell \in \mathbb{A}} [\hat{d}_\ell(x,y)]_K. \tag{5}$$

A group H endowed with the *profinite topology* is a topological group such that the set of all finite-index subgroups is a (closed/open) neighborhood base of the identity. A subgroup P is called *separable* if it is closed in the profinite topology. Equivalently, it is the intersection of all finite-index subgroups containing P. A group is called *residually finite* if the trivial subgroup is closed.

A maximal abelian subgroup of a residually finite group is separable (see [31, Proposition 1]). Note that a maximal elementary (ie virtually cyclic) group E in a relatively hyperbolic group H contains a maximal abelian group (of rank 1) as a finite-index subgroup. If H is residually finite, then E is a finite union of closed subsets; hence it is closed and thus separable.

We will use the following corollary in the proof of Theorem 1.5.

Corollary 3.3 *Assume that H is a residually finite relatively hyperbolic group. Then for any $K \gg 0$, there exists a finite-index subgroup $\dot{H} \le H$ acting on finitely many quasitrees T_i $(1 \le i \le n)$ such that every orbital map of the $\dot{H}$-action on $\prod_{i=1}^n T_i$ is a quasi-isometric embedding from $(\dot{H}, |\cdot|_{\hat{H}}^K)$ to $\prod_{i=1}^n T_i$.*

This corollary is essentially proved in [43], inspired by the arguments in the setting of mapping class groups [6]. We sketch the proof for the convenience of the reader.

Sketch of proof Recall that for any $\theta > 0$, a set $\mathbb{T}$ of (uniform) quasilines in a hyperbolic space with θ-bounded projection satisfies the projection axioms for a projection constant $\xi = \xi(\theta) > 0$. Let λ and c be the constants given by Lemma 2.30 with respect to the acylindrical action $H \curvearrowright \hat{H}$. For our purpose, we will choose θ to be the constant given by Proposition 2.29. Then the distance formula for the quasitree $\mathscr{C}_K(\mathbb{T})$ constructed from $\mathbb{T}$ holds for any $K \ge 4\xi$.

For a fixed large constant K, Lemma 3.2 provides an H-finite set of quasilines $\mathbb{A}$ such that (5) holds. We then use the separability to find a finite-index subgroup $\dot{H}$ of H such that $\mathbb{A}$ decomposes as a finite union of $\dot{H}$-invariant $\mathbb{T}_i$ each of which satisfies the projection axioms with projection constant ξ. To be precise, the stabilizer E of a quasiline ℓ in $\mathbb{A}$ is a maximal elementary subgroup of H and thus is separable in H if H is residually finite (since a maximal abelian group in a residually finite group is separable). By Proposition 2.29 and the paragraph after Lemma 2.1 in [6], the separability of E allows one to choose a finite-index subgroup $\dot{H}$ containing E such that any $\dot{H}$-orbit $\mathbb{T}_i$ in the collection of quasilines $H\ell$ satisfies the projection axioms with projection constant ξ. We take a common finite-index subgroup $\dot{H}$ for finitely many quasilines ℓ in $\mathbb{A}$ up to H-orbits and therefore have found all $\dot{H}$-orbit $\mathbb{T}_i$ such that their union covers $\mathbb{A}$.

Finally, it is straightforward to verify that the right-hand term of (5) coincides with the sum of distances over the finitely many quasitrees $T_i := \mathscr{C}_K(\mathbb{T}_i)$. Thus, the thick distance $d_{\hat{H}}^K(x, y)$ is quasi-isometric to the distance on a finite product of quasitrees. □

All our discussion generalizes to the geometric action of H on a geodesic metric space Y, since there exists an H-equivariant quasi-isometry between $\mathrm{Cay}(H, S)$ and Y. Therefore, replacing $\mathrm{Cay}(H, S)$ with Y, we have the same thick distance formula. This is the setup for CKA actions in next sections.

In next subsection, we obtain property (QT) for relatively hyperbolic groups provided peripheral subgroups do so.

3.2 Proof of property (QT) of relatively hyperbolic groups

Proof of Theorem 1.5 Recall that $\mathcal{P}$ is a finite set of subgroups. For each $P \in \mathcal{P}$, choose a full set E_P of left P-coset representatives in H such that $1 \in E_P$. For given P and $1 \le i \le n_P$, we define the collection of quasitrees

$$\mathbb{T}_P^i := \{ f T_i \mid f \in E_P \}$$

where T_i are quasitrees associated to P given by assumption. Then H preserves $\mathbb{T}_P^i$ by the following action: for any point $f(x) \in f T_i$ and $h \in H$,

$$h \cdot f(x) := f' p(x) \in f' T_i$$

where $p \in P$ is given by $hf = f'p$ for $f' \in E_P$.

We are now going to define projection maps $\{\pi_{fT_i}\}$ as follows.

By assumption, we fix an orbital embedding ι_P^i of P into T_i such that the induced map

$$\prod_{i=1}^{n_P} \iota_P^i : P \to \prod_{i=1}^{n_P} T_i$$

is a quasi-isometric embedding. We then define an equivariant family of orbital maps $\iota_{fP}^i : fP \to fT_i$ such that

$$\iota_{fP}^i(x) := f \iota_P^i (f^{-1} x) \quad \text{for all } x \in fP.$$

Then for any $h \in H$ and $x \in fP$, $h \cdot \iota_{fP}^i(x) = \iota_{f'P}^i(hx)$ where $f' \in E_P$ with $hf = f'p$ and $p \in P$.

Let π_{fP} be the shortest projection to the coset fP in H with respect to the word metric. For any two distinct $fT_i, f'T_i \in \mathbb{T}_P^i$, we set

$$\pi_{fT_i}(f'T_i) := \iota_{fP}^i(\pi_{fP}(f'P)).$$

Recall that $\mathbb{P} = \{ fP \mid f \in H, P \in \mathcal{P} \}$ satisfies the projection axioms with shortest projection maps $\{\pi_{fP}\}$. It is readily checked that the projection axioms pass to the collection $\mathbb{T}_P^i$ under equivariant Lipschitz maps $\{\iota_{fP}^i\}_{fP \in \mathbb{P}}$.

We can therefore build the projection complex for $\mathbb{T}_P^i$ for a fixed $K \gg 0$. By Proposition 2.27, the following distance holds for any $x', y' \in \mathcal{C}_K(\mathbb{T}_P^i)$:

$$|x' - y'|_{\mathcal{C}_K(\mathbb{T}_P^i)} \sim_K \sum_{T \in \mathbb{T}_P^i} [d_T(x', y')]_K. \tag{6}$$

Note that $\prod_{i=1}^{n_P} \iota_P^i : P \to \prod_{i=1}^{n_P} T_i$ is a quasi-isometric embedding for each $P \in \mathcal{P}$. Thus, for any $x, y \in G$ and $P \in \mathbb{P}$,

$$d_P(x, y) = |\pi_P(x) - \pi_P(y)|_P \sim \sum_{i=1}^{n_P} |\iota_P^i(\pi_P(x)) - \iota_P^i(\pi_P(y))|_{T_i}. \tag{7}$$

Setting $x' = \iota_P^i(\pi_P(x))$ and $y' = \iota_P^i(\pi_P(y))$ in (7), we deduce from (6) that

$$d_P(x,y) \preceq_K \sum_{i=1}^{n_P} |\iota_P^i(\pi_P(x)) - \iota_P^i(\pi_P(y))|_{\mathscr{C}_K(\mathbb{T}_P^i)}. \tag{8}$$

Recall from Lemma 3.1 that for any $x, y \in H$, we have

$$|x-y|_H \sim_K |x-y|_{\hat{H}}^K + \sum_{P\in\mathbb{P}} [d_P(x,y)]_K.$$

Note that the orbital map of any isometric action is Lipschitz. To prove property (QT) of H, it suffices to give an upper bound of $|x-y|_H$. Taking account of (8), it remains to construct a finite product of quasitrees to bound $|x-y|_{\hat{H}}^K$.

Since H is residually finite, by Corollary 3.3, there exists a finite-index subgroup, still denoted by H, and a finite product Y of quasitrees such that the orbital map Π_0 from H to Y gives a quasi-isometric embedding of H equipped with $|\cdot|_{\hat{H}}^K$-function into Y.

Recall that π_P is the shortest projection to $P \in \mathbb{P}$. For $1 \le i \le n_P$, define

$$\Pi_i : H \to \mathscr{C}_K(\mathbb{T}_P^i)$$

by sending an element $h \in H$ to $\iota_P^i(\pi_P(h))$. We then have n equivariant maps Π_i from H to quasitrees after reindexing, where $n := \sum_{P\in\mathscr{P}} n_P$.

Let $\Pi := \Pi_0 \times \prod_{i=1}^n \Pi_i$ be the map from H to $Y \times \prod_{i=1}^n \mathscr{C}_K(\mathbb{T}_P^i)$, where Y is the finite product of quasitrees as in the previous paragraphs. As previously mentioned, the product map Π gives an upper bound on $d_H(x,y)$, so is a quasi-isometric embedding of H. Therefore, H has property (QT). □

Remark 3.4 An immediate corollary of Theorem 1.5 is that the fundamental group of a finite volume hyperbolic 3-manifold has property (QT). An alternative proof is that $\pi_1(M)$ is virtually compact special by deep theorems of Agol [3] and Wise [53], and thus $\pi_1(M)$ has property (QT).

We say that the profinite topology on H induces a *full profinite topology* on a subgroup P if every finite-index subgroup of P contains the intersection of P with a finite-index subgroup of H.

Theorem 3.5 *Suppose that H is residually finite and each $P \in \mathscr{P}$ is separable. Assume furthermore that H induces the full profinite topology on each $P \in \mathscr{P}$. If each $P \in \mathscr{P}$ acts by isometry on a finite product of quasitrees without $\mathbb{R}$-factor such that orbital maps are quasi-isometric embeddings, then H has property (QT).*

Proof By [26, Corollary 1.3], there is a finite-index subgroup $\dot{P}$ of P acting on each quasitree T_i such that the diagonal action of $\dot{P}$ on $\prod_{i=1}^n T_i$ induces a quasi-isometric embedding orbital map $\prod_{i=1}^n \iota_{\dot{P}}^i$.

By the assumption, H induces the full profinite topology on $P \in \mathscr{P}$, so every finite-index subgroup of a separable subgroup P is also separable. Thus, there are finite-index subgroups $\dot{H}_P$ of H for $P \in \mathscr{P}$ such that $\dot{P} = \dot{H}_P \cap P$.

Consider the finite-index normal subgroup $\dot{H} := \bigcap\{h\dot{H}_P h^{-1} \mid P \in \mathcal{P}\}$ in H. Since $\dot{H}$ is normal in H, we see that $\dot{H} \cap hPh^{-1} \subset h\dot{P}h^{-1}$ is equivalent to $\dot{H} \cap P \subset \dot{P}$. The later holds by the choice of $\dot{H}_P$. Hence, for every $h \in H$, $\dot{H} \cap hPh^{-1}$ preserves the factors of the product decomposition. Note that $\dot{H}$ is hyperbolic relative to $\{\dot{H} \cap hPh^{-1} \mid h \in H\}$. The conclusion follows from Theorem 1.5. □

In subsequent sections (Sections 4, 5, 6 and 7), the proof of property (QT) of CKA groups will be discussed, which may be considered as the technical heart of this paper.

4 Coning-off CKA spaces

In this section, we recapitulate the content of [43, Section 5] and give an outline of the proof of Theorem 1.3.

Let $G \curvearrowright X$ be a CKA action where G is the fundamental group of an admissible graph of groups $\mathcal{G}$ (see Section 2.3), and let $G \curvearrowright T$ be the action of G on the associated Bass–Serre tree T of $\mathcal{G}$. Let T^0 and T^1 be the vertex and edge sets of T.

Let $\{F_e\}$ be the collection of boundary planes of the space Y_v (see Section 2.3). We note that the intersection of a boundary plane F_e of Y_v with the hyperbolic space $\overline{Y}_v$ is a line. We define the collection of lines $\mathbb{L}_v$ of the hyperbolic space $\overline{Y}_v$ as

$$\mathbb{L}_v = \{\ell_e := F_e \cap \overline{Y}_v \mid e_- = v\},$$

which shall be referred as *boundary lines*.

4.1 Construction of coned-off spaces

Recall that $T^0 = \mathcal{V}_1 \cup \mathcal{V}_2$ where $\mathcal{V}_i$ consists of vertices in T with pairwise even distances. Let $\dot{G} < G$ be the subgroup of index at most 2 preserving $\mathcal{V}_1$ and $\mathcal{V}_2$ given by Lemma 2.23.

Fix a large $r > 0$. A *hyperbolic r-cone* by definition is the metric completion of the (incomplete) universal cover of a punctured hyperbolic disk of radius r. Let $\mathbb{Y}_i = \{\overline{Y}_v \mid v \in \mathcal{V}_i\}$ be the collection of hyperbolic spaces and $\dot{\mathbb{Y}}_i = \{\dot{Y}_v \mid v \in \mathcal{V}_i\}$ be their coned-off spaces (which are uniformly hyperbolic for $r \gg 0$) by attaching hyperbolic r-cones along the boundary lines of $\overline{Y}_v$.

Note that $\dot{G}$ preserves $\mathbb{Y}_i$ and $\dot{\mathbb{Y}}_i$ by the action on the index $gY_v = Y_{gv}$ for any $g \in \dot{G}$. For each $w \in T^0$, let $\mathrm{St}(w)$ be the star of w in T with adjacent vertices as *extremities*. Then $\mathrm{St}(w)$ admits the action of G_w so that the stabilizers of the extremities are the corresponding edge groups.

Define $\dot{\mathcal{X}}_i$ to be the space obtained from the disjoint union of coned-off spaces $\dot{Y}_v$ $(v \in \mathcal{V}_i)$ with cone points identified with the extremities of the stars $\mathrm{St}(w)$ with $v \in \mathrm{Lk}(w)$. Endowed with induced length metric, the space $\dot{\mathcal{X}}_i$ is a Gromov-hyperbolic space.

Lemma 4.1 *Fix a sufficiently large $r > 0$ and $i \in \{1, 2\}$. The space $\dot{\mathcal{X}}_i$ is a δ-hyperbolic space where $\delta > 0$ only depends on the hyperbolicity constants of $\dot{Y}_v$ $(v \in \mathcal{V}_i)$.*

The subgroup $\dot{G}$ acts on $\dot{\mathscr{X}}_i$ with the following properties:

(1) *for each $v \in \mathscr{V}_i$, the stabilizer of $\dot{Y}_v$ is isomorphic to G_v and H_v acts coboundedly on $\dot{Y}_v$, and*

(2) *for each $w \in T^0 - \mathscr{V}_i$, G_w acts on $\mathrm{St}(w)$ in the same manner as the action on the Bass–Serre tree T.*

Proof Note that the stabilizers of the cone points of $\dot{Y}_v$ under the action of G_v on $\dot{Y}_v$ are the same as that of the extremities of stars $\mathrm{St}(w)$, which are both the edge groups G_e for $e = [v, w]$. By construction, the cone points of $\dot{Y}_v$ are identified with the extremities of stars $\mathrm{St}(w)$, so the actions of G_v on $\dot{Y}_v$ ($v \in \mathscr{V}_i$) and of G_w on $\mathrm{St}(w)$ ($w \in T^0 - \mathscr{V}_i$) extend over $\dot{\mathscr{X}}_i$, and hence $\dot{G}$ acts by isometries on $\dot{\mathscr{X}}_i$. □

Remark 4.2 Our construction of coned-off spaces is slightly different from the one in [43, Section 5.1], where the cone points are identified directly between different spaces $\dot{Y}_v$ and $\dot{Y}_{v'}$. Thus certain assumption on vertex groups is necessary in [43] to ensure an action on the coned-off space.

We now define the thick distance on $\dot{\mathscr{X}}_i$ ($i = 1, 2$) by taking the sum of thick distances through $\dot{Y}_v$ as follows.

If x is a point in a coned-off space $\dot{Y}_v \subset \dot{\mathscr{X}}_i$, we denote $\rho(x)$ by v (by abuse of notation). By the above tree-like construction, any path between $x, y \in \dot{\mathscr{X}}_i$ has to pass through in order a pair of boundary lines ℓ_v^- and ℓ_v^+ of $\overline{Y}_v$ for each $v \in [\rho(x), \rho(y)]$. By abuse of language, if x is not contained in a hyperbolic cone, set $\ell_v^- = x$ for $v = \rho(x)$. Similarly, if y is not contained in a hyperbolic cone, set $\ell_v^+ = y$ for $v = \rho(y)$. Let (x_v, y_v) be a pair of points in the boundary lines (ℓ_v^-, ℓ_v^+) such that $[x_v, y_v]$ is orthogonal to ℓ_v^- and ℓ_v^+. Recall that $|x_v - y_v|_{\dot{Y}_v}^K$ is the K-cut-off thick distance defined in (3).

Definition 4.3 For any $K \geq 0$, the *K-thick distance* between x and y is defined by

$$|x - y|_{\dot{\mathscr{X}}_i}^K := \sum_{v \in [\rho(x), \rho(y)] \cap \mathscr{V}_i} |x_v - y_v|_{\dot{Y}_v}^K. \tag{9}$$

Since $|\cdot|_{\dot{Y}_v}^K$ is H_v-invariant, we see that $|x - y|_{\dot{\mathscr{X}}_i}^K$ is $\dot{G}$-invariant.

Remark 4.4 The definition of $|\cdot|_{\dot{\mathscr{X}}_i}^K$ is designed to ignore the parts in hyperbolic cones between different pieces. One consequence is that perturbing x and y in hyperbolic cones does not change their K-thick distance.

4.2 Construct the collection of quasilines in $\dot{\mathscr{X}}_i$

If $E(\ell)$ denotes the stabilizer in H_v of a boundary line ℓ of $\overline{Y}_v$, then $E(\ell)$ is virtually cyclic and almost malnormal. Since $\{E(\ell)\}$ is H_v-finite by conjugacy, let $\mathbb{E}_v$ be a complete finite set of conjugacy representatives. By Lemma 2.16, H_v is hyperbolic relative to peripheral subgroups $\mathbb{E}_v$. Hence, the results in Section 3 apply here.

Let $\lambda, c > 0$ be the universal constants given by Lemma 2.30 applied to the actions of H_v on $\dot{Y}_v$ for all $v \in T^0$ (since there are only finitely many actions up to conjugacy). By convention, the quasilines in coned-off spaces are understood as (λ, c)-quasigeodesics in $\dot{\mathscr{X}}_i$ and $\dot{Y}_v$.

The coning-off construction has the following consequence [43, Lemma 5.14]: the shortest projection of any quasiline α in $\dot{Y}_v$ to a quasiline β in $\dot{Y}_{v'}$ has to pass through the cone point attached to $\dot{Y}_{v'}$, and thus has uniformly bounded diameter by $\theta = \theta(\lambda, c) > 0$.

For simplicity, we also assume that $\theta = \theta(\lambda, c) > 0$ satisfies the conclusion of Proposition 2.29. Consequently, this determines a constant $\xi = \xi(\theta) > 0$ such that any set of quasilines with θ-bounded projection satisfies the projection axioms with projection constant ξ.

Fix $K > \max\{4\xi, \theta\}$. For each $v \in \mathcal{V}$, there exists an H_v-finite collection of quasilines $\mathbb{A}_v$ in $\dot{Y}_v$ and a constant $N = N(\mathbb{A}_v, K)$ such that the $d^K_{\hat{H}_v}$-distance formula holds by Lemma 3.2.

Since $\dot{G}$ acts cofinitely on $\mathcal{V}_1$ and $\mathcal{V}_2$, we can assume $\mathbb{A}_w = g\mathbb{A}_v$ if $w = gv$ for $g \in \dot{G}$. Let

$$\mathbb{A}_i := \bigcup_{v \in \mathcal{V}_i} \mathbb{A}_v$$

for $i = 1, 2$, which are both $\dot{G}$-invariant. We now equip $\mathbb{A}_i$ with projection maps as the shortest projection maps between two quasilines in $\dot{\mathcal{X}}_i$ for $i = 1, 2$.

If γ is a quasiline in $\dot{\mathcal{X}}_i$ for $i = 1, 2$, denote by $\dot{d}_\gamma(x, y)$ the $|\cdot|_{\dot{\mathcal{X}}_i}$-diameter of the shortest projection of $x, y \in \dot{\mathcal{X}}_i$ to γ.

The following result shows that the thick distance is captured by the projections of $\mathbb{A}_i$. Recall that r is the radius of the hyperbolic cones in constructing $\dot{\mathcal{X}}_i$.

Proposition 4.5 [43, Proposition 5.9] *For any $x, y \in \dot{\mathcal{X}}_i$,*

$$|x - y|^K_{\dot{\mathcal{X}}_i} \sim_{r,K} \sum_{\gamma \in \mathbb{A}_i} [\dot{d}_\gamma(x, y)]_K + |\rho(x) - \rho(y)|_T. \tag{10}$$

In the next subsection, we construct a suitable finite subgroup of G such that it acts isometrically on a finite product of quasitrees $T_1, \dots, T_n$ under some assumptions on vertex groups. This allows rewriting the right-hand side of the distance formula (10) as the product distance of the T_i.

4.3 Isometric action of a suitable finite-index subgroup of G

In a group, two elements are *independent* if they do not have conjugate powers (see [52, Definition 3.2]).

Definition 4.6 A group H is *omnipotent* if for any nonempty set of pairwise independent elements $\{h_1, \dots, h_r\}$ $(r \geq 1)$ there is a integer $p \geq 1$ such that for every choice of positive natural numbers $\{n_1, \dots, n_r\}$, there is a finite quotient $H \to \hat{H}$ such that $\hat{h}_i$ has order $n_i p$ for each i.

Let $G \curvearrowright X$ be a CKA action, where G is the fundamental group of the admissible graph of groups $\mathcal{G}$ such that every vertex group G_v is a central extension of an omnipotent hyperbolic group. By Lemma 4.1, the finite-index subgroup $\dot{G}$ acts on $\dot{\mathcal{X}}_1 \times \dot{\mathcal{X}}_2 \times T$ which is equipped with the $\dot{G}$-invariant function $|\cdot|^K_{\dot{\mathcal{X}}_1} \times |\cdot|^K_{\dot{\mathcal{X}}_2} \times |\cdot|_T$. The main result of this subsection is the following.

Proposition 4.7 *The group $\dot{G}$ admits finitely many isometric actions on quasitrees T_i for $1 \le i \le n$ such that there exists a $\dot{G}$-equivariant quasi-isometric embedding from $\dot{\mathscr{X}}_1 \times \dot{\mathscr{X}}_2 \times T$ to $T_1 \times T_2 \times \cdots \times T_n \times T$.*

We emphasize here that $|\cdot|^K_{\dot{\mathscr{X}}_1} \times |\cdot|^K_{\dot{\mathscr{X}}_2} \times |\cdot|_T$ on the domain for the quasi-isometric embedding is not a distance function, but the target is equipped with product distance.

By [11, Theorem II.6.12], G_v contains a subgroup K_v intersecting trivially with $Z(G_v)$ such that the direct product $K_v \times Z(G_v)$ is a finite-index subgroup. Thus, the image of K_v in $G_v/Z(G_v)$ is of finite index in H_v and K_v acts geometrically on hyperbolic spaces $\overline{Y}_v$. Since H_v is omnipotent and then is residually finite, we can assume that K_v is torsion-free.

Recall the $\dot{G}$-invariant collection of quasilines in Section 4.2,

$$\mathbb{A}_i = \bigcup_{v \in \mathscr{V}_i} \mathbb{A}_v,$$

where $\mathbb{A}_v$ is the collection of quasilines such that d^K-distance formula holds by Lemma 3.2. By the residual finiteness of K_v, there exists a finite-index subgroup $\dot{K}_v$ such that $\mathbb{A}_v$ is partitioned into $\dot{K}_v$-invariant subcollections with projection constants ξ.

To prepare the proof, we need to introduce a compatible condition of gluing finite-index subgroups. A collection of finite-index subgroups $\{G'_e, G'_v \mid v \in \mathscr{G}^0, e \in \mathscr{G}^1\}$ is called *compatible* if whenever $v = e_-$, we have

$$G_v \cap G'_e = G'_v \cap G_e.$$

By [24, Theorem 7.51], a compatible collection of finite-index subgroups gives a finite-index subgroup of G. The following result says that upon taking finite-index subgroups, we can assume that each vertex group is a direct product in a CKA group.

Lemma 4.8 *Let $\{\dot{K}_v < K_v \mid v \in \mathscr{G}^0\}$ be a collection of finite-index subgroups. Then there exist finite-index subgroups $\ddot{K}_v$ of $\dot{K}_v$, G'_e of G_e and Z_v of $Z(G_v)$ such that the collection of finite-index subgroups $\{G'_e, G'_v = \ddot{K}_v \times Z_v \mid v \in \mathscr{G}^0, e \in \mathscr{G}^1\}$ is compatible.*

Assuming Lemma 4.8, we now complete the proof of Proposition 4.7.

Proof of Proposition 4.7 We pass to further finite-index subgroups $\ddot{K}_v < \dot{K}_v$ satisfying compatible conditions, which then gives a further indexed subgroup $\ddot{G} \subset \dot{G}$. For $i = 1, 2$, let us partition $\mathbb{A}_i = \bigcup_{k=1}^{n_i} \mathbb{A}^i_k$ into $\ddot{G}$-obits $\mathbb{A}^i_k$. By the construction of $\ddot{G}$, we know that $\ddot{G}$ intersects each vertex group G_v of the Bass–Serre tree in a (conjugate) subgroup $\ddot{K}_v$. Thus, for each k, $\mathbb{A}^i_k$ are the union of certain $\ddot{K}_v$-invariant subcollections where v are varied in $\mathscr{V}_i$.

Recall that $\mathbb{A}^i$ for $i = 1, 2$ satisfies the projection axioms with a uniform projection constant ξ in Section 4.2. We can then build the quasitrees $T^i_k := \mathscr{C}_K(\mathbb{A}^i_k)$ where $1 \le k \le n_i$. Setting $n = n_1 + n_2$, this thus yields isometric group actions of $\ddot{G}$ on quasitrees T_i $(1 \le i \le n)$.

We first construct a $\ddot{G}$-equivariant map Φ from $\dot{\mathscr{X}}_1 \times \dot{\mathscr{X}}_2 \times T$ to $T_1 \times T_2 \times \cdots \times T_n \times T$. By equivariance, it suffices to fix a basepoint in each $\mathscr{X}_1$, $\mathscr{X}_2$, T and T_i so that Φ sends basepoints to basepoints. The quasi-isometric embedding property follows from the distance formula (10), where the right-hand side is now replaced by the distance in the corresponding quasitrees.

Note that $\ddot{G}$ is of finite index in $\dot{G}$. By taking more copies of quasitrees T_i in the target, the map Φ can be made $\dot{G}$-equivariant. Indeed, if a finite-index subgroup $H \subset G$ acts on some space X then G acts on a finite product of $[G : H]$ copies of X without preserving the factors. The map Φ can be extended to these copies as well. □

Proof of Lemma 4.8 Assume that $\langle f_v \rangle = Z(G_v)$ for any $v \in \mathscr{G}^0$. Then for an oriented edge $e = [v, w]$ from v to w, the subgroup $\langle f_v, f_w \rangle$ is of finite index in G_e.

Note that $G_e \cong \mathbb{Z}^2$ admits a base $\{\hat{f}_v, \hat{b}_e\}$ where $\hat{f}_v$ is primitive so that f_v is some power of $\hat{f}_v$. Let $\pi_v : G_v \to H_v = G_v / Z(G_v)$. Thus, $\pi_v(G_e)$ is a direct product of a torsion group with $\langle b_e \rangle$ in H_v, where $b_e = \pi_v(\hat{b}_e)$ is a loxodromic element.

Similarly, let $\hat{f}_w, \hat{b}_{\bar{e}} \in G_e$ such that $\langle \hat{f}_w, \hat{b}_{\bar{e}} \rangle = G_e$. Keep in mind that for any integer $n \neq 0$,

$$\langle \hat{f}_v^n, \hat{b}_e^n \rangle = \langle \hat{b}_{\bar{e}}^n, \hat{f}_w^n \rangle$$

is of finite index in G_e.

We choose an integer $m \neq 0$ such that $\hat{b}_e^m \in \dot{K}_v$ for every vertex $v \in \mathscr{G}^0$ and every oriented edge e from $e_- = v$. Such an integer m exists since $\dot{K}_v$ injects into H_v as a finite-index subgroup, and $\mathscr{G}$ is a finite graph of groups.

Apply the omnipotence of H_v to the independent set of elements $\{b_e \mid e_- = v\}$. Let p_v be the constant given by Definition 4.6. Set

$$s := m \prod_{v \in \mathscr{G}^0} p_v.$$

Set $l_v = s/p_v$. Thus, for the collection $\{b_e \mid e_- = v\}$, there exists a finite quotient $\xi_v : H_v \to \overline{H}_v$ such that $\xi_v(b_e)$ has order $s = l_v p_v$ and $b_e^s \in \ker(\xi_v)$. Then $\ddot{K}_v := \dot{K}_v \cap \pi_v^{-1} \ker(\xi_v)$ is of finite index in $\dot{K}_v$. Recall that $\pi_v|_{K_v} : K_v \to H_v$ is injective (see the paragraph before Lemma 4.8). Since $\pi_v(\hat{b}_e^s) = b_e^s$ is loxodromic in H_v and $\hat{b}_e^s \in \dot{K}_v$ for $m|s$, we have that $\hat{b}_e^s$ is a loxodromic element in $\ddot{K}_v$.

For each oriented edge $e = [v, w] \in \mathscr{G}^1$, define

$$G'_v := \langle \hat{f}_v^s \rangle \times \ddot{K}_v, \quad G'_w := \langle \hat{f}_w^s \rangle \times \ddot{K}_w, \quad G'_e := \langle \hat{f}_v^s, \hat{b}_e^s \rangle = \langle \hat{b}_{\bar{e}}^s, \hat{f}_w^s \rangle < G'_v.$$

Let $g \in G_e \cap G'_v$ be any element so we can write $g = \hat{f}_v^{sm} k$ for some $m \in \mathbb{Z}$ and $k \in \ddot{K}_v$. Recall that $\pi_v(G_e)$ is a direct product of $\langle b_e \rangle$ and a torsion group, and $\ddot{K}_v$ is torsion-free. So

$$\pi_v(g) = \pi_v(k) \in \pi_v(G_e) \cap \pi_v(\ddot{K}_v)$$

is some power of b_e; $\pi_v(k) = b_e^l$ for some $l \in \mathbb{Z}$. Note that $b_e^l = \pi_v(k) \in \ker(\xi_v)$ so omnipotence implies that $s|l$, ie $l = ns$ for some $n \in \mathbb{Z}$. Since $b_e = \pi_v(\hat{b}_e)$ and $\pi_v \colon \ddot{K}_v \to H_v$ is injective, we obtain that $k = \hat{b}_e^{ns}$. Therefore, $g = \hat{f}_v^{sm}\hat{b}_e^{ns} \in G'_e$ which implies

$$G_v \cap G'_e = G'_v \cap G_e.$$

Therefore, the collection $\{G'_v, G'_e \mid v \in \mathcal{G}^0, e \in \mathcal{G}^1\}$ is verified to be compatible. □

4.4 Outline of the proof of Theorem 1.3

Let $G \curvearrowright X$ be a CKA action where G is the fundamental group of an admissible graph of groups $\mathcal{G}$ such that for every vertex group the central extension (1) has omnipotent hyperbolic quotient group. Recall that property (QT) is preserved undertaking finite-index subgroups (see Lemma 2.3). Upon passing to further indexed subgroups in Lemma 4.8, we can assume that $G_v = H_v \times \mathbb{Z}$, where H_v acts geometrically on $\overline{Y}_v$ and also we can assume $\dot{G} = G$. To show the property (QT) of G, we must find not only a suitable action on a finite product of quasitrees, but also ensure the distance of points in the image can recover word distance in the ambient group. We briefly describe here the strategy of the proof. Details are given in Sections 5 and 6.

Thanks to Proposition 4.7, we know that there exists a G-equivariant quasi-isometric embedding (note that $\dot{G} = G$)

$$\dot{\mathcal{X}}_1 \times \dot{\mathcal{X}}_2 \times T \to T_1 \times T_2 \times \cdots \times T_n \times T.$$

Here T_i (with $i \in \{1, 2, \dots, n\}$) is a quasitree. As the geometry of space $\dot{\mathcal{X}}_1 \times \dot{\mathcal{X}}_2 \times T$ does not capture the distance from vertical parts of X, there is no way finding a quasi-isometric embedding from the orbit Go to $\dot{\mathcal{X}}_1 \times \dot{\mathcal{X}}_2 \times T$. To overcome this obstacle, in Section 5, we will construct two additional quasitrees, denoted by $\mathcal{C}_K(\mathbb{F}_1)$ and $\mathcal{C}_K(\mathbb{F}_2)$, and will show that there is indeed a G-equivariant quasi-isometric embedding

$$\Phi \colon Go \to \mathcal{C}_K(\mathbb{F}_1) \times \mathcal{C}_K(\mathbb{F}_2) \times \dot{\mathcal{X}}_1 \times \dot{\mathcal{X}}_2 \times T$$

(Section 6 is devoted to constructing Φ and verifying G-equivariant quasi-isometric embedding of Φ). As a consequence, we obtain the desirable G-equivariant quasi-isometric embedding

$$Go \to \mathcal{C}_K(\mathbb{F}_1) \times \mathcal{C}_K(\mathbb{F}_2) \times T_1 \times T_2 \times \cdots \times T_n \times T$$

which entails property (QT) of G.

5 Projection system of fiber lines

Recall we partition $T^0 = \mathcal{V}_1 \cup \mathcal{V}_2$ where $\mathcal{V}_i$ consists of vertices in T with pairwise even distances. For convenience, we sometimes write $\mathcal{V} = \mathcal{V}_1$ and $\mathcal{W} = \mathcal{V}_2$. We note that property (QT) of a group is preserved under taking a finite-index subgroup (see Lemma 2.3). Thus passing to a finite-index subgroup (see Lemma 2.23) if necessary we could assume that G is torsion-free and preserves $\mathcal{V}_i$ with $i = 1, 2$.

Note that $e = [w, v]$ is an oriented edge from w towards v, and $\bar{e} = [v, w]$ is the oriented edge from v towards w. For each oriented edge e, let F_e be the corresponding boundary plane. It is clear that $F_e = F_{\bar{e}}$ does not depend on the orientation.

5.1 Desired quasilines

By Lemma 2.17, the CKA space X decomposes as the union of vertex spaces $\widetilde{Y}_v = N_D(Y_v)$ for $v \in T^0$, on which the vertex groups G_v act geometrically. The center $Z(G_v) \simeq \mathbb{Z}$ allows us to split Y_v as a metric product $\overline{Y}_v \times \mathbb{R}$. Upon passing to further finite-index subgroups in Lemma 4.8, we can assume that $G_v = H_v \times \mathbb{Z}$, where H_v acts geometrically on $\overline{Y}_v$. If the CKA action $G \curvearrowright X$ is not flip (as in [43]), the system of fiber lines $\mathbb{R}$ in $Y_v = \overline{Y}_v \times \mathbb{R}$ does not behave well with respect to the G-action. We introduce better geometric models for vertex subgroups in order to resolve the G-action of fiber lines. As in [29], these models are the metric product of $\overline{Y}_v$ with a quasiline.

We first explain the construction of the quasiline obtained from a quasimorphism. The following lemma is cited from Lemma 4.2 and the proof of Corollary 4.3 in [29]. We present their proof as it is short and crucial for our discussion.

Lemma 5.1 *Let H be a hyperbolic group relative to a finite collection of virtually cyclic subgroups $\{E_i \mid 1 \le i \le n\}$. Consider $G = H \times \mathbb{Z}$ and fix a set of elements $c_i \in E_i \times \mathbb{Z}$ for each $1 \le i \le n$ such that $\langle c_i \rangle$ has unbounded projection to E_i. Then there exist a generating set S of G and a (λ, λ)-quasi-isometry $\varphi\colon \mathrm{Cay}(G, S) \to \mathbb{R}$ such that the following holds.*

(1) *If $g\langle c_i \rangle$ and $g'\langle c_i \rangle$ are two $\langle c_i \rangle$-cosets for $g, g' \in E_i \times \mathbb{Z}$, then*

$$\lambda^{-1}|g\langle c_i \rangle - g'\langle c_i \rangle|_G - \lambda \le |\varphi(g\langle c_i \rangle) - \varphi(g'\langle c_i \rangle)| \le \lambda |g\langle c_i \rangle - g'\langle c_i \rangle|_G + \lambda$$

where $|g\langle c_i \rangle - g'\langle c_i \rangle|_G$ denotes the distance between two subsets in G equipped with a word metric relative to a finite generating set (*so not the distance on* $\mathrm{Cay}(G, S)$).

(2) *With the natural action of $G \to H$, the diagonal action of $G = H \times \mathbb{Z}$ on $H \times \mathrm{Cay}(G, S)$ is metrically proper and cobounded, where $\mathbb{Z} \subset G$ acts loxodromically on $\mathrm{Cay}(G, S)$ but $\langle c_i \rangle$ acts boundedly.*

In applications, the choice of elements c_i shall come from the fiber generator of the adjacent pieces. See Lemma 5.2 below.

Proof Let $\pi_H : G = H \times \mathbb{Z} \to H$ and $\pi_{\mathbb{Z}} : G = H \times \mathbb{Z} \to \mathbb{Z}$ be the natural projections. Let $t_i = \pi_H(c_i) \in E_i$ be the projection to H of the element c_i. We then choose a quasimorphism $\phi_i : H \to \mathbb{R}$ by [32] such that $\phi_i(t_i) = 1$ but $\phi_i(E_k) = 0$ if $E_k \ne E_i$. Define the quasimorphism of $G \to \mathbb{R}$ as follows: for any $x \in G$,

$$\varphi(x) := \pi_{\mathbb{Z}}(x) - \sum_{i=1}^{n} \pi_{\mathbb{Z}}(c_i) \cdot (\phi_i \circ \pi_H(x)).$$

By definition, φ takes the constant value on $\langle c_i\rangle$-cosets. Moreover, the distance $|g\langle c_i\rangle - g'\langle c_i\rangle|_G$ is bi-Lipschitz to $|\varphi(g\langle c_i\rangle) - \varphi(g'\langle c_i\rangle)|$ with a constant depending only on $\langle c_i\rangle$.

To find the generating set S, notice that the homogenization of φ (still denoted by φ) has a bounded distance to the original one. As φ is unbounded, there exists $h \in G$ such that $\{\varphi(h^n) = n\varphi(h) \mid n \in \mathbb{Z}\}$ is an infinite cyclic subgroup. Let $S := \varphi^{-1}([0, 2\varphi(h)])$ be a (possibly infinite) subset of G. One can prove that S generates G, and $\varphi\colon G \to \mathbb{R}$ induces a desired quasi-isometry $\varphi\colon \mathrm{Cay}(G, S) \to \mathbb{R}$. See [1, Lemma 4.15] for details. $\square$

5.2 New geometric model for vertex spaces

Recall that G acts on the Bass–Serre tree T with finitely many vertex orbits. Let $\{v_1, v_2, \dots, v_n\} \subset T$ be the full set of vertex representatives, and let S_{v_i} be the (infinite) generating set for G_{v_i} given by Lemma 5.1. Then G_{v_i} acts on the quasiline $\mathfrak{fl}(v_i) := \mathrm{Cay}(G_{v_i}, S_{v_i})$. Let v be an arbitrary vertex in T, so that $v = gv_i$ for some $g \in G$ and $i \in \{1, 2, \dots, n\}$. By equivariance, we define the quasiline $\mathfrak{fl}(v) := g\mathfrak{fl}(v_i) = g\,\mathrm{Cay}(G_{v_i}, S_{v_i})$, and the action of $G_{gv_i} = gG_{v_i}g^{-1}$ on $g\mathfrak{fl}(v_i)$ is induced from the action of G_{v_i} on $\mathfrak{fl}(v_i)$.

Consider the word metric on G given by a finite generating set of G including a finite generating set of G_{v_i} for each representative vertex v_i. Equipping each vertex group G_v with a word metric, the inclusion of G_v into G is a quasi-isometric embedding since Y_v is quasi-isometrically embedded in the CAT(0) space X.

Write $X_v := \overline{Y}_v \times \mathfrak{fl}(v)$ for the new geometric model for G_v. By Lemma 5.1, the diagonal action $G_v \curvearrowright X_v$ is metrically proper and cobounded, and hence the induced orbital map

$$G_v \to G_v o' \subset X_v$$

is a G_v-equivariant quasi-isometry for any basepoint $o' = (o'_1, o'_2) \in X_v$.

Let us fix a basepoint $o = (o_1, o_2) \in Y_v$. As G_v acts freely and geometrically on $Y_v = \overline{Y}_v \times \mathbb{R}$, let

$$G_v o \to G_v$$

be a bijective G_v-equivariant quasi-isometry, a quasi-inverse to the orbital map of $G_v \curvearrowright Y_v$.

Choose the same first coordinate $o_1 = o'_1$ for the above basepoints o and o'. Define a G_v-equivariant map $\Lambda_v\colon Y_v \to X_v$ as the composite of the above two G-equivariant maps

$$\Lambda_v\colon Y_v = \overline{Y}_v \times \mathbb{R} \to G_v \to X_v = \overline{Y}_v \times \mathfrak{fl}(v).$$

Define the horizontal and vertical projection maps

$$\Lambda_v^{\mathrm{hor}}\colon Y_v \to \overline{Y}_v, \quad \Lambda_v^{\mathrm{ver}}\colon Y_v \to \mathfrak{fl}(v) \tag{11}$$

as the composites of the map Λ_v with the projections to the factor $\overline{Y}_v$ and $\mathfrak{fl}(v)$ respectively. For the product space $X_v = \overline{Y}_v \times \mathfrak{fl}(v)$, we define similarly the horizontal distance and vertical distances $|\cdot|^{\mathrm{hor}}_{X_v}$ and $|\cdot|^{\mathrm{ver}}_{X_v}$. In terms of these notations, we have for any $x, y \in Y_v$,

$$|\Lambda_v(x) - \Lambda_v(y)|^{\mathrm{hor}}_{X_v} = |\Lambda^{\mathrm{hor}}_v(x) - \Lambda^{\mathrm{hor}}_v(y)|_{\overline{Y}_v},$$
$$|\Lambda_v(x) - \Lambda_v(y)|^{\mathrm{ver}}_{X_v} = |\Lambda^{\mathrm{ver}}_v(x) - \Lambda^{\mathrm{ver}}_v(y)|_{\mathfrak{fl}(v)}.$$

We now derive a few important facts from Lemma 5.1 about Λ_v.

Recall that each piece Y_v of the CKA space X splits as a metric product $\overline{Y}_v \times \mathbb{R}$. In this context, a *fiber line* in Y_v refers to a subset $\{x\} \times \mathbb{R}$ of Y_v where $x \in \overline{Y}_v$.

Let $\pi_v \colon Y_v \to \overline{Y}_v$ be the natural projection map coming from the splitting $Y_v = \overline{Y}_v \times \mathbb{R}$. We remark that π_v and Λ^{hor}_v are not the same.

Lemma 5.2 *There exists a uniform constant $\lambda > 0$ such that Λ_v is a (λ, λ)-quasi-isometry: for any $x, y \in Y_v$,*

$$\frac{1}{\lambda}|\Lambda_v(x) - \Lambda_v(y)|_{X_v} - \lambda \le |x - y|_{Y_v} \le \lambda|\Lambda_v(x) - \Lambda_v(y)|_{X_v} + \lambda.$$

Moreover, let Y_w be the adjacent piece of Y_v in the CKA space X. Let ℓ and ℓ' be lines in the plane $P = Y_v \cap Y_w$ such that ℓ and ℓ' are fibers in Y_w. Then the following hold:

1. $\mathrm{diam}(\Lambda^{\mathrm{ver}}_v(\ell)) \le \lambda$. *In other words, $\Lambda_v(\ell) \subset \overline{Y}_v \cap B(a, \lambda)$ in $\overline{Y}_v \times \mathfrak{fl}(v)$ for some $a \in \mathfrak{fl}(v)$.*
2. *Let $p \in Y_v = \overline{Y}_v \times \mathbb{R}$ be any point and $\pi_v(p)$ be the projection of p into the factor $\overline{Y}_v$. Then $|\pi_v(p) - \Lambda^{\mathrm{hor}}_v(p)|_{\overline{Y}_v} \le \lambda$.*
3. *Denote by $|\ell - \ell'|_{Y_v}$ the distance between ℓ and ℓ' in Y_v. Then*

$$\lambda^{-1}|\ell - \ell'|_{Y_v} - \lambda \le \mathrm{diam}_{\mathfrak{fl}(v)}(\Lambda^{\mathrm{ver}}_v(\ell) \cup \Lambda^{\mathrm{ver}}_v(\ell')) \le \lambda|\ell - \ell'|_{Y_v} + \lambda.$$

Proof We first prove (2). Choose the fixed basepoints $o = (o_1, o_2)$ in Y_v and $o' = (o'_1, o'_2)$ in $\overline{Y}_v \times \mathfrak{fl}(v)$ such that their projections into the factor $\overline{Y}_v$ are the same: $o_1 = o'_1 \in \overline{Y}_v$. Take any point $p = (a, t)$ in $Y_v = \overline{Y}_v \times \mathbb{R}$, so $\pi_v(p) = a$. By our definition of the G_v-equivariant quasi-inverse $Y_v \to G_v$, there exists a group element $g \in G_v$ such that $|go - p|_{Y_v} \le \lambda$ for some uniform constant λ. We write $g = (h, n)$ in $H_v \times \mathbb{Z}$. Note that G_v acts on $\overline{Y}_v \times \mathfrak{fl}(v)$ diagonally; thus the image of the group element $g = (h, n)$ under the composition map

$$G_v \to \overline{Y}_v \times \mathfrak{fl}(v) \to \overline{Y}_v$$

is $h \cdot o_1$, where the first one is the orbital map and the second one is the projection map. If Y_v is equipped with L^1-metric, it follows that $|ho_1 - a|_{\overline{Y}_v} \le |go - p|_{Y_v} \le \lambda$. As the map Λ_v descends to the map $\overline{Y}_v \to \overline{Y}_v$ sending a to $h(o_1)$, our claim is confirmed:

$$|\Lambda^{\mathrm{hor}}_v(x) - \Lambda^{\mathrm{hor}}_v(y)|_{\overline{Y}_v} - \lambda \le d^h(x, y) \le \lambda + |\Lambda^{\mathrm{hor}}_v(x) - \Lambda^{\mathrm{hor}}_v(y)|_{\overline{Y}_v}.$$

For part (1), as there are only finitely many isometric types of Y_v of X, we only need to prove that $\mathrm{diam}(\Lambda_v^{\mathrm{ver}}(\ell)) \le \lambda$ for one given Y_v. Indeed, recall that $\Lambda_v^{\mathrm{ver}}\colon Y_v \to \mathfrak{fl}(v)$ factors through

$$Y_v \to X_v = \overline{Y}_v \times \mathfrak{fl}(v)$$

as the natural projection $X_v \to \mathfrak{fl}(v)$. The latter agrees with the quasimorphism $\varphi\colon G_v \to \mathbb{R}$ up to a bounded error in the proof of Lemma 5.1, vanishing on the center $Z(G_w)$. If $B(a,\lambda)$ denotes the ball at some element $a \in \mathfrak{fl}(v)$ with radius λ, it follows that $Z(G_w)o \subset \overline{Y}_v \times B(a,\lambda)$. Every fiber line ℓ in Y_w lies in a uniform neighborhood of the orbit of a $Z(G_w)$-coset. Our second claim is thus verified.

Part (3) is clear from our construction. □

5.3 Projection maps

Recall $T^0 = \mathcal{V}_1 \cup \mathcal{V}_2$ where $\mathcal{V}_i$ consists of vertices in T with pairwise even distances. Let

$$\mathbb{F}_1 = \{\mathfrak{fl}(v) \mid v \in \mathcal{V}_1\}, \quad \mathbb{F}_2 = \{\mathfrak{fl}(w) \mid w \in \mathcal{V}_2\}.$$

It remains to define a family of projection maps for them.

Definition 5.3 (projection maps in $\mathbb{F}_i$) Let $e_1 = [v,w]$, $e_2 = [w,v_2]$ denote the first two (oriented) edges in $[v,v']$. Let $F_{e_1} = Y_v \cap Y_w$ and $F_{e_2} = Y_{v_2} \cap Y_w$ be the two boundary planes of Y_w. Let $\mathcal{S}_{e_1e_2}$ be the strip in Y_w joining two boundary plane F_{e_1} and F_{e_2} of Y_w (see Section 2.3.1 for the definition of strips). We note that $\mathcal{S}_{e_1e_2} \cap F_{e_1}$ is a line in F_{e_1} that is parallel to a fiber in Y_w. We then define the *projection from* $\mathfrak{fl}(v')$ *into* $\mathfrak{fl}(v)$ to be

$$\Pi_{\mathfrak{fl}(v)}(\mathfrak{fl}(v')) := \Lambda_v^{\mathrm{ver}}(\mathcal{S}_{e_1e_2} \cap F_{e_1}),$$

where Λ_v^{ver} defined in (11) is the vertical projection to the quasiline in $X_v = \overline{Y}_v \times \mathfrak{fl}(v)$.

Lemma 5.4 *Let $\lambda > 0$ be the constant given by Lemma 5.2. Let a, b and c be distinct vertices in $\mathcal{V}_i$ with $i = 1,2$. If $d_T(a,[b,c]) \ge 2$ then $\Pi_{\mathfrak{fl}(a)}(\mathfrak{fl}(c)) = \Pi_{\mathfrak{fl}(a)}(\mathfrak{fl}(b)) \le \lambda$.*

Proof Let $[b,a]$ and $[c,a]$ be the geodesics in the tree T connecting b and c to w respectively. Let $e \cdot e'$ be the last two edges in $[b,a]$ (that is also the last two edges in $[c,a]$). Let $\mathcal{S}_{ee'}$ be the strip in Y_{e_+} connecting two boundary planes F_e and $F_{e'}$ of Y_{e_+} By our definition of projection maps, we have that $\Pi_{\mathfrak{fl}(a)}(\mathfrak{fl}(c)) = \Pi_{\mathfrak{fl}(a)}(\mathfrak{fl}(b)) = \Lambda_a^{\mathrm{ver}}(\mathcal{S}_{ee'} \cap P_{e'}) \le \lambda$. □

5.4 Projection axioms

We are now going to verify that $\mathbb{F}_i$ $(i = 1,2)$ with the above-defined projection maps in Definition 5.3 satisfy the projection axioms (see Definition 2.24). For each vertex $v \in T$, let $\mathbb{L}_v$ be the collection of boundary lines in the hyperbolic space $\overline{Y}_v$ defined at the beginning of Section 4. Let ℓ_1, ℓ_2 and ℓ_3 be three distinct boundary lines in $\mathbb{L}_v$. We write

$$d_{\ell_1}(\ell_2,\ell_3) = \mathrm{diam}(\pi_{\ell_1}(\ell_2) \cup \pi_{\ell_1}(\ell_3))$$

where $\pi_{\ell_i}(\ell_j)$ is the shortest projection of ℓ_j to ℓ_i in the CAT(0) hyperbolic space $\overline{Y}_v$ (note that $\overline{Y}_v$ is a hyperbolic space since H_v acts geometrically on $\overline{Y}_v$ and H_v is a nonelementary hyperbolic group). Recall that

$$d_{\mathfrak{fl}(v_1)}(\mathfrak{fl}(v_2), \mathfrak{fl}(v_3)) := \operatorname{diam}\bigl(\Pi_{\mathfrak{fl}(v_1)}(\mathfrak{fl}(v_2)) \cup \Pi_{\mathfrak{fl}(v_1)}(\mathfrak{fl}(v_3))\bigr).$$

Lemma 5.5 *There exists a uniform constant $\lambda > 0$ such that the following holds. Let v_1, v_2 and v_3 be distinct vertices in $\mathcal{V}_1$ such that v_1, v_2 and v_3 are in $\mathrm{Lk}(o)$ for some vertex o in $\mathcal{V}_2$. Let e_i denote the edge $[v_i, o]$ with $i = 1, 2, 3$ and let F_{e_i} be the plane in X associated to e_i. For each $i = 1, 2, 3$, let ℓ_i denote the boundary line of $\overline{Y}_o$ that is the projection of F_{e_i} into $\overline{Y}_o$. Then*

$$\frac{1}{\lambda} d_{\ell_1}(\ell_2, \ell_3) - \lambda \le d_{\mathfrak{fl}(v_1)}(\mathfrak{fl}(v_2), \mathfrak{fl}(v_3)) \le \lambda d_{\ell_1}(\ell_2, \ell_3) + \lambda.$$

Proof Let $\mathcal{S}_{e_1e_2}$ and $\mathcal{S}_{e_1e_3}$ be the strips in Y_o connecting the planes F_{e_1} to F_{e_2} and F_{e_1} to F_{e_3} respectively. We denote the line $\mathcal{S}_{e_1e_2} \cap F_{e_1}$ by ℓ and denote the line $\mathcal{S}_{e_1e_3} \cap F_{e_1}$ by ℓ'. Note that both lines ℓ and ℓ' are fibers in Y_o. Recall that by our definition of projection maps, we have $\Pi_{\mathfrak{fl}(v_1)}(\mathfrak{fl}(v_2)) = \Lambda^{\mathrm{ver}}_{v_1}(\ell)$ and $\Pi_{\mathfrak{fl}(v_1)}(\mathfrak{fl}(v_3)) = \Lambda^{\mathrm{ver}}_{v_1}(\ell')$. By part (3) of Lemma 5.2, for some $\lambda > 0$, we have that

$$\frac{1}{\lambda}|\ell - \ell'| - \lambda \le \operatorname{diam}\bigl(\Pi_{\mathfrak{fl}(v_1)}(\mathfrak{fl}(v_2)) \cup \Pi_{\mathfrak{fl}(v_1)}(\mathfrak{fl}(v_3))\bigr) \le \lambda|\ell - \ell'| + \lambda.$$

Note that $|\ell - \ell'| = d_{\ell_1}(\ell_2, \ell_3)$ (indeed, let α and β be the shortest geodesics joining ℓ_2 to ℓ_1 and ℓ_3 to ℓ_1 respectively; then ℓ and ℓ' are the product $\alpha_+ \times \mathbb{R}$ and $\beta_+ \times \mathbb{R}$ of endpoints of α and β, respectively, with the $\mathbb{R}$ direction in $Y_o = \overline{Y}_o \times \mathbb{R}$). Combining the above inequalities, we obtain a constant $\lambda' = \lambda'(\lambda) > 0$ still denoted by λ such that

$$\frac{1}{\lambda} d_{\ell_1}(\ell_2, \ell_3) - \lambda \le \operatorname{diam}\bigl(\Pi_{\mathfrak{fl}(v_1)}(\mathfrak{fl}(v_2)) \cup \Pi_{\mathfrak{fl}(v_1)}(\mathfrak{fl}(v_3))\bigr) \le \lambda d_{\ell_1}(\ell_2, \ell_3) + \lambda. \qquad \square$$

We are now going to prove the following.

Lemma 5.6 *There exists a constant $\xi > 0$ such that for each $i \in \{1, 2\}$, the collection $\mathbb{F}_i$ with projection maps $\pi_{\mathfrak{fl}(v)}$ satisfies the projection axioms with projection constant ξ.*

Proof We verify in order the projection axioms (see Definition 2.24) for the projection maps defined on $\mathbb{F}_1$. The case for $\mathbb{F}_2$ is symmetric. The constant ξ will be defined explicitly during the proof.

Axiom 1 Let $\lambda > 0$ be the constant given by Lemma 5.2. Since $\mathcal{S}_{e_1e_2} \cap F_{e_1}$ is a fiber line in Y_w, it follows from Lemma 5.2 that $\operatorname{diam} \Lambda^{\mathrm{ver}}_v(\mathcal{S}_{e_1e_2} \cap F_{e_1}) \le \lambda$. Thus $\operatorname{diam}\bigl(\Pi_{\mathfrak{fl}(v)}(\mathfrak{fl}(v'))\bigr) \le \lambda$. Axiom 1 in Definition 2.24 is verified.

Axiom 2 Let u, v and w be distinct vertices in $\mathcal{V}_1$. We will show that there exists $\xi \ge 0$ sufficiently large that if $d_{\mathfrak{fl}(w)}(\mathfrak{fl}(u), \mathfrak{fl}(v)) > \xi$, then $d_{\mathfrak{fl}(u)}(\mathfrak{fl}(w), \mathfrak{fl}(v)) \le \xi$ or $d_{\mathfrak{fl}(v)}(\mathfrak{fl}(w), \mathfrak{fl}(u)) \le \xi$. The constant ξ will be defined explicitly during the proof. Since $d_{\mathfrak{fl}(w)}(\mathfrak{fl}(u), \mathfrak{fl}(v)) > \xi$, it follows from Lemma 5.4 that there is some restriction on w, ie w is either lies on $[u, v]$ or $d_T(w, [u, v]) = 1$.

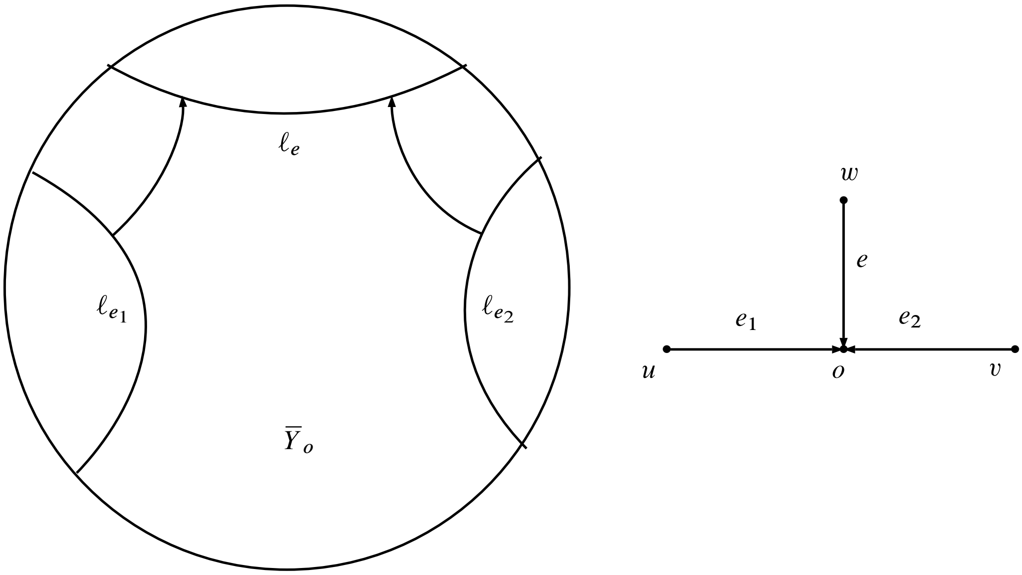


Figure 2: Verification of Axiom 2.

Case 1 Suppose w lies on $[u,v]$. Since $u,w,v\in\mathcal{V}_1$, we have $d_T(u,[w,v])\geq 2$ and $d_T(v,[u,w])\geq 2$. Axiom 2 thus follows from Lemma 5.4.

Case 2 Suppose $d_T(w,[u,v])=1$. Without loss of generality, we can assume that u, v and w lie in the same link $\mathrm{Lk}(o)$ for some vertex o in $\mathcal{V}_2$. Indeed, let $o\in[u,v]$ be adjacent to w and $u',v'\in\mathrm{Lk}(o)\cap[u,v]$. It is clear by definition that $\pi_{\mathfrak{fl}(u)}(\mathfrak{fl}(v'))=\pi_{\mathfrak{fl}(u)}(\mathfrak{fl}(v))$ and $\pi_{\mathfrak{fl}(v)}(\mathfrak{fl}(u'))=\pi_{\mathfrak{fl}(v)}(\mathfrak{fl}(u))$. As a result, we can thus assume that $u=u'$ and $v=v'$ lie in the link $\mathrm{Lk}(o)$.

Recall that $\overline{Y}_o$ is a δ-hyperbolic space whose boundary lines $\mathbb{L}_o$ satisfy the projection axioms for a constant ξ_0 [48]. We claim that $\xi=\xi_0$ is the desired constant for Axiom 2.

Write $e=[w,o]$, $e_1=[u,o]$ and $e_2=[v,o]$. Let ℓ_e, ℓ_{e_1} and ℓ_{e_2} be the corresponding boundary lines of $\overline{Y}_o$ to the oriented edges e, e_1 and e_2. By Lemma 5.5, we have

$$\frac{1}{\lambda}d_{\ell_e}(\ell_{e_1},\ell_{e_2})-\lambda\leq d_{\mathfrak{fl}(w)}(\mathfrak{fl}(u),\mathfrak{fl}(v))\leq\lambda d_{\ell_e}(\ell_{e_1},\ell_{e_2})+\lambda.$$

As $\mathbb{L}_o$ satisfies the projection axioms, we see that if $d_{\ell_e}(\ell_{e_1},\ell_{e_2})>\xi_0$, then $d_{\ell_{e_1}}(\ell_e,\ell_{e_2})\leq\xi_0$. Using Lemma 5.5 again, we have that

$$\frac{1}{\lambda}d_{\ell_{e_1}}(\ell_e,\ell_{e_2})-\lambda\leq d_{\mathfrak{fl}(u)}(\mathfrak{fl}(w),\mathfrak{fl}(v))\leq\lambda d_{\ell_{e_1}}(\ell_e,\ell_{e_2})+\lambda.$$

Let ξ be a constant such that $\xi>\lambda\xi_0+\lambda$. It follows from the above inequalities that

$$d_{\mathfrak{fl}(u)}(\mathfrak{fl}(w),\mathfrak{fl}(v))=\mathrm{diam}\big(\Pi_{\mathfrak{fl}(u)}(\mathfrak{fl}(w))\cup\Pi_{\mathfrak{fl}(u)}(\mathfrak{fl}(v))\big)\leq\xi,$$

so Axiom 2 is verified.

Axiom 3 For $u \neq v \in \mathcal{V}_1$, the set

$$\{w \in \mathcal{V}_1 \mid d_{\mathfrak{fl}(w)}(\mathfrak{fl}(u), \mathfrak{fl}(v)) > \xi\}$$

is a finite set.

Indeed, by Lemma 5.4, such a w is either contained in the interior of $[u, v]$ or $d(w, [u, v]) = 1$. The first case yields only $(d(u, v) - 1)$ choices for w. We now consider the case $d(w, [u, v]) = 1$. Since u, v and w have pairwise even distance, there exists $o \in \mathcal{W} \cap [u, v]^0$ and two vertices u' and v' on $[u, v]$ adjacent to o such that $u', v', w \in \mathrm{Lk}(o)$. By the projection axioms of boundary lines $\mathbb{L}_o$ of $\overline{Y}_o$, the set of w satisfying $d_{\mathfrak{fl}(w)}(\mathfrak{fl}(u), \mathfrak{fl}(v)) > \xi$ is finite. Thus, in both cases, the set of such w is finite. □

Lemma 5.7 *For each $i = 1, 2$, the collection $\mathbb{F}_i = \{\mathfrak{fl}(v) \mid v \in \mathcal{V}_i\}$ admits an action of the group G such that*

$$\Pi_{g\mathfrak{fl}(v)}(g\mathfrak{fl}(u)) = g\Pi_{\mathfrak{fl}(v)}(\mathfrak{fl}(u))$$

for any $v, u \in \mathcal{V}_i$ and any $g \in G$.

Proof First, let us recall some discussion in the beginning of Section 5.2. Recall that $\{v_1, v_2, \dots, v_n\} \subset T$ is the full set of vertex representatives of T and for each representative vertex $v_1, v_2, \dots, v_n$ of T, the quasiline $\mathfrak{fl}(v_j)$ is the Cayley graph $\mathrm{Cay}(G_{v_j}, S_{v_j})$ for some generating set S_{v_j} of G_{v_j} (see Lemma 5.1). Let v be an arbitrary vertex in T; then $v = gv_i$ for some $g \in G$ and $i \in \{1, 2, \dots, n\}$. The quasiline $\mathfrak{fl}(v)$ is given by $g\mathfrak{fl}(v_i) = g\,\mathrm{Cay}(G_{v_i}, S_{v_i})$, and the action of $G_{gv_i} = gG_{v_i}g^{-1}$ on $g\mathfrak{fl}(v_i)$ is induced from the action of G_{v_i} on $\mathfrak{fl}(v_i)$. We are now going to show that

$$\Pi_{g\mathfrak{fl}(v)}(g\mathfrak{fl}(u)) = g\Pi_{\mathfrak{fl}(v)}(\mathfrak{fl}(u)).$$

Recall that the family of maps $\Lambda^{\mathrm{ver}}_{gv}\colon Y_{gv} = gY_v \to g\mathfrak{fl}(v)$ are G-equivariant: $\Lambda^{\mathrm{ver}}_{gv}(gx) = g\Lambda^{\mathrm{ver}}_v(x)$ for all $x \in Y_v$. Let e_1 and e_2 be the first two edges in the geodesic $[v, u]$ with $v = (e_1)_-$ and $(e_1)_+ = (e_2)_-$. By Definition 5.3 of projection map, we have that

$$\begin{aligned}\Pi_{\mathfrak{fl}(gv)}(\mathfrak{fl}(gu)) &= \mathrm{diam}\big(\Lambda^{\mathrm{ver}}_{gv}(\mathcal{S}_{ge_1ge_2} \cap F_{ge_1})\big)\\ &= \mathrm{diam}\big(\Lambda^{\mathrm{ver}}_{gv}(g(\mathcal{S}_{e_1e_2} \cap F_{e_1}))\big)\\ &= \mathrm{diam}\big(g\Lambda^{\mathrm{ver}}_{v}(\mathcal{S}_{e_1e_2} \cap F_{e_1})\big)\\ &= g\,\mathrm{diam}\big(\Lambda^{\mathrm{ver}}_{v}(\mathcal{S}_{e_1e_2} \cap F_{e_1})\big) = g\Pi_{\mathfrak{fl}(v)}(\mathfrak{fl}(u))\end{aligned}$$

for any $g \in G$. □

Definition 5.8 Let $\xi > 0$ be the projection constant given by Lemma 5.6, so the collection of quasilines $\mathbb{F}_i = \{\mathfrak{fl}(v) \mid v \in \mathcal{V}_i\}$ with $i = 1, 2$ satisfies the projection axioms. For any fixed $K > 4\xi$, we obtain the unbounded quasitrees of metric spaces $\mathcal{C}_K(\mathbb{F}_1)$ and $\mathcal{C}_K(\mathbb{F}_2)$ (see Section 2.4). Combining Lemma 5.7 with [5, Section 4.4], the spaces $\mathcal{C}_K(\mathbb{F}_1)$ and $\mathcal{C}_K(\mathbb{F}_2)$ are quasitrees and admit unbounded isometric actions $G \curvearrowright \mathcal{C}_K(\mathbb{F}_1)$ and $G \curvearrowright \mathcal{C}_K(\mathbb{F}_2)$. The quasitrees $\mathcal{C}_K(\mathbb{F}_1)$ and $\mathcal{C}_K(\mathbb{F}_2)$ are called *vertical quasitrees* hereafter.

6 Distance formulas in the CKA space X

Let $\mathscr{C}_K(\mathbb{F}_1)$ and $\mathscr{C}_K(\mathbb{F}_2)$ be the vertical quasitrees in Definition 5.8. Let $\dot{\mathscr{X}}_1$ and $\dot{\mathscr{X}}_2$ be the coned-off spaces defined in Section 4.1. According to the outline of the proof of Theorem 1.3 in Section 4.4, the last step to prove property (QT) of G is to show that there is a G-equivariant quasi-isometric embedding

$$\Phi\colon Go \to \mathscr{C}_K(\mathbb{F}_1) \times \mathscr{C}_K(\mathbb{F}_2) \times \dot{\mathscr{X}}_1 \times \dot{\mathscr{X}}_2 \times T.$$

This section is devoted to constructing such a desired map Φ and verifying it is a quasi-isometric embedding.

We list here notation that will be used in the rest of this section.

- We fix an edge $[v_0, w_0]$ in the Bass–Serre tree T such that $v_0 \in \mathscr{V}_1$. Let $o \in X$ be a basepoint in the common boundary plane $F_{[v_0,w_0]}$ between two pieces Y_{v_0} and Y_{w_0}.
- Assume that $x = o \in Y_{v_0}$ and $y = go \in Y_{v_{2n}}$ for some $g \in G$ and $v_{2n} = go$. We list the vertices on the geodesic $[v_0, v_{2n}]$ by $\{v_0, v_1, \dots, v_{2n}\}$ where $v_{2i} \in \mathscr{V}_1$ and $v_{2i+1} \in \mathscr{V}_2$. Let $e_{i+1} = [v_i, v_{i+1}]$ be the oriented edge towards v_{i+1}. By definition of special paths, let $p_i := \mathscr{S}_{e_{i-1}e_i} \cap \mathscr{S}_{e_i e_{i+1}}$ be the intersection of two strips with $p_0 : x = o$ and $p_{2n+1} = y = go$.
- Let α be the geodesic edge path in the Bass–Serre tree T connecting v_0 to v_{2n}. And let $w_1 \in \mathscr{V}_2$ be a vertex adjacent to v_{2n}. Set

$$\tilde{\alpha} := e_0 \cup \alpha \cup e_{2n+1}$$

  where $e_0 = [w_0, v_0]$ and $e_{2n+1} = [v_{2n}, w_1]$. It is possible that $e_0 = \bar{e}_1$ and $e_{2n+1} = \bar{e}_{2n}$, ie $\tilde{\alpha}$ contains backtracking at e_0 and e_{2n}.

6.1 Construction of the desired map Φ

It is a product of the following four maps with the index map ρ in Definition 2.19.

- We define $\vartheta_1\colon Go \to \mathscr{C}_K(\mathbb{F}_1)$ as follows. Recall that each quasiline $\mathfrak{fl}(v)$ for $v \in \mathscr{V}_1$ embeds as a convex subset into $\mathscr{C}_K(\mathbb{L}_1)$ and $\Lambda_v^{\mathrm{ver}}\colon G_v o \to \mathfrak{fl}(v)$ is a G_v-equivariant map. For every $g \in G$, we set $\vartheta_1(go) := \Lambda_{gv_0}^{\mathrm{ver}}(go) = g\Lambda_{v_0}^{\mathrm{ver}}(o)$. The second equality follows by G_v-equivariance.
- Similarly, define $\vartheta_2\colon Go \to \mathscr{C}_K(\mathbb{F}_2)$ by $\vartheta_2(go) := \Lambda_{gw_0}^{\mathrm{ver}}(go) = g\Lambda_{w_0}^{\mathrm{ver}}(o)$ for every $g \in G$.
- Define $\vartheta_3(o) := \pi_{\overline{Y}_{v_0}}(o)$ and extend the definition by equivariance so that $\vartheta_3(go) := g\vartheta_3(o)$ for any $g \in G$. We thus obtain a G-equivariant map $\vartheta_3\colon Go \to \dot{\mathscr{X}}_1$.
- Choose $\vartheta_4(o)$ to be the cone point of the hyperbolic cone attached to the boundary line $\ell_{[v_0,w_0]}$ of $\overline{Y}_{w_0}$. We then extend $\vartheta_4(go) = g\vartheta_4(o)$ for any $g \in G$ so that $g\vartheta_4(o)$ is the corresponding cone point to $\ell_{[gv_0,gw_0]}$ of $\overline{Y}_{gw_0}$. We thus obtain a G-equivariant map $\vartheta_4\colon Go \to \dot{\mathscr{X}}_2$.

We then define

$$(\clubsuit)\qquad \Phi\colon Go \to \mathscr{C}_K(\mathbb{F}_1)\times\mathscr{C}_K(\mathbb{F}_2)\times(\dot{\mathscr{X}}_1, d^K_{\dot{\mathscr{X}}_1})\times(\dot{\mathscr{X}}_2, d^K_{\dot{\mathscr{X}}_2})\times T$$

by

$$\Phi := \vartheta_1\times\vartheta_2\times\vartheta_3\times\vartheta_4\times\rho$$

where $\dot{\mathscr{X}}_i$ for $i=1,2$ are equipped with the K-thick distance $d^K_{\dot{\mathscr{X}}_i}$ (not genuine distance) defined in (9), and the other three spaces are equipped with length metric. By abuse of language, we call the sum of the distances over the factors the L^1-metric on the product space.

The remainder of this section is to verify the following.

Proposition 6.1 *The map Φ in ($\clubsuit$) is a G-equivariant quasi-isometric embedding.*

Idea of the proof of Proposition 6.1 Since the orbital map of any isometric action is Lipschitz (see eg [11, Lemma I.8.18]), we will only need to give a linear upper bound on $|x-y|_X$. Recall from (2) in Section 2.3.2, for any $x,y\in X$,

$$|x-y|_X \sim |x-y|^{\mathrm{hor}}_X + |x-y|^{\mathrm{ver}}_X$$

where $|x-y|^{\mathrm{hor}}_X = \sum_{i=0}^{2n}|p_i-p_{i+1}|^{\mathrm{hor}}_{Y_{v_i}}$ and $|x-y|^{\mathrm{ver}}_X = \sum_{i=0}^{2n}|p_i-p_{i+1}|^{\mathrm{ver}}_{Y_{v_i}}$.

Recall from Section 5.2, we build a new geometric model X_v of Y_v for each vertex v in the Bass–Serre tree T. Namely, we have a G_v-equivariant quasi-isometric map $\Lambda_v\colon Y_v=\overline{Y}_v\times\mathbb{R}\to X_v=\overline{Y}_v\times\mathfrak{fl}(v)$. For $x,y\in Go$, we shall accordingly replace $|x-y|^{\mathrm{ver}}_X$ by the quantity

$$V(x,y) := \sum_{0\le i\le 2n} |\Lambda^{\mathrm{ver}}_{v_i}(p_i)-\Lambda^{\mathrm{ver}}_{v_i}(p_{i+1})|_{\mathfrak{fl}(v_i)}. \tag{12}$$

To be precise, we first prove in Lemma 6.2 that

$$|x-y|_X \le \epsilon\big(|\rho(x)-\rho(y)|_T + |x-y|^{\mathrm{hor}}_X + V(x,y)\big),$$

and then we find suitable upper bounds of $V(x,y)$ (see Proposition 6.4) and $|x-y|^{\mathrm{hor}}_X$ (see Lemma 6.6).

6.2 Verifying Φ is a quasi-isometric embedding

In this section, we will verify that the map Φ in ($\clubsuit$) is a quasi-isometric embedding.

6.2.1 Upper bound of the distance $|x-y|_X$ on X

Lemma 6.2 *Let $x,y\in Go$. The exists a constant $\epsilon>0$ such that*

$$|x-y|_X \le \epsilon\big(|\rho(x)-\rho(y)|_T + |x-y|^{\mathrm{hor}}_X + V(x,y)\big). \tag{13}$$

Proof Recall that $p_0=x$ and $p_{2n+1}=y$. Using the triangle inequality we have

$$|x-y|_X \le \sum_{i=0}^{2n}|p_i-p_{i+1}|_{Y_{v_i}}.$$

Note that $2n = |\rho(x)-\rho(y)|_T$ and $|x-y|_X^{\mathrm{hor}} = \sum_{i=0}^{2n} |p_i - p_{i+1}|_{Y_{v_i}}^{\mathrm{hor}}$. The proof is then complete by summing over $0 \le i \le 2n$ the following inequality (14).

Claim *There exists a uniform constant $\epsilon' > 0$ such that for any $i \in \{0, 1, \dots, 2n\}$,*

$$|p_i - p_{i+1}|_{Y_{v_i}} \le \epsilon' + \epsilon' |p_i - p_{i+1}|_{Y_{v_i}}^{\mathrm{hor}} + \epsilon' |\Lambda_{v_i}^{\mathrm{ver}}(p_i) - \Lambda_{v_i}^{\mathrm{ver}}(p_{i+1})|_{\mathfrak{fl}(v_i)}. \tag{14}$$

Proof of the claim Indeed, since $\Lambda_{v_i} : Y_{v_i} \to X_{v_i}$ is a quasi-isometry by Lemma 5.2, we then have

$$|p_i - p_{i+1}|_{Y_{v_i}} \sim_\lambda |\Lambda_{v_i}(p_i) - \Lambda_{v_i}(p_{i+1})|_{X_{v_i}}.$$

Using part (2) of Lemma 5.2 we have that

$$|\Lambda_{v_i}^{\mathrm{hor}}(p_i) - \Lambda_{v_i}^{\mathrm{hor}}(p_{i+1})|_{\overline{Y}_{v_i}} \sim_\lambda |p_i - p_{i+1}|_{Y_{v_i}}^{\mathrm{hor}}.$$

It implies that

$$\begin{aligned} |\Lambda_{v_i}(p_i)) - \Lambda_{v_i}(p_{i+1})|_{X_{v_i}} &\sim_{\sqrt{2}} |\Lambda_{v_i}^{\mathrm{hor}}(p_i) - \Lambda_{v_i}^{\mathrm{hor}}(p_{i+1})|_{\overline{Y}_{v_i}} + |\Lambda_{v_i}^{\mathrm{ver}}(p_i) - \Lambda_{v_i}^{\mathrm{ver}}(p_{i+1})|_{\mathfrak{fl}(v_i)} \\ &\sim_\lambda |p_i - p_{i+1}|_{Y_{v_i}}^{\mathrm{hor}} + |\Lambda_{v_i}^{\mathrm{ver}}(p_i) - \Lambda_{v_i}^{\mathrm{ver}}(p_{i+1})|_{\mathfrak{fl}(v_i)} \end{aligned}$$

where the first coarse equality holds by definition of $\Lambda_{v_i}^{\mathrm{hor}}$ and $\Lambda_{v_i}^{\mathrm{ver}}$. Hence there exists a uniform constant $\epsilon' > 0$ such that the inequality (14) holds. □

The lemma is proved. □

6.2.2 Preparation for upper bounds of $V(x, y)$ and $|x-y|_X^{\mathrm{hor}}$ Fix $K \ge 4\xi$ where the constant $\xi > \lambda$ is given by Lemma 5.6. Let $\mathscr{C}_K(\mathbb{F}_1)$ and $\mathscr{C}_K(\mathbb{F}_2)$ be the vertical quasitrees given by Definition 5.8. With $i \in \{1, 2\}$, Proposition 2.27 gives the distance formula

$$|\vartheta_i(x) - \vartheta_i(y)|_{\mathscr{C}_K(\mathbb{F}_i)} \sim_K \sum_{\mathfrak{fl}(w) \in \mathbb{F}_i} [d_{\mathfrak{fl}(w)}(\vartheta_i(x), \vartheta_i(y))]_K. \tag{$\circledast$}$$

To give an appropriate upper bound of $V(x, y)$, we need the following two technical lemmas (Lemmas 6.3 and 6.5).

Lemma 6.3 *For any $v_{2i} \in [v_0, v_{2n}]$ with $0 \le i \le n$, we have*

$$d_{\mathfrak{fl}(v_{2i})}(\vartheta_1(x), \vartheta_1(y)) \sim_\lambda |\Lambda_{v_{2i}}^{\mathrm{ver}}(p_{2i}) - \Lambda_{v_{2i}}^{\mathrm{ver}}(p_{2i+1})|_{\mathfrak{fl}(v_{2i})}. \tag{15}$$

For any $0 \le i \le n-1$, we have

$$d_{\mathfrak{fl}(v_{2i+1})}(\vartheta_2(x), \vartheta_2(y)) \sim_\lambda |\Lambda_{v_{2i+1}}^{\mathrm{ver}}(p_{2i+1}) - \Lambda_{v_{2i+1}}^{\mathrm{ver}}(p_{2i+2})|_{\mathfrak{fl}(v_{2i+1})}. \tag{16}$$

Proof We first prove (15) for the case $0 < i < n$. The cases $i = 0$ or $i = n$ are similar.

Note that $\ell_1 := \mathscr{S}_{e_{2i-1}e_{2i}} \cap F_{e_{2i}}$ is a fiber line of $Y_{v_{2i-1}}$ containing p_{2i}, and similarly,

$$\ell_2 := \mathscr{S}_{e_{2i+1}e_{2i+2}} \cap F_{e_{2i+1}}$$

contains p_{2i+1}. By Definition 5.3 of projection maps, we have

$$\Pi_{\mathfrak{fl}(v_{2i})}(\mathfrak{fl}(v_0)) = \Lambda^{\mathrm{ver}}_{v_{2i}}(\ell_1) \quad \text{and} \quad \Pi_{\mathfrak{fl}(v_{2i})}(\mathfrak{fl}(v_{2n})) = \Lambda^{\mathrm{ver}}_{v_{2i}}(\ell_2).$$

Let $\lambda > 0$ be the constant given by Lemma 5.2, so the fiber lines ℓ_1 and ℓ_2 are sent by $\Lambda^{\mathrm{ver}}_{v_{2i}}$ into $\mathfrak{fl}(v_{2i})$ as subsets of diameter at most λ:

$$\operatorname{diam} \Lambda^{\mathrm{ver}}_{v_{2i}}(\ell_1), \operatorname{diam} \Lambda^{\mathrm{ver}}_{v_{2i}}(\ell_2) \le \lambda.$$

By definition of ϑ_1, we have $\vartheta_1(x) = \Lambda^{\mathrm{ver}}_{v_0}(x) \in \mathfrak{fl}(v_0)$ and $\vartheta_1(y) = \Lambda^{\mathrm{ver}}_{v_{2n}}(y) \in \mathfrak{fl}(v_{2n})$. Thus,

$$d_{\mathfrak{fl}(v_{2i})}(\vartheta_1(x), \vartheta_1(y)) \sim_\lambda d_{\mathfrak{fl}(v_{2i})}(\mathfrak{fl}(v_0), \mathfrak{fl}(v_{2n})).$$

As $p_{2i} \in \ell_1$ and $p_{2i+1} \in \ell_2$, we obtain

$$d_{\mathfrak{fl}(v_{2i})}(\mathfrak{fl}(v_0), \mathfrak{fl}(v_{2n})) \sim_\lambda |\Lambda^{\mathrm{ver}}_{v_{2i}}(p_{2i}) - \Lambda^{\mathrm{ver}}_{v_{2i}}(p_{2i+1})|_{\mathfrak{fl}(v_{2i})},$$

completing the proof of (15).

We are now going to prove (16). If $w_0 \ne v_1$ or $1 \le i \le n-1$, the same proof for (15) proves (16). We now consider $w_0 = v_1$ and $i = 0$. In this case, we note that $e_0 = \bar{e}_1$. By definition, we have that $\vartheta_2(x) = \vartheta_2(o) = \Lambda^{\mathrm{ver}}_{w_0}(o) \in \mathfrak{fl}(w_0)$, so we obtain $\Pi_{\mathfrak{fl}(v_1)}(\vartheta_2(x)) = \vartheta_2(x)$. Recall that $\mathcal{S}_{xe_1}$ is the strip in Y_{v_0} over the shortest arc from x to F_{e_1} (see construction of special path). As $x \in F_{e_0} = F_{e_1}$, we have $\ell_1 := \mathcal{S}_{xe_1}$ is a fiber line of Y_{v_0} that passes through x and also p_1. Thus, $\vartheta_2(x) \in \Pi_{\mathfrak{fl}(v_1)}(\ell_1)$.

Recall that $\mathcal{S}_{xe_1}$ is the strip in Y_{v_0} over the shortest arc from x to F_{e_1} (see construction of special path). As $x \in F_{e_0} = F_{e_1}$, we have that $\ell_1 := \mathcal{S}_{xe_1}$ is a fiber line of Y_{v_0} that passes through x and also p_1. Thus, $\vartheta_2(x) \in \Pi_{\mathfrak{fl}(v_1)}(\ell_1)$. Let $\ell_2 = \mathcal{S}_{e_2e_3} \cap F_{e_2}$ be the fiber line on Y_{v_2} that passes through p_2. If $w_1 = v_1$, then $\alpha = [v_0, v_1][v_1, v_2]$ and $y \in F_{e_2}$. By the same reason, ℓ_2 passes through y, so $\vartheta_2(y) \in \Pi_{\mathfrak{fl}(v_2)}(\ell_2)$. If $w_1 \ne v_1$, the projection $\Pi_{\mathcal{L}_{v_1}}(\vartheta_2(y))$ must be contained in $\Pi_{\mathfrak{fl}(v_1)}(\ell_2)$. In both cases, we have

$$d_{\mathfrak{fl}(v_1)}(\vartheta_2(x), \vartheta_2(y)) \sim_\lambda \operatorname{diam}(\Lambda^{\mathrm{ver}}_{v_1}(\ell_1) \cup \Lambda^{\mathrm{ver}}_{v_1}(\ell_2))$$

where we use $\operatorname{diam}(\Lambda^{\mathrm{ver}}_{v_1}(\ell_1)), \operatorname{diam}(\Lambda^{\mathrm{ver}}_{v_1}(\ell_2)) \le \lambda$ by Lemma 5.2. For $p_1 \in \ell_1$ and $p_2 \in \ell_2$, we obtain

$$\operatorname{diam}(\Lambda^{\mathrm{ver}}_{v_1}(\ell_1) \cup \Lambda^{\mathrm{ver}}_{v_1}(\ell_2)) \sim_\lambda |\Lambda^{\mathrm{ver}}_{v_1}(p_1) - \Lambda^{\mathrm{ver}}_{v_1}(p_2)|_{\mathfrak{fl}(v_1)},$$

concluding the proof of (16). □

Let us recall the notation from Section 2.4. Let $x \in \mathfrak{fl}(v), z \in \mathfrak{fl}(u) \in \mathbb{F}_i$.

If $\mathfrak{fl}(v) \ne \mathfrak{fl}(u) \ne \mathfrak{fl}(w)$ then

$$d_{\mathfrak{fl}(w)}(x, z) := d_{\mathfrak{fl}(w)}(\mathfrak{fl}(v), \mathfrak{fl}(u)).$$

If $\mathfrak{fl}(w) = \mathfrak{fl}(v)$ and $\mathfrak{fl}(w) \ne \mathfrak{fl}(u)$, define $d_{\mathfrak{fl}(w)}(x, z) := \operatorname{diam}(\pi_{\mathfrak{fl}(w)}(x, \mathfrak{fl}(u)))$.

If $\mathfrak{fl}(v) = \mathfrak{fl}(u) = \mathfrak{fl}(w)$, let $d_{\mathfrak{fl}(w)}(x, z)$ be the distance in $\mathfrak{fl}(w)$.

6.2.3 Upper bound of $V(x,y)$ Let ϑ_1 and ϑ_2 be the maps defined in Section 6.1. We now have prepared all ingredients for the proof of the following result.

Proposition 6.4 *Let $x, y \in Go$ and $\alpha := [\rho(x), \rho(y)]$ be the geodesic in T. Then*

$$V(x,y) \preceq_K \sum_{j=1,2} \Bigg(\sum_{v \in \alpha \cap \mathcal{V}_j} [d_{\mathfrak{fl}(v)}(\vartheta_j(x), \vartheta_j(y))]_K \Bigg) + d_T(\rho(x), \rho(y)). \tag{17}$$

Proof The goal is to recover the sum on the right side of (12), that is

$$V(x,y) = \sum_{0 \le i \le 2n} |\Lambda^{\mathrm{ver}}_{v_i}(p_i) - \Lambda^{\mathrm{ver}}_{v_i}(p_{i+1})|_{\mathfrak{fl}(v_i)}$$

via the maps ϑ_1 and ϑ_2. By Lemma 6.3, we have the desired inequalities (15) for even indices $v_{2i} \in [v_0, v_{2n}] \cap \mathcal{V}_1$ with $0 \le i \le n$, that is

$$d_{\mathfrak{fl}(v_{2i})}(\vartheta_1(x), \vartheta_1(y)) \sim_\lambda |\Lambda^{\mathrm{ver}}_{v_{2i}}(p_{2i}) - \Lambda^{\mathrm{ver}}_{v_{2i}}(p_{2i+1})|_{\mathfrak{fl}(v_{2i})}.$$

By Lemma 6.3, the inequalities (16) recover the odd indices $v_{2i+1} \in [v_0, v_{2n}] \cap \mathcal{V}_2$ with $0 \le i \le n-1$ in (12), that is,

$$d_{\mathfrak{fl}(v_{2i+1})}(\vartheta_2(x), \vartheta_2(y)) \sim_\lambda |\Lambda^{\mathrm{ver}}_{v_{2i+1}}(p_{2i+1}) - \Lambda^{\mathrm{ver}}_{v_{2i+1}}(p_{2i+2})|_{\mathfrak{fl}(v_{2i+1})}.$$

Plugging inequalities (15) and (16) into (12), and using the term $|\rho(x) - \rho(y)|_T$ to count the additive errors in this process completes the proof of the desired inequality (17). Applying then the K-cutoff function $[\,\cdot\,]_K$ does not affect the inequalities. □

6.2.4 Upper bound of $|x-y|^{\mathrm{hor}}_X$ The horizontal distance d^h defined in (2) of the special path γ from x to y records the totality of the projected distances to the base hyperbolic spaces $\overline{Y}_v$:

$$\begin{aligned} |x-y|^{\mathrm{hor}}_X &= |x-p_1|^{\mathrm{hor}}_{Y_{v_1}} + |p_1 - p_2|^{\mathrm{hor}}_{Y_{v_2}} + \cdots + |p_{2n} - y|^{\mathrm{hor}}_{Y_{v_{2n}}} \\ &= |\vartheta_3(x) - F_{e_1}|_{Y_{v_0}} + \sum_{i=1}^{2n-1} |F_{e_i} - F_{e_{i+1}}|_{Y_{v_i}} + |F_{e_{2n}} - \vartheta_3(y)|_{Y_{v_{2n}}} \end{aligned}$$

where the map ϑ_3 defined in Section 6.1 sends a point in $Y_v = \overline{Y}_v \times \mathbb{R}$ to the hyperbolic base $\overline{Y}_v$.

Before moving on, let us introduce more notation to represent the horizontal distance. Let $x_0 = \vartheta_3(x)$, $y_0 \in F_{e_1}$ and $x_{2n} \in F_{e_{2n}}$, $y_{2n} = \vartheta_3(y)$ be such that $[x_0, y_0]$ is orthogonal to F_{e_1}, and $[x_{2n}, y_{2n}]$ to $F_{e_{2n}}$. Choose $x_i \in F_{e_i}$ and $y_i \in F_{e_{i+1}}$ so that $[x_i, y_i]$ is a geodesic in $\overline{Y}_{v_i}$ orthogonal to F_{e_i} and $F_{e_{i+1}}$. Thus,

$$|x-y|^{\mathrm{hor}}_X = \sum_{i=0}^{2n} |x_i - y_i|_{\overline{Y}_{v_i}}. \tag{18}$$

Recall that $\dot{\mathscr{X}}_1$ and $\dot{\mathscr{X}}_2$ are the coned-off spaces defined in Section 4.1. By Definition 4.3 of the K-thick distance of $\dot{\mathscr{X}}_j$ for any $K > 0$, and the remark after it, we have

$$\sum_{i=0}^{2n} |x_i - y_i|^K_{\dot{Y}_{v_i}} = |\vartheta_3(x) - \vartheta_3(y)|^K_{\dot{\mathscr{X}}_1} + |\vartheta_4(x) - \vartheta_4(y)|^K_{\dot{\mathscr{X}}_2} \tag{19}$$

where $|\cdot|^K_{\dot{Y}_{v_i}}$ defined in (3) is the K-thick distance on the coned-off space $\dot{Y}_{v_i}$. The map ϑ_4 defined in Section 6.1 sends a point go in Go to the hyperbolic cone point to the boundary line $\ell_{g[v_0,w_0]}$ (recall that o is chosen on a common boundary plane $F_{[v_0,w_0]}$).

Hence, the K-thick distance (19) differs from the horizontal distance (18) by the amount coned-off on boundary lines. The purpose of this subsection is to recover the loss in the coned-off from the projection system of fiber lines.

To prove Lemma 6.6, we need the following lemma.

Lemma 6.5 *Let ϑ_2 be the map given by Section 6.1. Let v be a vertex in $\mathrm{Lk}(v_{2i})-\tilde{\alpha}$ and let $e=[v,v_{2i}]$. Let ℓ_e, $\ell_{e_{2i}}$, and $\ell_{\bar{e}_{2i+1}}$ be the boundary lines of $\widetilde{F}_{v_{2i}}$ associated to distinct edges e, e_{2i} and $\bar{e}_{2i+1}$ respectively. Then we have*

$$d_{\mathfrak{fl}(v)}(\vartheta_2(x),\vartheta_2(y)) \sim_\lambda d_{\ell_e}(\ell_{e_{2i}},\ell_{\bar{e}_{2i+1}}). \tag{20}$$

Proof Note that ℓ_e, $\ell_{e_{2i}}$ and $\ell_{\bar{e}_{2i+1}}$ are the projection of planes F_e, $F_{e_{2i}}$ and $F_{e_{2i+1}}$ of $Y_{v_{2i}}$ into the factor $Y_{v_{2i}}$. We prove (20) case by case, according to the configuration of e_0, e_{2n+1} with α.

Case 1 Suppose $0<i<n$. By Definition 5.3, the projection of $\vartheta_2(x)=\Lambda^{\mathrm{ver}}_{w_0}(o)\in\mathfrak{fl}(w_0)$ to $\mathfrak{fl}(v)$ is the same as that of $\mathfrak{fl}(v_1)$ to $\mathfrak{fl}(v)$, and the projection of $\vartheta_2(y)\in\mathfrak{fl}(w_1)$ to $\mathfrak{fl}(v)$ is the same as that of $\mathfrak{fl}(v_{2n-1})$ to $\mathfrak{fl}(v)$. That is to say, $d_{\mathfrak{fl}(v)}(\vartheta_2(x),\vartheta_2(y))=d_{\mathfrak{fl}(v)}(\mathfrak{fl}(v_1),\mathfrak{fl}(v_{2n-1}))$. Hence, (20) follows by Lemma 5.5: $d_{\mathfrak{fl}(v)}(\mathfrak{fl}(v_1),\mathfrak{fl}(v_{2n-1}))\sim_\lambda d_{\ell_e}(\ell_{e_{2i}},\ell_{\bar{e}_{2i+1}})$ for any $v\in\mathrm{Lk}(v_{2i})-\tilde{\alpha}$.

Case 2 Suppose $i=0$ or $i=n$. We only consider the case $i=0$ and analyze the configuration of w_0 with α. The analyze for the case for $i=n$ and w_1 is symmetric.

Case 2.1 Suppose $w_0\neq v_1$. In this case $e_0\cdot\alpha$ is a geodesic from w_0 to v_{2n}. By Definition 5.3 of projection maps, no matter whether $\bar{e}_{2n+1}=e_{2n}$ holds or not, the projection of $\vartheta_2(x)\in\mathfrak{fl}(w_0)$ to $\mathfrak{fl}(v)$ is the same as that of $\mathfrak{fl}(w_0)$ to $\mathfrak{fl}(v)$, and the projection of $\vartheta_2(y)\in\mathfrak{fl}(w_1)$ to $\mathfrak{fl}(v)$ is the same as that of $\mathfrak{fl}(v_{2n-1})$ to $\mathfrak{fl}(v)$. By Lemma 5.5, we have $d_{\mathfrak{fl}(v)}(\mathfrak{fl}(w_0),\mathfrak{fl}(v_{2n-1}))\sim_\lambda d_{\ell_e}(\ell_{e_{2i}},\ell_{\bar{e}_{2i+1}})$.

Case 2.2 Suppose $w_0=v_1$. No matter whether $w_0=w_1$ or not, we have

$$d_{\mathfrak{fl}(v)}(\vartheta_2(x),\vartheta_2(y))\le \Pi_{\mathfrak{fl}(v)}(\mathfrak{fl}(w_0))\le\xi$$

where ξ is the projection constant given by Lemma 5.6. On the right side of (20), $d_{\ell_e}(\ell_{e_{2i}},\ell_{\bar{e}_{2i+1}})$ is bounded above by ξ for $i=0$ (as $e_0=\bar{e}_1$). Thus (20) holds as well in this case. □

Lemma 6.6 *For any $x,y\in Go$, we have*

$$|x-y|^{\mathrm{hor}}_X \preceq_K \sum_{i=0}^{2n}|x_i-y_i|^K_{\dot{Y}_{v_i}} + \sum_{i=0}^{2n}\sum_{w\in\mathrm{Lk}(v_i)-\alpha}[d_{\mathfrak{fl}(w)}(\vartheta_j(x),\vartheta_j(y))]_K$$

where the index $j=1$ is chosen if i is odd, otherwise $j=2$.

Proof We consider (18) for the horizontal distance $|x-y|_X^{\mathrm{hor}}$. Let $\mathbb{L}_{v_i}$ be the set of boundary lines of $\overline{Y}_{v_i}$ corresponding to the set of oriented edges $e \in \mathrm{St}(v_i)$ (ie the collection $\{F_e \cap \overline{Y}_{v_i} \mid e \in \mathrm{St}(v_i)\}$). By Lemma 3.1, for each $0 \le i \le 2n$, we have

$$|x_i - y_i|_{\overline{Y}_{v_i}} \sim_K |x_i - y_i|^K_{\dot{Y}_{v_i}} + \sum_{\ell_e \in \mathbb{L}_{v_i}} [d_{\ell_e}(x_i, y_i)]_K \tag{21}$$

for any sufficiently large $K \gg 0$.

Let $e = [w, v_i] \in \mathrm{St}(v_i)$ and $\ell_e \in \mathbb{L}_{v_i}$ be the corresponding boundary line of $\overline{Y}_{v_i}$. Set $j = 1$ if i is odd, otherwise $j = 2$.

If $e = e_i$ or $e = \bar{e}_{i+1}$ for $1 \le i \le 2n-1$, then

$$d_{\ell_e}(x_i, y_i) \le \xi$$

since $[x_i, y_i]$ is orthogonal to ℓ_e.

We remark that when $i = 0$ (the case $i = 2n$ is similar), it is possible that $[x_0, y_0]$ may not be perpendicular to ℓ_e. However, we have

$$d_{\ell_e}(x_0, y_0) \preceq d_{\mathfrak{fl}(w_0)}(\vartheta_2(x), \vartheta_2(y)).$$

Otherwise, if $e \neq e_i$ and $e \neq \bar{e}_{i+1}$ for $1 \le i \le 2n-1$, we have $e \notin \alpha$ for which the following holds by Lemma 6.5 for $j = 2$ and by Lemma 5.5 for $j = 1$:

$$d_{\mathfrak{fl}(w)}(\vartheta_j(x), \vartheta_j(y)) \sim d_{\ell_e}(x_i, y_i).$$

Note that $A \le \lambda B + C$ with $B \ge K \ge C$ implies $[A]_K \preceq_K [B]_K$. Thus, for each $0 \le i \le 2n$, we deduce from (21) that

$$|x_i - y_i|_{\overline{Y}_{v_i}} \sim_K |x_i - y_i|^K_{\dot{Y}_{v_i}} + \sum_{w \in \mathrm{Lk}(v_i) - \alpha} [d_{\mathfrak{fl}(w)}(\vartheta_j(x), \vartheta_j(y))]_K \tag{22}$$

for any $K \gg 0$, where $j = 1$ if i is odd, and $j = 2$ otherwise. We sum up (22) over $v_i \in \alpha$ to get the horizontal distance $d^h(x, y)$ in (18):

$$|x-y|_X^{\mathrm{hor}} = \sum_{i=0}^{2n} |x_i - y_i|_{\overline{Y}_{v_i}} \preceq_K \sum_{i=0}^{2n} |x_i - y_i|^K_{\dot{Y}_{v_i}} + \sum_{i=0}^{2n} \sum_{w \in \mathrm{Lk}(v_i) - \alpha} [d_{\mathfrak{fl}(w)}(\vartheta_j(x), \vartheta_j(y))]_K. \qquad \square$$

We now have prepared all ingredients in the proof of Proposition 6.1.

Proof of Proposition 6.1 Since ρ and ϑ_i (with $i \in \{1, 2, 3, 4\}$) are G-equivariant maps, it follows that Φ is a G-equivariant map. Since the orbital map of any isometric action is Lipschitz (see eg [11, Lemma I.8.18]), it suffices to give an upper bound on $d(x, y)$.

Let $\epsilon > 0$ be the constant given by Lemma 6.2, so that

$$|x-y|_X \le \epsilon\big(|\rho(x) - \rho(y)|_T + |x-y|_X^{\mathrm{hor}} + V(x, y)\big).$$

Appropriate upper bounds of the vertical distance $V(x,y)$ and the horizontal distance $|x-y|_X^{\mathrm{hor}}$ have been already treated in Proposition 6.4 and Lemma 6.6 respectively. They are

$$V(x,y) \preceq_K \sum_{j=1,2}\Bigg(\sum_{v\in\alpha\cap\mathcal{V}_j}[d_{\mathfrak{fl}(v)}(\vartheta_j(x),\vartheta_j(y))]_K\Bigg)+|\rho(x)-\rho(y)|_T$$

and

$$|x-y|_X^{\mathrm{hor}} \preceq_K \sum_{i=0}^{2n}|x_i-y_i|_{\dot{Y}_{v_i}}^K+\sum_{i=0}^{2n}\sum_{w\in\mathrm{Lk}(v_i)-\alpha}[d_{\mathfrak{fl}(w)}(\vartheta_j(x),\vartheta_j(y))]_K$$

where the index j depends on i: $j=1$ if i is odd, otherwise $j=2$. The above two inequalities yield

$$|x-y|_X^{\mathrm{hor}}+V(x,y)\preceq_K |\rho(x)-\rho(y)|_T+\sum_{i=0}^{2n}|x_i-y_i|_{\dot{Y}_{v_i}}^K+\sum_{i=0}^{2n}\sum_{w\in\mathrm{Lk}(v_i)}[d_{\mathfrak{fl}(w)}(\vartheta_j(x),\vartheta_j(y))]_K.$$

By (⊛), we have

$$\sum_{i=0}^{2n}\sum_{w\in\mathrm{Lk}(v_i)}[d_{\mathfrak{fl}(w)}(\vartheta_j(x),\vartheta_j(y))]_K\preceq_K|\vartheta_1(x)-\vartheta_1(x)|_{\mathscr{C}_K(\mathbb{F}_1)}+|\vartheta_2(x)-\vartheta_2(x)|_{\mathscr{C}_K(\mathbb{F}_2)}.$$

It follows that

$$|x-y|_X^{\mathrm{hor}}+V(x,y)\preceq_K |\rho(x)-\rho(y)|_T+\sum_{i=0}^{2n}|x_i-y_i|_{\dot{Y}_{v_i}}^K+\sum_{i=1}^{2}|\vartheta_i(x)-\vartheta_i(x)|_{\mathscr{C}_K(\mathbb{F}_i)}.$$

Plugging the thick distance formula (19) into the above inequality, we obtain

$$\begin{aligned}|x-y|_X^{\mathrm{hor}}+V(x,y)\preceq_K |\rho(x)-\rho(y)|_T&+|\vartheta_3(x)-\vartheta_3(y)|_{\dot{\mathscr{X}}_1}^K+|\vartheta_4(x)-\vartheta_4(y)|_{\dot{\mathscr{X}}_2}^K\\&+|\vartheta_1(x)-\vartheta_1(x)|_{\mathscr{C}_K(\mathbb{F}_1)}+|\vartheta_2(x)-\vartheta_2(x)|_{\mathscr{C}_K(\mathbb{F}_2)}.\end{aligned}$$

As $|x-y|_X\le\epsilon(|\rho(x)-\rho(y)|+|x-y|_X^{\mathrm{hor}}+V(x,y))$, it is a consequence from the above inequality that the map $\Phi=\vartheta_1\times\vartheta_2\times\vartheta_3\times\vartheta_4\times\rho$ in (♣) is a G-equivariant quasi-isometric embedding from X to $\mathscr{C}_K(\mathbb{F}_1)\times\mathscr{C}_K(\mathbb{F}_2)\times\dot{\mathscr{X}}_1\times\dot{\mathscr{X}}_2\times T$. □

7 Proof of Theorem 1.3

Let $G\curvearrowright X$ be a CKA action such that for every vertex group the central extension (1) has omnipotent hyperbolic quotient group. Let $\dot{G}<G$ be the subgroup of the index at most 2 preserving $\mathcal{V}_1$ and $\mathcal{V}_2$ given by Lemma 2.23. Upon passing to further finite-index subgroups in Lemma 4.8, we obtain a finite-index subgroup G' of $\dot{G}$ such that the results in Sections 5 and 6 hold for G'. We caution the reader that at the beginning of Section 5 we assume that each vertex group of G is a direct product, this assumption may not hold for the original G, but holds in the finite-index subgroup G' of G.

As G' is a subgroup of $\dot{G}$, it follows from Proposition 4.7 that there exists a G'-equivariant quasi-isometric embedding

$$\eta\colon(\dot{\mathscr{X}}_1\times\dot{\mathscr{X}}_2\times T,d_{\dot{\mathscr{X}}_1}^K\times d_{\dot{\mathscr{X}}_2}^K\times d_T)\to T_1\times T_2\times\cdots\times T_n\times T.$$

Applying Proposition 6.1 to G', we have a G'-equivariant quasi-isometric embedding

$$\Phi\colon G'o \to \mathscr{C}_K(\mathbb{F}_1) \times \mathscr{C}_K(\mathbb{F}_2) \times (\dot{\mathscr{X}}_1, d^K_{\dot{\mathscr{X}}_1}) \times (\dot{\mathscr{X}}_2, d^K_{\dot{\mathscr{X}}_2}) \times T.$$

It implies that $(\mathrm{id}_{\mathscr{C}_K(\mathbb{F}_1)} \times \mathrm{id}_{\mathscr{C}_K(\mathbb{F}_2)} \times \eta) \circ \Phi$ is a G'-equivariant quasi-isometric embedding from $G' \cdot o$ to the finite product of quasitrees $\mathscr{C}_K(\mathbb{F}_1) \times \mathscr{C}_K(\mathbb{F}_1) \times T_1 \times T_2 \times \cdots \times T_n \times T$. Thus G' has property (QT), implying G has property (QT).

8 Applications: property (QT) of 3-manifold groups

In this section, we apply results obtained in previous sections to give a complete characterization of property (QT) of all finitely generated 3-manifold groups (Theorem 1.1). Note that property (QT) is a commensurability invariant. Hence, we can always assume that all 3-manifolds are compact and orientable (by taking Scott's compact core and double cover).

Let M be a compact, connected, orientable, irreducible 3-manifold with empty or tori boundary. M is called *geometric* if its interior admits geometric structures in the sense of Thurston; those are S^3, $\mathbb{E}^3$, $\mathbb{H}^3$, $S^2 \times \mathbb{R}$, $\mathbb{H}^2 \times \mathbb{R}$, $\widetilde{\mathrm{SL}(2,\mathbb{R})}$, Nil and Sol. If M is not geometric, then M is called a *nongeometric* 3*-manifold*. By geometric decomposition of 3-manifolds, there is a nonempty minimal union $\mathscr{T} \subset M$ of disjoint essential tori and Klein bottles, unique up to isotopy, such that each component of $M \backslash \mathscr{T}$ is either a Seifert fibered piece or a hyperbolic piece. M is called *graph manifold* if all the pieces of $M \backslash \mathscr{T}$ are Seifert fibered pieces, otherwise it is a *mixed manifold*.

We remark here that the geometric decomposition is slightly different from the torus decomposition, but they are closely related (if M has no decomposing Klein bottle, then these two decompositions agree with each other). Such a difference can be got rid of by passing to some finite cover of M. Since we are only interested in virtual properties of 3-manifolds in this paper, we can always assume that these two decompositions agree with each other (on some finite cover of M). For this reason, we will only use the term torus decomposition in the remainder of this section.

8.1 Property (QT) of geometric 3-manifolds

Proposition 8.1 *The fundamental group $\pi_1(M)$ of a geometric 3-manifold M has property (QT) if and only if M does not support Sol or Nil geometry.*

Proof We are going to prove the necessity. Assume that $\pi_1(M)$ has property (QT). By Lemma 2.5, $\pi_1(M)$ does not contain any distorted element, while the fundamental group of a 3-manifold with Nil geometry or Sol geometry contains quadratically/exponentially distorted elements (for example, see [41, Proposition 1.2]). Hence, M does not support Sol or Nil geometry.

Now, we are going to prove sufficiency. If M supports geometry $\mathbb{E}^3$, S^3 or $S^2 \times \mathbb{R}$, then $\pi_1(M)$ is virtually abelian so has property (QT). If the geometry of M is $\mathbb{H}^2 \times \mathbb{R}$ then M is virtually covered by $\Sigma \times S^1$ for some hyperbolic surface Σ. Note that $\pi_1(\Sigma)$ is a residually finite hyperbolic group so it has property (QT) by [6, Theorem 1.1]. Hence, $\pi_1(\Sigma) \times \mathbb{Z}$ has property (QT). Since $\pi_1(\Sigma) \times \mathbb{Z}$ is a finite-index subgroup of $\pi_1(M)$, it follows that $\pi_1(M)$ has property (QT) by Lemma 2.3. If M supports geometry $\mathbb{H}^3$, $\pi_1(M)$ is virtually compact special by deep theorems of Agol [3] and Wise [53]; thus $\pi_1(M)$ has property (QT) since it is undistorted in a right-angled Artin group. Note that if the boundary of M is empty, then $\pi_1(M)$ is a residually finite hyperbolic group. As a result, it can be inferred that $\pi_1(M)$ possesses property (QT) as an alternative argument, according to [6, Theorem 1.1].

Finally, we need to show that if M supports $\widetilde{\mathrm{SL}(2,\mathbb{R})}$ geometry then $\pi_1(M)$ has property (QT). To see this, by passing to a finite cover if necessary, we could assume that M is a nontrivial circle bundle over a closed surface Σ with $\chi(\Sigma) < 0$. Let $1 \to K \to \pi_1(M) \to \pi_1(\Sigma) \to 1$ be the short exact sequence associated with the circle bundle where K is the normal cyclic subgroup of $\pi_1(M)$ generated by a fiber. Let $\pi : \pi_1(M) \to \pi_1(\Sigma)$ be the surjective homomorphism in the above short exact sequence. Note that the short exact sequence does not split since M is supporting $\widetilde{\mathrm{SL}(2,\mathbb{R})}$ geometry. According to the first paragraph in the proof of [29, Corollary 4.3], there exists a generating set $\mathscr{S}$ of $G = \pi_1(M)$ such that $\mathscr{L} := \mathrm{Cay}(G, \mathscr{S})$ is a quasiline. Moreover, the diagonal action of G on $\pi_1(\Sigma) \times \mathscr{L}$ is metrically proper and cobounded, and thus its orbital map is a quasi-isometry. Since $\pi_1(\Sigma)$ is a residually finite hyperbolic group, it follows from [6] that $\pi_1(\Sigma)$ has property (QT). Hence there exists a finite product of quasitrees $\prod_{i=1}^n T_i$ such that $\pi_1(\Sigma) \curvearrowright \prod_{i=1}^n T_i$ such that its orbital map is a quasi-isometric embedding. It is easy to see that the orbital map of the diagonal action $G \curvearrowright \prod_{i=1}^n T_i \times \mathscr{L}$ of G on the product $\prod_{i=1}^n T_i \times \mathscr{L}$ is a quasi-isometric embedding. Therefore $\pi_1(M)$ has property (QT). $\square$

8.2 Property (QT) of nongeometric 3-manifolds

In this section, we are going to prove Theorem 1.2. Recall that a nongeometric 3-manifold is either a graph manifold or a mixed manifold.

8.2.1 Property (QT) of graph manifolds Let M be a graph manifold. Since property (QT) is preserved under taking finite-index subgroups (see Lemma 2.3), we only need to show that a finite cover of M has property (QT). By passing to a finite cover, we can assume that each Seifert fibered piece in the JSJ decomposition of M is a trivial circle bundle over a hyperbolic surface of genus at least 2, and the intersection numbers of fibers of adjacent Seifert pieces have absolute value 1 (see [34, Lemma 2.1]). Also we can assume that the underlying graph of the graph manifold M is bipartite since any nonbipartite graph manifold is double covered by a bipartite one.

We note that $\pi_1(M)$ is an admissible group in the sense of Definition 2.12. However, it is not always true that $\pi_1(M)$ can act geometrically on a CAT(0) space, so property (QT) in this case does not follow immediately from Theorem 1.3. Indeed, if M is a graph manifold with nonempty boundary then it always admits a Riemannian metric of nonpositive curvature (see [35]). In particular, $\pi(M) \curvearrowright \widetilde{M}$ is a CKA action, and thus property (QT) of $\pi_1(M)$ follows from Theorem 1.3. However, many closed graph manifolds are shown to not support any Riemannian metric of nonpositive curvature (see [35]).

We remark here that the CAT(0) metric on the CKA space X in Sections 5 and 6 is not really essential in the proofs. Below we will make certain modifications on some steps to run the proof of Theorem 1.3 for closed graph manifolds.

Metrics on M

We now are going to describe a convenient metric on M introduced by Kapovich and Leeb [34]. For each Seifert component $M_v = F_v \times S^1$ of M, we choose a hyperbolic metric on the base surface F_v such that all boundary components are totally geodesic of unit length, and then equip each Seifert component $M_v = F_v \times S^1$ with the product metric d_v such that the fibers have length one. Metrics d_v on M_v induce the product metrics on $\widetilde{M}_v$ which by abuse of notations is also denoted by d_v.

Let M_v and M_w be adjacent Seifert components in the closed graph manifold M, and let $T \subset M_v \cap M_w$ be a JSJ torus. Each metric space $(\widetilde{T}, d_v)$ and $(\widetilde{T}, d_w)$ is a Euclidean plane. After applying a homotopy to the gluing map, we may assume that at each JSJ torus T, the gluing map ϕ from the boundary torus $\overleftarrow{T} \subset M_v$ to the boundary torus $\overrightarrow{T} \subset M_w$ is affine in the sense that the identity map $(\widetilde{T}, d_v) \to (\widetilde{T}, d_w)$ is affine. We now have a product metric on each Seifert component $M_v = F_v \times S^1$. These metrics may not agree with each other on the JSJ tori but the gluing maps are bilipschitz (since they are affine). The product metrics on the Seifert components induce a length metric on the graph manifold M denoted by d (see [12, Section 3.1] for details). Moreover, there exists a positive constant L such that on each Seifert component $M_v = F_v \times S^1$ we have

$$\frac{1}{L} d_v(x, y) \leq d(x, y) \leq L d_v(x, y)$$

for all x and y in M_v. (See [45, Lemma 1.8] for a detailed proof of the last claim.) Metric d on M induces metric on $\widetilde{M}$, which is also denoted by d (by abuse of notation). Then for all x and y in $\widetilde{M}_v$ we have

$$\frac{1}{L} d_v(x, y) \leq d(x, y) \leq L d_v(x, y).$$

Remark 8.2 Note that the space $(\widetilde{M}, d)$ may not be a CAT(0) space but $\pi_1(M)$ acts geometrically on $(\widetilde{M}, d)$ via deck transformations.

In Section 2.3.2, we define special paths on a CAT(0) space X. In this section, although $(\widetilde{M}, d)$ is no longer a CAT(0) space, we are still able to define special paths in $(\widetilde{M}, d)$. The construction is similar to Section 2.3.2 with slight changes.

Special paths on $\widetilde{M}$

Lift the JSJ decomposition of the graph manifold M to the universal cover $\widetilde{M}$, and let T be the tree dual to this decomposition of $\widetilde{M}$ (ie the Bass–Serre tree of $\pi_1(M)$). For every pair of adjacent edges e_1 and e_2 in T, let v be the common vertex of e_1 and e_2. Let ℓ and ℓ' be two boundary lines of $\widetilde{F}_v$ corresponding to the edges e_1 and e_2 respectively. Let $\gamma_{e_1 e_2}$ be the shortest geodesic joining ℓ and ℓ' in $(\widetilde{M}_v, d_v)$. This geodesic determines an Euclidean strip $\mathcal{S}_{e_1 e_2} := \gamma_{e_1 e_2} \times \mathbb{R}$ in $(\widetilde{M}_v, d_v)$. Let x be a point in $(\widetilde{M}_v, d_v)$ and e be an edge with an endpoint v. The minimal geodesic from x to the plane F_e also define a strip $\mathcal{S}_{xe} := \gamma_{xe} \times \mathbb{R}$ in $(\widetilde{M}_v, d_v)$ where $\gamma_{xe} \subset \widetilde{F}_v$ is the projection to $\widetilde{F}_v$ of this minimal geodesic.

Now, let x and y be any two points in the universal cover $\widetilde{M}$ of M such that x and y belong to the interiors of pieces $\widetilde{M}_v$ and $\widetilde{M}_v'$ respectively. If $v = v'$ then we define a *special path* in X connecting x and y to be the geodesic $[x, y]$ in $(\widetilde{M}, d)$. Otherwise, let $e_1 \cdots e_n$ be the geodesic edge path connecting v and v'. For notational purpose, we write $e_0 := x$ and $e_{n+1} := y$. Let $p_i \in F_{e_i}$ be the intersection point of the strips $\mathcal{S}_{e_{i-1} e_i}$ and $\mathcal{S}_{e_i e_{i+1}}$. The *special path* connecting x and y is the concatenation of the geodesics

$$[x, p_1] \cdot [p_1, p_2] \cdots [p_n, y].$$

We label $p_0 := x$ and $p_{n+1} := y$.

Proposition 8.3 *If M is a graph manifold, then $\pi_1(M)$ has property (QT).*

Proof If M is a nonpositively curved graph manifold (for example, when M has nonempty boundary) then the fact that $\pi_1(M)$ has property (QT) is followed from Theorem 1.3. The only case that does not follow directly from Theorem 1.3 is when M is a closed graph manifold (recall many closed graph manifolds are nonpositively curved but many are not). Since the metric d on $\widetilde{M}$ restricted to each piece $\widetilde{M}_v$ is L-bilipschitz equivalent to d_v, so the inequalities in Section 6 are slightly changed by a uniform multiplicative constant. For example, the statement $a \asymp_K b$ (or $a \preceq_K b$) in Section 6 will be changed to $a \asymp_{K'} b$ (or $a \preceq'_K b$) for some constant K' depending on K. Thus, the proof, in this case, is performed along the same lines as the proof of Theorem 1.3. □

8.2.2 Property (QT) of mixed 3-manifolds Recall that a nongeometric 3-manifold with empty or tori boundary is either a graph manifold or a mixed 3-manifold. The case of graph manifold has been addressed in Section 8.2.1. In this section, we address the mixed 3-manifold case.

Proposition 8.4 *The fundamental group of a mixed 3-manifold has property (QT).*

The fundamental group of a mixed 3-manifold has a natural relatively hyperbolic structure as follows: Let $M_1, \ldots, M_k$ be the maximal graph manifold pieces, isolated Seifert fibered components of the JSJ-decomposition of M, and $S_1, \ldots, S_l$ be the tori in M not contained in any M_i. The fundamental group $G = \pi_1(M)$ is hyperbolic relative to the set of parabolic subgroups

$$\mathcal{P} = \{\pi_1(M_p) \mid 1 \le p \le k\} \cup \{\pi_1(S_q) \mid 1 \le q \le l\}$$

(see [8; 22]).

The following lemma provides many separable subgroups in $\pi_1(M)$, generalizing [50, Lemma 3.3]. The proof uses a recent result of the second author and Sun in [41] where the authors show that separability and distortion of subgroups in 3-manifold groups are closely related.

Lemma 8.5 *Let M be a compact, orientable, irreducible 3-manifold with empty or tori boundary, with nontrivial torus decomposition and that does not support the Sol geometry. If H is a finitely generated, undistorted subgroup of $\pi_1(M)$, then H is separable in $\pi_1(M)$.*

Proof Let M_H be the covering space of M corresponding to $H \leq \pi_1(M)$. Generalizing a notion called "almost fiber part" in [36], an embedded (possibly disconnected) subsurface $\Phi(H)$ in M_H called an "almost fiber surface" is introduced in [49]. Sun [49, Theorem 1.3] shows that all information about the separability of H can be obtained by examining the almost fibered surface.

In [41], the authors introduce a notion called "modified almost fibered surface" (denoted by $\hat{\Phi}(H)$) that slightly modifies the original definition of almost fibered surface in [49] and show that information about the distortion of H in G can be also be obtained by examining the "modified almost fibered surface". We refer the reader to [49] for the definition of the almost fiber surface and to [41] for the definition of the modified almost fiber surface. The precise definitions are not needed here, so we only state here some facts from [41] that will be used later in the proof.

The torus decomposition of M induces the torus decomposition of $\Phi(H)$. Let $\Phi(H)$ and $\hat{\Phi}(H)$ be the almost fiber surface and modified almost fiber surface of H respectively.

(1) Both the almost fiber surface $\Phi(H)$ and the modified almost fiber surface $\hat{\Phi}(H)$ are (possibly disconnected) subsurfaces of M_H.

(2) The almost fiber surface $\Phi(H)$ has some piece that is homeomorphic to the annulus and parallel to the boundary of $\Phi(H)$. We delete these annulus pieces from $\Phi(H)$ to get the modified almost fiber surface, and we denote it by $\hat{\Phi}(H)$.

The surface $\Phi(H)$ (resp. $\hat{\Phi}(H)$) has a natural graph of spaces structure with the dual graph denoted by $\Gamma_{\Phi(H)}$ (resp. $\Gamma_{\hat{\Phi}(H)}$). By [41, Theorem 1.4], every component S of the modified almost fiber surface $\hat{\Phi}(H)$ must contain only one piece (otherwise, the distortion of H in $\pi_1(M)$ is at least quadratic, this contradicts the fact that H is undistorted in $\pi_1(M)$). This fact combined with (2) implies that the graph $\Gamma_{\Phi(H)}$ is a union of trees. By [49, Theorem 1.3] (or see also [50, Theorem 3.2] for a statement) tells us that whenever $\Gamma_{\Phi(H)}$ does not contain a simple cycle then H is separable. As shown above, we are in this case; hence we conclude that the subgroup H is separable in $\pi_1(M)$. □

Proof of Proposition 8.4 Let $M_1, \dots, M_k$ be the collection of maximal graph manifold components and Seifert fibered pieces in the geometric decomposition of M. Let $S_1, \dots, S_\ell$ be the tori in the boundary of M that bound a hyperbolic piece, and let $T_1, \dots, T_m$ be the tori in the JSJ decomposition of M that separate two hyperbolic components of the JSJ decomposition. Then $\pi_1(M)$ is hyperbolic relative to

$$\mathbb{P} = \{\pi_1(M_p)\}_{p=1}^{k} \cup \{\pi_1(S_q)\}_{q=1}^{\ell} \cup \{\pi_1(T_r)\}_{r=1}^{m}$$

(see [8; 22]).

We are going to show that $G = \pi_1(M)$ satisfies all conditions in Theorem 1.5.

Claim 1 *$\pi_1(M)$ induces the full profinite topology on each $P \in \mathcal{P}$.*

Indeed, it is well known that the fundamental groups of all compact 3-manifolds are residually finite; thus $\pi_1(M)$ is residually finite. Since each peripheral subgroup P is undistorted in $\pi_1(M)$, it follows from Lemma 8.5 that P is separable in $\pi_1(M)$. Again, by Lemma 8.5, each finite-index subgroup of P is also separable in $\pi_1(M)$. By [47, Lemma 2.8], $\pi_1(M)$ induces the full profinite topology on P.

Claim 2 *For each peripheral subgroup $P \in \mathbb{P}$, there exists a finite-index subgroup P' of P acting isometrically on a finite number of quasitrees so that the diagonal action of P' on the finite product of these quasitrees induces quasi-isometric embeddings on orbital maps.*

Indeed, if $P = \pi_1(T_r)$ or $P = \pi_1(S_q)$ for some r or q then $\pi_1(P) = \mathbb{Z}^2$; we let $P' := P$. If $P = \pi_1(M_j)$ for some Seifert component $M_j = F_j \times S^1$ then $P = \pi_1(F_j) \times \mathbb{Z}$. In this case, as F_j is a hyperbolic surface with nonempty boundary, $\pi_1(F_j)$ is a free group; hence we choose $P' = P$ as $\pi_1(F_j)$ is a quasitree. The last case we must consider is that $P = \pi_1(M_j)$ where M_j is a maximal graph manifold component. Passing to an appropriate finite cover $M'_j \to M_j$ we can assume that $\pi_1(M'_j)$ acts on a finite number of quasitrees (but they are not quasilines) $T_1, T_2, \dots, T_n$ so that the orbital map induced from the diagonal action $\pi_1(M_j) \curvearrowright \prod_{1=1}^n T_i$ is a quasi-isometric embedding (see Proposition 8.3). Claim 2 is confirmed. We then repeat the proof of Theorem 3.5 (the second and third paragraph) to show that P satisfies the hypothesis of Theorem 1.5.

In summary, we have verified the hypotheses in Theorem 1.5 for $G = \pi_1(M)$, so mixed 3-manifold groups have property (QT). □

Proof of Theorem 1.2 Let M be a compact orientable irreducible 3-manifold with empty or tori boundary, with nontrivial torus decomposition, and that does not support the Sol geometry. Such a 3-manifold M is either a graph manifold or a mixed manifold. The graph manifold case and mixed manifold case have been addressed in Propositions 8.3 and 8.4, respectively, and hence the theorem is proved. □

8.3 Property (QT) of finitely generated 3-manifolds

Proposition 8.6 *Let M be a compact, orientable, irreducible, ∂-irreducible 3-manifold such that it has a boundary component of genus at least 2. Then $\pi_1(M)$ has property (QT).*

Proof We consider the following two cases:

Case 1 M has trivial torus decomposition. In this case, M supports a geometrically finite hyperbolic structure with infinite volume. We paste hyperbolic 3-manifolds with totally geodesic boundaries to M to get a finite volume hyperbolic 3-manifold N. By the covering theorem (see [16]) and the subgroup tameness theorem (see [2; 15]), a finitely generated subgroup of the finite volume hyperbolic 3-manifold N is either a virtual fiber surface subgroup or undistorted. By the construction of N, the subgroup $\pi_1(M) \le \pi_1(N)$ could not be a virtual fiber surface subgroup, and thus $\pi_1(M)$ must be undistorted in $\pi_1(N)$. Since $\pi_1(N)$ has property (QT), it follows that $\pi_1(M)$ has property (QT) (see Lemma 2.3).

Case 2 We now assume that M has nontrivial torus decomposition. By [49, Section 6.3], we paste hyperbolic 3-manifolds with totally geodesic boundaries to M to get a 3-manifold N with empty or tori boundary. The new manifold N satisfies the following properties.

(1) M is a submanifold of N with incompressible tori boundary.

(2) The torus decomposition of M also gives the torus decomposition of N.

(3) Each piece of M with a boundary component of genus at least 2 is contained in a hyperbolic piece of N.

In particular, it follows from (2) and (3) that N is a mixed 3-manifold. The subgroup $\pi_1(M)$ sits nicely in $\pi_1(N)$. By the proof of Case 1.2 in the proof of [41, Theorem 1.3], we have that $\pi_1(M)$ is undistorted in $\pi_1(N)$ (even more than that, $\pi_1(M)$ is strongly quasiconvex in $\pi_1(N)$ (see [42]). Note that $\pi_1(N)$ has property (QT) by Proposition 8.4. Since $\pi_1(M)$ is undistorted in $\pi_1(N)$ and $\pi_1(N)$ has property (QT), it follows that $\pi_1(M)$ has property (QT). □

We now give the proof of Theorem 1.1 which gives a complete characterization of property (QT) for finitely generated 3-manifolds groups.

Proof of Theorem 1.1 Since M is a compact, orientable 3-manifold, it decomposes into irreducible, ∂-irreducible pieces $M_1, \dots, M_k$ by the sphere-disc decomposition. In particular, $\pi_1(M)$ is the free product $\pi_1(M_1) * \pi_1(M_2) * \cdots * \pi_1(M_k) * F_r$ for some free group F_r. We remark here that $\pi_1(M)$ is hyperbolic relative to the collection $\mathbb{P} = \{P_1, \dots, P_k, F_r\}$ where $P_i := \pi_1(M_i)$.

We are going to prove the necessity. Assume that $\pi_1(M)$ has property (QT). Since $\pi_1(M_i)$ is undistorted in $\pi_1(M)$, it follows that $\pi_1(M_i)$ has property (QT) (see Lemma 2.3). By Proposition 8.1, M_i does not support Sol and Nil geometry.

Now, we are going to prove sufficiency. Assume that there is no piece M_i that supports either Sol or Nil geometry. We would like to show that $\pi_1(M)$ has property (QT). In this case, we observe that each peripheral subgroup $P \in \mathbb{P}$ has property (QT). Indeed, a free group $P = F_r$ of course has property (QT), so let us now assume that $P = \pi_1(M_i)$ for some $1 \le i \le k$. If M_i has a boundary component of genus at least 2 then property (QT) of $\pi_1(M_i)$ follows from Proposition 8.6. Otherwise, M_i has empty or tori boundary. Then the property (QT) of $\pi_1(M_i)$ follows from Proposition 8.1 for geometric manifolds, Proposition 8.3 for graph manifolds, and Proposition 8.4 for mixed graph manifolds.

We are going to show that $G = \pi_1(M)$ satisfies all conditions in Theorem 1.5. The proof is similar to the proof of Proposition 8.4 with minor changes.

Claim 1 *$\pi_1(M)$ induces the full profinite topology on each $P_i \in \mathcal{P}$.*

It is well known that the fundamental groups of all compact 3-manifolds are residually finite, thus $\pi_1(M)$ is residually finite and its finite-index subgroups are residually finite as well. Any finite-index subgroup

H of $P_i = \pi_1(M_i)$ is separable in the free product $G = P_1 * P_2 * \cdots * P_k * F_r$ by [13, Theorem 1.1]. Hence it follows from [47, Lemma 2.8] that G induces the full profinite topology on P_i.

Claim 2 *For each peripheral subgroup $P \in \mathbb{P}$, there exists a finite-index subgroup P' of P acting isometrically on a finite number of quasitrees, so that the diagonal action of P' on the finite product of these quasitrees induces quasi-isometric embeddings on orbital maps.*

Indeed, the claim obviously holds for $P = F_r$ or $P = \mathbb{Z}^2$. The claim also holds for $P = \pi_1(M_i)$ where M_i is a geometric 3-manifold. The case of graph manifolds is proved in Claim 2 of the proof of Proposition 8.4. The only case left is when M_i is a mixed 3-manifold or M_i has a boundary component with genus at least 2. It has been shown in Proposition 8.6 that if M_i has a boundary component with genus at least 2 then it is an undistorted subgroup in a mixed 3-manifold. Therefore it suffices to consider only the mixed 3-manifold case. Recall that in the proof of Proposition 8.4, we show that there exists a finite-index subgroup of $\pi_1(M_i)$ such that it is a relatively hyperbolic group, satisfying the conditions of Theorem 1.5, and thus Claim 2 is confirmed.

With Claims 1 and 2, we use the same argument as in the proof of Theorem 3.5 (see the second and third paragraph) to find a finite-index normal subgroup G' of G such that G' is hyperbolic relative to a collection of subgroups satisfying the hypotheses in Theorem 1.5, and thus G' has property (QT). Therefore, $\pi_1(M)$ has property (QT) since G' is a finite-index subgroup of $\pi_1(M)$ and G' does have property (QT). □

References

[1] **C Abbott**, **S H Balasubramanya**, **D Osin**, *Hyperbolic structures on groups*, Algebr. Geom. Topol. 19 (2019) 1747–1835 MR Zbl

[2] **I Agol**, *Tameness of hyperbolic 3-manifolds*, preprint (2004) arXiv math/0405568

[3] **I Agol**, *The virtual Haken conjecture*, Doc. Math. 18 (2013) 1045–1087 MR Zbl

[4] **J Bajpai**, *Omnipotence of surface groups*, master's thesis, McGill University (2007) Available at `https://central.bac-lac.gc.ca/.item?id=MR51068`

[5] **M Bestvina**, **K Bromberg**, **K Fujiwara**, *Constructing group actions on quasi-trees and applications to mapping class groups*, Publ. Math. Inst. Hautes Études Sci. 122 (2015) 1–64 MR Zbl

[6] **M Bestvina**, **K Bromberg**, **K Fujiwara**, *Proper actions on finite products of quasi-trees*, Ann. H. Lebesgue 4 (2021) 685–709 MR Zbl

[7] **M Bestvina**, **K Bromberg**, **K Fujiwara**, **A Sisto**, *Acylindrical actions on projection complexes*, Enseign. Math. 65 (2019) 1–32 MR Zbl

[8] **H Bigdely**, **D T Wise**, *Quasiconvexity and relatively hyperbolic groups that split*, Michigan Math. J. 62 (2013) 387–406 MR Zbl

[9] **B H Bowditch**, *Tight geodesics in the curve complex*, Invent. Math. 171 (2008) 281–300 MR Zbl

[10] **B H Bowditch**, *Relatively hyperbolic groups*, Internat. J. Algebra Comput. 22 (2012) art. id. 1250016 MR Zbl

[11] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR Zbl

[12] **D Burago**, **Y Burago**, **S Ivanov**, *A course in metric geometry*, Grad. Stud. in Math. 33, Amer. Math. Soc., Providence, RI (2001) MR Zbl

[13] **R G Burns**, *On finitely generated subgroups of free products*, J. Austral. Math. Soc. 12 (1971) 358–364 MR Zbl

[14] **J O Button**, *Groups acting purely loxodromically on products of hyperbolic graphs*, preprint (2020) arXiv 2009.10575

[15] **D Calegari**, **D Gabai**, *Shrinkwrapping and the taming of hyperbolic* 3*-manifolds*, J. Amer. Math. Soc. 19 (2006) 385–446 MR Zbl

[16] **R D Canary**, *A covering theorem for hyperbolic* 3*-manifolds and its applications*, Topology 35 (1996) 751–778 MR Zbl

[17] **P-E Caprace**, **Y Cornulier**, **N Monod**, **R Tessera**, *Amenable hyperbolic groups*, J. Eur. Math. Soc. 17 (2015) 2903–2947 MR Zbl

[18] **G R Conner**, *Discreteness properties of translation numbers in solvable groups*, J. Group Theory 3 (2000) 77–94 MR Zbl

[19] **Y de Cornulier**, *Dimension of asymptotic cones of Lie groups*, J. Topol. 1 (2008) 342–361 MR Zbl

[20] **C B Croke**, **B Kleiner**, *Spaces with nonpositive curvature and their ideal boundaries*, Topology 39 (2000) 549–556 MR Zbl

[21] **C B Croke**, **B Kleiner**, *The geodesic flow of a nonpositively curved graph manifold*, Geom. Funct. Anal. 12 (2002) 479–545 MR Zbl

[22] **F Dahmani**, *Combination of convergence groups*, Geom. Topol. 7 (2003) 933–963 MR Zbl

[23] **A Dranishnikov**, **T Januszkiewicz**, *Every Coxeter group acts amenably on a compact space*, Topology Proc. 24 (1999) 135–141 MR Zbl

[24] **C Druţu**, **M Kapovich**, *Geometric group theory*, Amer. Math. Soc. Colloq. Publ. 63, Amer. Math. Soc., Providence, RI (2018) MR Zbl

[25] **C Druţu**, **M Sapir**, *Tree-graded spaces and asymptotic cones of groups*, Topology 44 (2005) 959–1058 MR Zbl

[26] **T Foertsch**, **A Lytchak**, *The de Rham decomposition theorem for metric spaces*, Geom. Funct. Anal. 18 (2008) 120–143 MR Zbl

[27] **V Gerasimov**, **L Potyagailo**, *Quasiconvexity in relatively hyperbolic groups*, J. Reine Angew. Math. 710 (2016) 95–135 MR Zbl

[28] **M F Hagen**, **P Przytycki**, *Cocompactly cubulated graph manifolds*, Israel J. Math. 207 (2015) 377–394 MR Zbl

[29] **M Hagen**, **J Russell**, **A Sisto**, **D Spriano**, *Equivariant hierarchically hyperbolic structures for* 3*-manifold groups via quasimorphisms*, Ann. Inst. Fourier (Grenoble) (online publication July 2024)

[30] **F Haglund**, **D T Wise**, *Special cube complexes*, Geom. Funct. Anal. 17 (2008) 1551–1620 MR Zbl

[31] **E Hamilton**, *Abelian subgroup separability of Haken* 3*-manifolds and closed hyperbolic n-orbifolds*, Proc. Lond. Math. Soc. 83 (2001) 626–646 MR Zbl

[32] **M Hull**, **D Osin**, *Induced quasicocycles on groups with hyperbolically embedded subgroups*, Algebr. Geom. Topol. 13 (2013) 2635–2665 MR Zbl

[33] **M Kapovich**, **B Leeb**, *Actions of discrete groups on nonpositively curved spaces*, Math. Ann. 306 (1996) 341–352 MR Zbl

[34] **M Kapovich**, **B Leeb**, 3*-manifold groups and nonpositive curvature*, Geom. Funct. Anal. 8 (1998) 841–852 MR Zbl

[35] **B Leeb**, 3*-manifolds with(out) metrics of nonpositive curvature*, Invent. Math. 122 (1995) 277–289 MR Zbl

[36] **Y Liu**, *A characterization of virtually embedded subsurfaces in* 3*-manifolds*, Trans. Amer. Math. Soc. 369 (2017) 1237–1264 MR Zbl

[37] **J M Mackay**, **A Sisto**, *Embedding relatively hyperbolic groups in products of trees*, Algebr. Geom. Topol. 13 (2013) 2261–2282 MR Zbl

[38] **J F Manning**, *Geometry of pseudocharacters*, Geom. Topol. 9 (2005) 1147–1185 MR Zbl

[39] **J F Manning**, *Quasi-actions on trees and property (QFA)*, J. Lond. Math. Soc. 73 (2006) 84–108 MR Zbl

[40] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves, II: Hierarchical structure*, Geom. Funct. Anal. 10 (2000) 902–974 MR Zbl

[41] **H T Nguyen**, **H Sun**, *Subgroup distortion of* 3*-manifold groups*, Trans. Amer. Math. Soc. 373 (2020) 6683–6711 MR Zbl

[42] **H T Nguyen**, **H C Tran**, **W Yang**, *Quasiconvexity in* 3*-manifold groups*, Math. Ann. 381 (2021) 405–437 MR Zbl

[43] **H T Nguyen**, **W Yang**, *Croke–Kleiner admissible groups: property (QT) and quasiconvexity*, Michigan Math. J. 73 (2023) 971–1019 MR Zbl

[44] **D Osin**, *Acylindrically hyperbolic groups*, Trans. Amer. Math. Soc. 368 (2016) 851–888 MR Zbl

[45] **G Paulik**, *Gluing spaces and analysis*, PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn (2004) Available at `https://hdl.handle.net/20.500.11811/2126`

[46] **P Przytycki**, **D T Wise**, *Mixed* 3*-manifolds are virtually special*, J. Amer. Math. Soc. 31 (2018) 319–347 MR Zbl

[47] **A W Reid**, *Profinite rigidity*, from "Proceedings of the International Congress of Mathematicians, II" (B Sirakov, P N de Souza, M Viana, editors), World Sci., Hackensack, NJ (2018) 1193–1216 MR Zbl

[48] **A Sisto**, *Projections and relative hyperbolicity*, Enseign. Math. 59 (2013) 165–181 MR Zbl

[49] **H Sun**, *A characterization on separable subgroups of* 3*-manifold groups*, J. Topol. 13 (2020) 187–236 MR Zbl

[50] **H Sun**, *All finitely generated* 3*-manifold groups are Grothendieck rigid*, Groups Geom. Dyn. 17 (2023) 385–402 MR Zbl

[51] **J Tidmore**, *Cocompact cubulations of mixed* 3*-manifolds*, Groups Geom. Dyn. 12 (2018) 1429–1460 MR Zbl

[52] **D T Wise**, *Subgroup separability of graphs of free groups with cyclic edge groups*, Q. J. Math. 51 (2000) 107–129 MR Zbl

[53] **D T Wise**, *The structure of groups with a quasiconvex hierarchy*, Ann. of Math. Stud. 209, Princeton Univ. Press (2021) MR Zbl

[54] **W-y Yang**, *Statistically convex-cocompact actions of groups with contracting elements*, Int. Math. Res. Not. 2019 (2019) 7259–7323 MR Zbl

School of Mathematics, Hunan University
Changsha, China

Department of Mathematics, FPT University
Da Nang, Vietnam

Beijing International Center for Mathematical Research, Peking University
Beijing, China

hansz@hnu.edu.cn, nthoang.math@gmail.com, wyang@math.pku.edu.cn

Received: 10 May 2022 Revised: 16 August 2023

Algebraic & Geometric Topology **25**:1 (2025) 161–205
DOI: 10.2140/agt.2025.25.161
Published: 5 March 2025

On positive braids, monodromy groups and framings

LIVIO FERRETTI

We associate to every positive braid a group, generalizing the geometric monodromy group of an isolated plane curve singularity. If the closure of the braid is a knot, we identify the corresponding group with a framed mapping class group. In particular, this gives a well defined knot invariant. As an application, we obtain that the geometric monodromy group of an irreducible singularity is determined by the genus and the Arf invariant of the associated knot.

57K10; 32S55, 57K20

1 Introduction

Singularity theory is a genuine source of examples and inspiration for knot theory. Since the topological type of an isolated plane curve singularity is determined by an associated link, it is possible to understand properties of the singularity from a knot theoretical viewpoint, and knot theory has been successfully applied to solve algebraic questions. In another direction, links of singularities form an interesting class of links, with special properties and invariants that follow from the whole machinery of singularity theory. It is often unclear which of those properties are inherently algebraic and which ones could be generalized to wider classes of knots and links. Among other invariants, the fundamental group of the discriminant complement and the geometric monodromy group have drawn much attention but have proved to be hard to investigate.

In [6], Baader and Lönne associate to any positive braid an abstract group defined by generators and relations, which they call the secondary braid group. Their motivation comes from the similarities between the combinatorial structure of positive braids and that of isolated plane curve singularities. In particular, they prove that for braids of type *ADE* and for braids of minimal braid index whose closure is a torus link $T_{p,q}$, the secondary braid group is isomorphic to the fundamental group of the discriminant complement of the corresponding singularities (simple singularities in the former case, Brieskorn–Pham singularities $f(x,y) = x^p + y^q$ in the latter; see [21] by Lönne). However, because of difficulties in dealing with conjugation in the positive braid monoid, they can prove that the secondary braid group is a well defined link invariant only for positive braids whose closure contains a positive half twist.

Inspired by their construction and in analogy with the definition of the geometric monodromy group of a singularity, in this article we associate to any positive braid β a group $MG(\beta)$, which we call the monodromy group of the positive braid, defined as a subgroup of the mapping class group of the unique genus minimizing Seifert surface of the closure $\hat{\beta}$, generated by the Dehn twists around some natural family of curves. The monodromy group of a positive braid is a quotient of Baader's and Lönne's secondary braid group which contains the monodromy diffeomorphism of the positive braid. Moreover, it is a generalization of the geometric monodromy group of an isolated plane curve singularity to the setting of positive braids.

Theorem 1.1 *Let $f : \mathbb{C}^2 \to \mathbb{C}$ define an isolated plane curve singularity and $L(f)$ be the link of f. Then there exists a positive braid β representing $L(f)$ such that the geometric monodromy group of f is equal to $MG(\beta)$.*

In [24], Portilla Cuadrado and Salter proved that the geometric monodromy group of any singularity of genus at least 5 and not of type A_n and D_n is a framed mapping class group, ie the stabilizer of some canonical framing on the Milnor fibre associated to the singularity and, among other things, they use this result for deducing the noninjectivity of the geometric monodromy representation. Following their approach, the main result of this paper is an identification of the monodromy group of a positive braid β whose closure is a knot with a framed mapping class group on the genus minimizing surface Σ_β.

Theorem 1.2 *Let β be a prime positive braid not of type A_n and whose closure is a knot. Then, for all but finitely many such braids, there exists a framing ϕ_β on Σ_β such that the monodromy group $MG(\beta)$ is equal to the framed mapping class group $\mathrm{Mod}(\Sigma_\beta, \phi_\beta)$.*

For the definition of a positive braid of type A_n, see Section 2. It is important to mention that the infinite family of braids of type A_n that we exclude from Theorem 1.2 is in fact one of the only cases where the monodromy group was already explicitly known: it is isomorphic to the Artin group of the corresponding type; see [23] by Perron and Vannier. Those groups are not isomorphic to any framed mapping class group, so their exclusion is a necessity, rather than a limitation of any sort.

Of course, as a consequence of Theorems 1.1 and 1.2, in the restricted context of singularities we immediately obtain that the geometric monodromy group of an irreducible singularity is controlled by a framing. In fact, as explained in Remark 5.7, one can see that our proof of Theorem 1.2 also applies to many links, including links of singularities not of type A_n and D_n, thus recovering the results of [24] up to finitely many exceptions. On the other hand, there are some infinite families of positive braid links for which our methods do not seem to work; see Remark 5.8. In spite of the increased combinatorial difficulty, working in the more general setting of positive braids has some advantages, as we will now explain.

Since the topological type of a singularity is completely determined by its link, a priori every topological invariant of a singularity should be somehow readable from the link. For instance, the Milnor number corresponds to the minimal first Betti number, while the multiplicity corresponds to the braid index [28]. However, this translation is often far from straightforward. Now, it turns out that framed mapping class groups are determined by the value of the framing on the boundary components of the surface and a certain Arf invariant associated to the framing. In the case of a surface Σ with connected boundary, the value of the framing on the boundary is always equal to the Euler characteristic of Σ, so that the framed mapping class group is determined simply by the genus of Σ and the Arf invariant of the framing. Working with positive braids, we are able to identify the Arf invariant of the framing with the classical Arf invariant of the boundary knot. We thus obtain the following corollaries, expressing the geometric monodromy group of an irreducible singularity in terms of well known invariants of its knot.

Corollary 1.3 *Let β be a prime positive braid not of type A_n and whose closure is a knot K. Up to finitely many exceptions, the monodromy group of β is an invariant of K, determined by its genus and Arf invariant.*

Corollary 1.4 *Let f define an irreducible isolated plane curve singularity that is not of type A_n and $K(f)$ be the knot of the singularity. For all but finitely many such singularities, the geometric monodromy group of f is determined by the genus and the Arf invariant of $K(f)$.*

It is important to point out that the monodromy group of a positive braid is proved to be an invariant of the braid closure only if the latter is connected; for braids whose closure is disconnected, the strongest invariance result is Corollary 2.7.

From a purely knot theoretical viewpoint, Theorem 1.2 might seem disappointing. It implies that, if the closure of a positive braid is a knot (up to finitely many exceptions), its monodromy group is an invariant of the knot, but a rather useless one: it is hard to compute, but determined by two classical and much easier invariants, a natural number and a mod 2 class. Its interest lies in negative results such as Corollary 1.4. The geometric monodromy group, which was typically considered a rich yet hard to investigate invariant of a plane curve singularity, turns out, in the case of irreducible singularities, to be determined by two simple knot invariants, and the question whether two irreducible singularities have the same geometric monodromy group can be answered by a direct and easy computation, using existing formulas for the Arf invariant of a knot. Of course, for each fixed genus there are many different irreducible singularities, so there will be different singularities with the same geometric monodromy group. We believe that for big enough genus both values of the Arf invariant are realized, so that there would be exactly two geometric monodromy groups.

The study of the monodromy group of a positive braid naturally has its place in the context of finitely generated subgroups of the mapping class group, and in particular subgroups generated by Dehn twists around a family of curves with prescribed intersection pattern. Those subgroups are interesting by themselves from a mapping class group theoretical viewpoint, but also appear naturally in different contexts, such as singularity theory or in the study of Lefschetz fibrations. The question of what groups can arise in this way is completely solved in the case of two Dehn twists; see for example Chapter 3 of [11], by Farb and Margalit, but is in general widely open. In [23] Perron and Vannier, interested in the geometric monodromy group of singularities, proved that if the intersection pattern of the curves is a Dynkin diagram of type A_n or D_n, the group generated by the Dehn twists is isomorphic to the Artin group of corresponding type, and conjectured this to be true for general graphs. This was later disproved by Labruère [19] and Wajnryb [27], whose results show that the only Artin groups whose Dynkin diagram is a tree and that geometrically embed in the mapping class group are precisely the ones of type A_n and D_n. Notice that, contrary to what Wajnryb claimed, the Artin groups of type $\tilde{A}_n$, ie whose Dynkin diagram is a cycle, do geometrically embed in the mapping class group, as recently proved by Ryffel in [26]. The theory of framed mapping class groups seems to suggest that, at least if the intersection pattern is in some sense rich enough, those finitely generated subgroups are controlled by a framing on the surface. Theorem 1.2 is an example of such a result.

As a final remark, although in this paper we concentrate only on positive braids, they are not the only natural class of links generalizing links of singularities to which one could try to associate a monodromy group. A'Campo's divide links form another such interesting family; see Section 3. More generally, it is known that the Milnor fibre of an isolated plane curve singularity can be constructed by a sequence of positive Hopf plumbings such that the core curves of the Hopf bands coincide with a distinguished basis of vanishing cycles of the singularity, the Dehn twist around which generate the geometric monodromy group; see [17] by Ishikawa. As explained in Remark 2.3, this is also the case for the monodromy group of a positive braid. Going one step further, for a general sequence of positive Hopf plumbings, one could define a monodromy group as the group generated by all the Dehn twists around the core curves of the Hopf bands. We expect that, at least for knots, results similar to Theorem 1.2 should hold in this more general setting. This is not difficult to see for Hopf plumbings with intersection pattern a tree and whose boundary is a knot of sufficiently big genus.

Structure of the paper In Section 2 we define the monodromy group of a positive braid and prove some basic invariance properties. In Section 3 we recall some basics of singularity theory and, using A'Campo's theory of divides, we prove Theorem 1.1. In Section 4 we discuss the general theory of framed mapping class groups and construct the framing appearing in Theorem 1.2. Finally, Section 5 is the technical part of the paper, in which we prove Theorem 1.2. This basically consists of a lengthy case distinction that allows us to apply general results about framed mapping class groups.

Acknowledgements I wish to thank Sebastian Baader for suggesting the topic and guiding me through this project. I am also very grateful to Livio Liechti for the several interesting discussions, and in particular for pointing out the connection to framed mapping class groups. Finally, thanks to Nick Salter, Pablo

Portilla Cuadrado and Michael Lönne for their interesting comments, and to the anonymous referee for the useful suggestions that greatly improved the exposition.

2 The monodromy group of a positive braid

Let B_N^+ be the monoid of positive braids on N strands and $\beta \in B_N^+$. We will usually represent such a braid with a *brick diagram*, a plane graph with N vertical lines connected by horizontal segments corresponding to the crossings. Since all the crossings are positive, one can reconstruct the braid from the brick diagram. It is well known that, if β is nonsplit, its closure $\hat{\beta}$ is a fibred link, whose fibre surface can be constructed by taking a disk for each strand of β and, for each generator σ_i in β, gluing a half-twisted band between the i^{th} and $(i+1)^{\text{th}}$ disks. The brick diagram of β naturally embeds in this surface as a retract. Let us denote this fibre surface by Σ_β, and let g be its genus and r the number of boundary components. On Σ_β there is a standard family of $2g+r-1$ curves γ_i, oriented counterclockwise, which are in one-to-one correspondence with the *bricks*, ie the innermost rectangles, of the brick diagram of β and form a basis of the first homology of Σ_β. See Figure 1 for an example of Σ_β with the corresponding curves for $\beta = \sigma_3\sigma_1\sigma_2\sigma_1^2\sigma_3\sigma_2$. The intersection pattern of those standard curves can be read off directly from the brick diagram, in the so called *linking graph*:

Definition 2.1 Let β be a positive braid word. Its *linking graph* is a graph whose vertices are the bricks of the brick diagram of β; two vertices are connected by an edge if and only if the corresponding bricks are arranged as the two bricks of the braids σ_i^3, $\sigma_i\sigma_{i+1}\sigma_i\sigma_{i+1}$ or $\sigma_{i+1}\sigma_i\sigma_{i+1}\sigma_i$.

Notice that two vertices of the linking graph are connected with an edge if and only if the corresponding curves intersect each other. Linking graphs of positive braids were studied in great detail in [5]. Here it is worth mentioning that since positive braid links are visually prime by [10], a positive braid link is prime if and only if the linking graph is connected. In what follows, we will say that a positive braid link is of

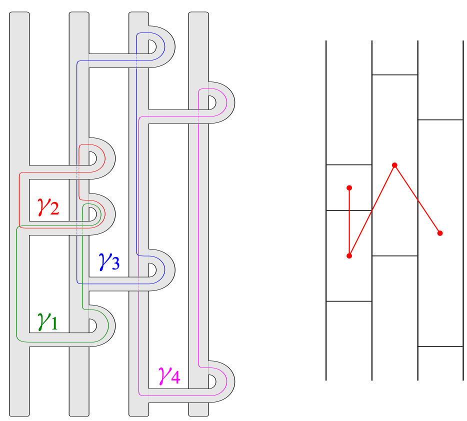


Figure 1: The fibre surface of $\sigma_3\sigma_1\sigma_2\sigma_1^2\sigma_3\sigma_2$, its brick diagram and the corresponding linking graph.

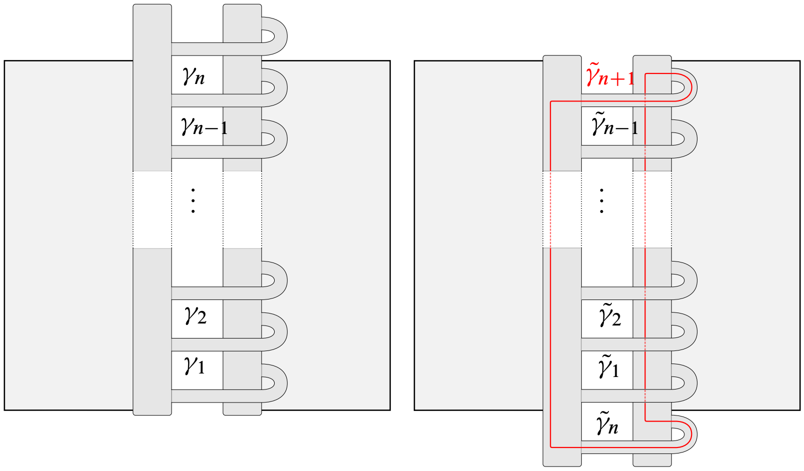


Figure 2: The isotopy between $\Sigma_{\beta\sigma_i}$ and $\Sigma_{\sigma_i\beta}$.

type A_n (resp. D_n) if it isotopic to the closure of the braid σ_1^{n+1} (resp. $\sigma_1^{n-2}\sigma_2\sigma_1^2\sigma_2$). Those braids have as linking graph the simply laced Dynkin diagram of type A_n or D_n.

Definition 2.2 Let β be a positive braid. The *monodromy group* $MG(\beta)$ is the subgroup of the mapping class group of Σ_β generated by all the Dehn twists around the curves γ_i, $i = 1, \dots, 2g+r-1$, ie

$$MG(\beta) = \langle T_{\gamma_1}, \dots, T_{\gamma_{2g+r-1}} \rangle \leq \mathrm{Mod}(\Sigma_\beta).$$

Remark 2.3 As we just said, if a positive braid β is nonsplit, then its closure is fibred, and Σ_β is the fibre surface. In fact, this surface can be constructed by a sequence of plumbings of positive Hopf bands, and the curves γ_i are precisely the core curves of those Hopf bands. The monodromy group of β therefore somehow reflects this plumbing structure.

Example 2.4 As already mentioned, it follows from [23] that if $\beta = \sigma_1^{n+1}$ then $MG(\beta)$ is isomorphic to the Artin group of type A_n. Similarly, for $\beta = \sigma_1^{n-2}\sigma_2\sigma_1^2\sigma_2$, $MG(\beta)$ is isomorphic to the Artin group of type D_n.

From the definition, it is clear that $MG(\beta)$ is invariant under far-commutativity (ie $\sigma_i\sigma_j = \sigma_j\sigma_i$ for $|i-j| \geq 2$) and positive Markov move.

Proposition 2.5 (elementary conjugation invariance) *Let β be a positive braid on N strands. Then for all $1 \leq i \leq N-1$, $MG(\beta\sigma_i) \simeq MG(\sigma_i\beta)$.*

Proof Consider the fibre surfaces of $\beta\sigma_i$ and $\sigma_i\beta$. Those surfaces are isotopic, by sliding the topmost band between the i^{th} and $(i+1)^{\text{th}}$ disks along the back of the disks and bringing it in the lowermost position. Note that this isotopy restricts to the identity outside of the i^{th} column. The surfaces $\Sigma_{\beta\sigma_i}$ and $\Sigma_{\sigma_i\beta}$ can hence be schematically represented as in Figure 2, where we drew the i^{th} column and the light grey boxes on the two sides represent the remaining parts of the surface.

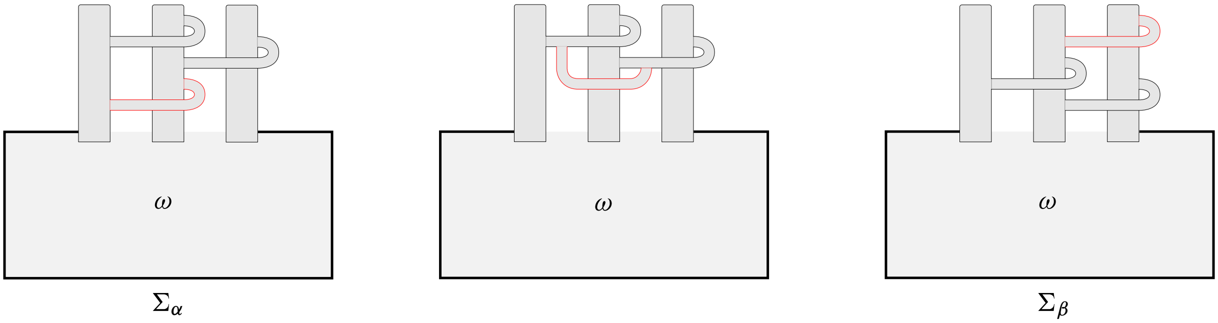


Figure 3: The isotopy between Σ_α and Σ_β.

Let us number the standard curves of the i^{th} column as in Figure 2. The isotopy will send each γ_i, $i = 1, \dots, n-1$ to the corresponding $\tilde{\gamma}_i$, $i = 1, \dots, n-1$ and transform γ_n into the red curve $\tilde{\gamma}_{n+1}$. All that we have to prove is then that we can generate the Dehn twists around the curves $\tilde{\gamma}_1, \dots, \tilde{\gamma}_{n-1}, \tilde{\gamma}_{n+1}$ using $\tilde{\gamma}_1, \dots, \tilde{\gamma}_{n-1}, \tilde{\gamma}_n$, and vice-versa. But we note that

$$\tilde{\gamma}_{n+1} = T_{\tilde{\gamma}_{n-1}} \cdots T_{\tilde{\gamma}_2} T_{\tilde{\gamma}_1}(\tilde{\gamma}_n),$$

so that for $h = T_{\tilde{\gamma}_{n-1}} \cdots T_{\tilde{\gamma}_2} T_{\tilde{\gamma}_1}$ we have

$$T_{\tilde{\gamma}_{n+1}} = h T_{\tilde{\gamma}_n} h^{-1}$$

and the result is proved. □

Proposition 2.6 (braid relation invariance) *Let α and β be two positive braids related by a braid relation, then $MG(\alpha) \simeq MG(\beta)$.*

Proof Up to elementary conjugation, we can suppose that $\alpha = \omega\sigma_i\sigma_{i+1}\sigma_i$ and $\beta = \omega\sigma_{i+1}\sigma_i\sigma_{i+1}$, where ω is a positive braid on N strands and $1 \le i \le N-2$. At the level of surfaces Σ_α and Σ_β the braid relation can be realized by an isotopy as in Figure 3. It is clear that all the standard curves γ_i are fixed by this isotopy but the ones (at most two) passing through the slid band.

- There is a generator σ_i in ω: in this case, there are two curves on Σ_α which are modified by the isotopy. Let us call them γ_1 and γ_2, as in Figure 4. We see that, after the isotopy, γ_1 is transformed into the corresponding $\tilde{\gamma}_1$, while γ_2 becomes $T_{\tilde{\gamma}_1}^{-1}(\tilde{\gamma}_2)$. All the other standard curves are fixed. Therefore, we get that $MG(\alpha) \simeq MG(\beta)$.
- There is no σ_i in ω: in this case, the only curve modified by the isotopy is γ_1, which as before is transformed into $\tilde{\gamma}_1$. Again, we directly have that $MG(\alpha) \simeq MG(\beta)$. □

The following corollary now follows directly by an observation of Orevkov about Garside's solution of the conjugacy problem in the braid group, saying that, in the presence of a positive half twist, two conjugate positive braids can be related by a sequence of braid relations and elementary conjugations; see [6, Section 6].

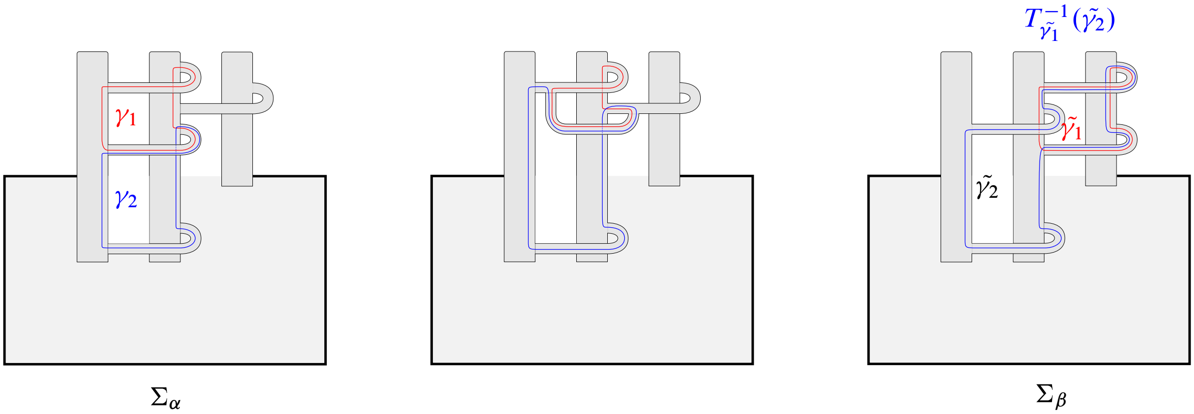


Figure 4: First case of braid relation invariance.

Corollary 2.7 *Let α and β be positive braids such that the closures are braid isotopic and contain a positive half twist. Then $MG(\alpha) \simeq MG(\beta)$.*

3 Divides and monodromy of singularities

The monodromy group of a positive braid is a generalization of the geometric monodromy group of an isolated plane curve singularity. In this section, we will make this statement more precise.

Let $f : \mathbb{C}^2 \to \mathbb{C}$ define an isolated plane curve singularity. For a suitably small radius $r > 0$, the sphere $\partial(B^4_r) \subset \mathbb{C}^2$ intersects the singular curve $C = f^{-1}(0)$ transversally, so that the intersection $L(f) = C \cap \partial(B^4_r)$ is a link in $S^3 = \partial(B^4_r)$, called the link of the singularity. It is well known that the isotopy type of $L(f)$ completely determines the topological type of the singularity. Moreover, in [22] Milnor proved that the map $f/|f| : S^3 \setminus L(f) \to S^1$ is a fibration. Singularity links are therefore fibred links, with fibre a surface $\Sigma(f)$ called the *Milnor fibre*. It turns out that all the singularity links are iterated torus links, and in particular positive braid links. The fibration induces a monodromy diffeomorphism of the fibre, which is only defined up to isotopy and therefore defines a mapping class in $\mathrm{Mod}(\Sigma(f))$, called the *geometric monodromy* of the singularity. The geometric monodromy is an important invariant, which determines the topology of the singularity and has been intensively studied in the context of singularity theory.

By the study of the deformations of the singularity, the geometric monodromy can be "promoted" to the so called *geometric monodromy group* of the singularity. It is a subgroup of $\mathrm{Mod}(\Sigma(f))$ generated by the Dehn twists around some specific curves on the Milnor fibre, called *vanishing cycles*. The geometric monodromy can be expressed as a product of those generators and is therefore an element of the geometric monodromy group. We will not discuss the original definition of the geometric monodromy group of a singularity since, although classic, it would require quite some background knowledge in singularity theory and will not be useful for our purposes. However, there exists an easy combinatorial model for the Milnor fibre of a singularity which allows us to directly define the geometric monodromy group in terms of explicit generators. This was constructed by A'Campo using the theory of divides.

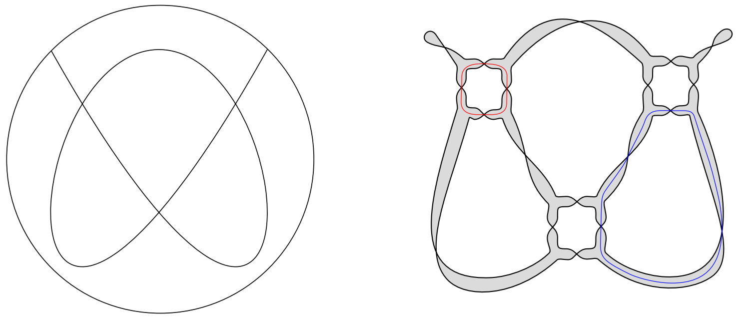


Figure 5: A divide and the associated surface with some of the vanishing cycles. The corresponding link is the torus knot $T_{3,4}$.

Definition 3.1 A *divide* $\mathcal{D}$ is a generic relative immersion of finitely many intervals in the unit disk $(D^2, \partial(D^2))$.

Here, *generic* means that the only singularities are double points and that the intervals meet the boundary $\partial(D^2)$ transversally. Examples of divides can be seen in Figures 5 and 6.

Divides were first introduced by A'Campo [1; 2] and Gusein-Zade [15; 14], who independently proved that they could be associated in a natural way to singularities and used them for studying properties of the monodromy. Later on, in [4; 3], A'Campo associated to any divide $\mathcal{D}$ a link $L(\mathcal{D})$, constructed as follows. Consider the tangent bundle of the unit disk, $TD^2 = \{(x, v) \mid x \in D^2, v \in T_x D^2\}$. The sphere S^3 can be seen as the unit sphere in TD^2,

$$S^3 = \{(x, v) \in TD^2 \mid |x|^2 + |v|^2 = 1\}.$$

Now let $\mathcal{D} \subset D^2$ be a divide, the link of $\mathcal{D}$ is defined as

$$L(\mathcal{D}) = \{(x, v) \in S^3 \mid x \in \mathcal{D}, v \in T_x\mathcal{D}\} \subset S^3.$$

This gives a link whose number of components is equal to the number of intervals in the divide. In the same papers, A'Campo proved that if the divide is connected the link is fibred and that if the divide was obtained from a singularity the associated link $L(\mathcal{D})$ is ambient isotopic to the link of the singularity. In this latter case, he also provided an easy graphical algorithm to construct a model of the Milnor fibre on which a system of vanishing cycles is visible. We say that a *face* of a divide $\mathcal{D}$ is a connected component of $D^2 \setminus \mathcal{D}$ which does not intersect the boundary of D^2. Let n be the number of intervals in $\mathcal{D}$, δ be the number of crossings and r the number of faces. The Milnor fibre will be a surface with first Betti number $\mu = \delta + r$ and n boundary components. The distinguished vanishing cycles will be given by one curve per crossing and one curve per face. The surface is constructed as follows: first, replace every crossing of $\mathcal{D}$ with a small circle, to get a trivalent graph. Now, realize every edge of this new graph by a half-twisted band. This will give a surface composed of twisted cylinders, corresponding to the crossings of $\mathcal{D}$, connected by half-twisted bands corresponding to the edges of $\mathcal{D}$. The vanishing cycle associated to a crossing will be given by the core curve of the corresponding cylinder, the vanishing cycle of a face will be given by the core curves of the bands bounding the face. An example of this construction is given in Figure 5.

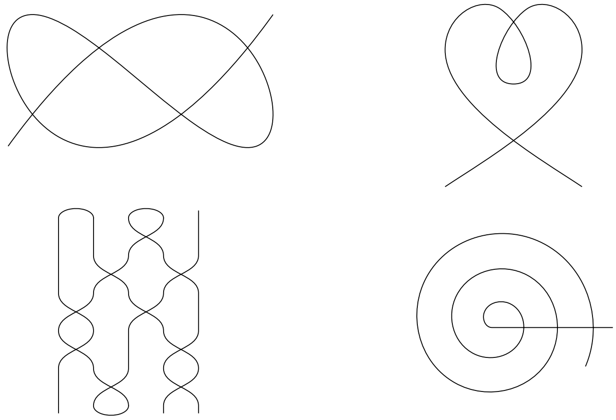


Figure 6: The divides on the left are ordered Morse, the divides on the right are not.

Remark 3.2 A'Campo's construction only leads to a combinatorial model of the Milnor fibre which is not embedded. A graphical procedure to construct a diagram of the link of a divide and the associated embedded fibre surface has been given by Hirasawa in [16].

Definition 3.3 Let f be an isolated plane curve singularity, $\mathcal{D}$ a divide associated to f and $\Sigma(f)$ the surface constructed from $\mathcal{D}$ with the previous procedure. The *geometric monodromy group* of f is the subgroup of $\mathrm{Mod}(\Sigma(f))$ generated by the Dehn twists around the vanishing cycles constructed on $\Sigma(f)$. This does not depend on the choice of the divide $\mathcal{D}$.

As we have already mentioned, links of singularities are closures of positive braids. Since fibre surfaces of fibred links are unique, the Milnor fibre of a singularity f is ambient isotopic to the fibre surface Σ_β of any positive braid β representing $L(f)$. We therefore now have two a priori distinct subgroups of $\mathrm{Mod}(\Sigma_\beta) = \mathrm{Mod}(\Sigma(f))$, the geometric monodromy group of f and the monodromy group of β. Theorem 1.1 says that those two groups coincide for at least one choice of β.

To prove Theorem 1.1, we will explicitly find an isotopy between the Milnor fibre constructed from a divide and the surface of an appropriate positive braid and identify the vanishing cycles on this braid surface. In order to do so, we need to use a divide from which the positive braid is somehow visible.

Definition 3.4 A divide $\mathcal{D} \subset D^2$ is an *ordered Morse divide* if there is a diameter of D^2 such that the orthogonal projection on this diameter is Morse when restricted to $\mathcal{D}$, all the local maxima (resp. minima) have the same critical value b (resp. a) with $b > a$ and all the crossings are mapped in the open interval (a, b).

Basically, a divide is ordered Morse (with respect to a given direction) if no local maxima or minima lie in an interior face of the divide. Examples of such divides are given in Figure 6.

Remark 3.5 In the literature, ordered Morse divides are sometimes called *scannable divides*.

Figure 7: Hirasawa's construction of the embedded fibre surface of an ordered Morse divide.

Ordered Morse divides were introduced by Couture and Perron [9], who used a generalization of those to construct a representative braid for any divide link. In particular, ordered Morse divides give positive braid links. Notice that every singularity has an associated divide which is ordered Morse (in fact, the divides originally constructed by A'Campo and Gusein-Zade are ordered Morse; see [9]). The result of Couture and Perron can be obtained geometrically: if we apply the algorithm of [16] to an ordered Morse divide, we get exactly the fibre surface of a positive braid. This was done in [13] for Lissajous divides and torus links, but the same procedure works for an arbitrary ordered Morse divide. The construction of the fibre surface is shown in Figure 7: one just has to replace the crossings and minima/maxima of the divide with the corresponding pieces of surface and glue them together following the pattern of the divide. Here we use that all the minima and maxima of the divide are in the exterior face: for general divides the fibre surface is more complicated.

Remark 3.6 The diagrams in Figure 7 are the mirror image of those obtained by Hirasawa in [16]. This is due to the different choice of orientation of S^3: Hirasawa uses the orientation induced by the trivialization $T\mathbb{R}^2 = \{(x, v) \mid x \in \mathbb{R}^2, v \in T_x\mathbb{R}^2\} \cong \mathbb{R}^2 \times \mathbb{R}^2$; we use the identification $T\mathbb{R}^2 \cong \mathbb{C}^2$, where the plane $\mathbb{R}^2$ is identified with the real part of $\mathbb{C}^2$, since this allows one to correctly identify the link of a singularity with the link of a corresponding divide.

Proof of Theorem 1.1 Let f be an isolated plane singularity and $\mathcal{D}$ an associated ordered Morse divide. Let Σ be the embedded surface constructed following [16], as explained above. It is an embedded fibre surface whose boundary is the link $L(\mathcal{D}) = L(f)$. To see that this is indeed the fibre surface of a positive braid, we just need to perform the isotopies shown in Figure 8(2a), getting a collection of disks connected by half-twisted bands, and slide all the bands to the front. Let us remark that an ordered Morse divide is

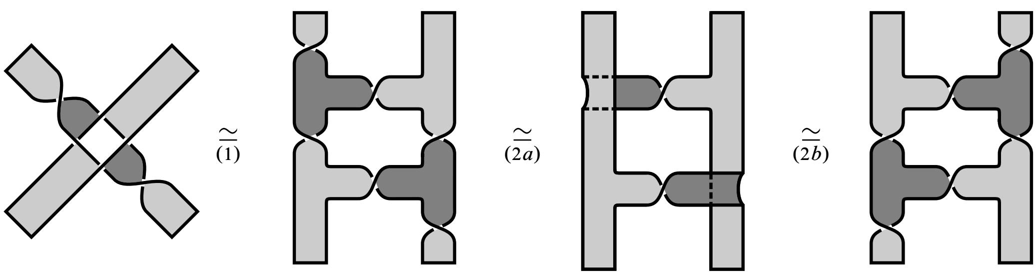


Figure 8: A sequence of isotopies.

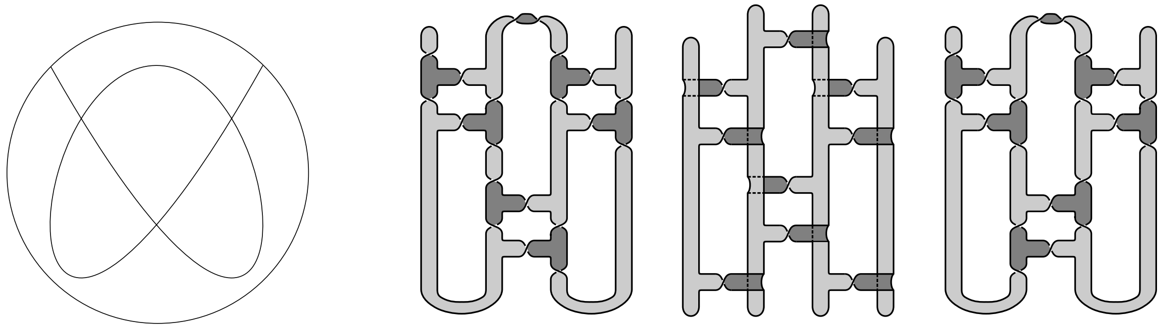


Figure 9: An example of the isotopies of Theorem 1.1.

formed of N parallel lines (where N is the number of points in the preimage of a regular value of the Morse projection) connected by the crossings and the minima/maxima. The braid obtained will have N strands, a crossing of $\mathcal{D}$ gives a pair of generators while every maximum/minimum gives one generator.

By further performing the isotopies of Figure 8(2b) around all the crossings of $\mathcal{D}$ corresponding to generators σ_i for even i, we can now directly identify Σ with an embedded version of A'Campo's model of the Milnor fibre. A system of vanishing cycles is therefore visible on the braid surface Σ. Those cycles are not exactly the same as the generators of the monodromy group of the braid, but the same arguments as in the proof of Proposition 2.5 show that the two groups are indeed the same. □

Example 3.7 In Figure 9, we see an example of the isotopies used in the previous proof. On the left, we start with a divide $\mathcal{D}$; we then construct the Seifert surface following Hirasawa's algorithm. After applying the isotopies of Figure 8(2a), we obtain the surface Σ_β of a positive braid, namely $\beta = (\sigma_1\sigma_2\sigma_3)^3$. On the right, we performed the isotopy of Figure 8(2b) around the central crossing of $\mathcal{D}$. In that way, we clearly see that the surface is composed of twisted cylinders corresponding to the crossings of $\mathcal{D}$ and connected by *half-twisted* bands, as required by A'Campo's construction (compare with Figure 5). Notice that it is not relevant that this last step is performed around all the crossings of $\mathcal{D}$ corresponding to generators σ_i for *even* i as opposed to *odd* i; what matters is that it alternates, in order the get the required half-twisting of the bands become visible.

4 Framings

We will now briefly recall the basics of the theory of framed surfaces, concentrating in particular on the action of the mapping class group on such structures, as investigated in [8; 25]. In what follows, we will adhere to the notations and conventions of [8], but we will restrict only to the case of surfaces with connected boundary. Let $\Sigma = \Sigma_{g,1}$ be a connected, compact, oriented surface of genus g with one boundary component. A framing ϕ on Σ is a trivialization of the tangent bundle $T\Sigma$. With the fixed orientation (and a choice of a Riemannian metric), a framing is determined by a nowhere-vanishing vector field ξ_ϕ on Σ. Two framings are *isotopic* if the associated vector fields are isotopic through nowhere-vanishing vector fields.

To a framing one can associate a *winding number function*, computing the holonomy of a simple closed curve. If $c\colon \mathbb{S}^1 \to \Sigma$ is a $\mathcal{C}^1$ embedding, one can define

$$\phi(c) = \int_{\mathbb{S}^1} d\angle\big(\dot{c}(t), \xi_\phi(c(t))\big) \in \mathbb{Z}.$$

This defines a map from the set of simple closed curves on Σ to $\mathbb{Z}$, which is clearly invariant under isotopy of ϕ and c. It is not hard to see that the converse also holds: the isotopy class of a framing on Σ is determined by its winding number function, and actually by the value on finitely many curves (see [8, Lemma 2.2; 25, Proposition 2.4]). Thanks to this, we will use the term "framing" indifferently to refer to the isotopy class of the vector field ξ_ϕ or to the associated winding number function ϕ.

Remark 4.1 Since we are only considering surfaces with connected boundary, it follows from the Poincaré–Hopf index theorem that for any framing ϕ on Σ, if the boundary $\partial\Sigma$ is oriented with the surface on its left, $\phi(\partial\Sigma) = \chi(\Sigma)$.

The mapping class group of Σ acts on the set of isotopy classes of framings by pullback, via $f \cdot \phi(c) = \phi(f^{-1}(c))$, for $f \in \mathrm{Mod}(\Sigma)$ and c a simple closed curve.

Definition 4.2 Let (Σ, ϕ) be a framed surface. The *framed mapping class group*

$$\mathrm{Mod}(\Sigma, \phi) = \{f \in \mathrm{Mod}(\Sigma) \mid f \cdot \phi = \phi\}$$

is the stabilizer of the isotopy class of ϕ.

Of particular interest is the action of Dehn twists.

Lemma 4.3 [8, Lemma 2.4] *Let (Σ, ϕ) be a framed surface and a, x oriented simple closed curves on Σ. Then*

$$\phi(T_a(x)) = \phi(x) + \langle x, a\rangle \phi(a),$$

where $\langle \cdot, \cdot \rangle$ denotes the algebraic intersection number.

We say that a nonseparating simple closed curve a on (Σ, ϕ) is *admissible* if $\phi(a) = 0$. As a consequence of Lemma 4.3 we have that a nonseparating simple closed curve $a \subset \Sigma$ is admissible if and only if the corresponding Dehn twist preserves ϕ. Calderon and Salter proved that, for big enough genus, the framed mapping class group is generated by those admissible twists:

Proposition 4.4 [8, Proposition 5.11] *If (Σ, ϕ) is a framed surface of genus $g \geq 5$,*

$$\mathrm{Mod}(\Sigma, \phi) = \langle T_a \mid a \text{ admissible for } \phi\rangle.$$

But more is true. The framed mapping class group is generated by finitely many admissible twists around curves with prescribed intersection pattern. Again following [8]:

Definition 4.5 Let $\mathcal{C} = \{c_1, \dots, c_k\}$ be a collection of curves on a surface Σ, pairwise in minimal position and intersecting at most once. We say that such a configuration

- *spans the surface* if Σ deformation retracts onto the union of curves in $\mathcal{C}$;
- is *arboreal* if its intersection graph is a tree, and E*-arboreal* if moreover it contains the Dynkin diagram E_6 as a subtree.

Definition 4.6 Let $\mathcal{C} = \{c_1, \dots, c_k, c_{k+1}, \dots, c_l\}$ be a collection of curves on a surface Σ and denote by S_j a regular neighbourhood of $\{c_1, \dots, c_j\}$. We say that $\mathcal{C}$ is an *h-assemblage of type E* if

- $\{c_1, \dots, c_k\}$ is an E-arboreal spanning configuration on a subsurface $S \subset \Sigma$ of genus h;
- For $j > k$, $c_j \cap S_{j-1}$ is a single arc;
- $S_l = \Sigma$.

Proposition 4.7 [8, Theorem B] *Let (Σ, ϕ) be a framed surface and $\mathcal{C} = \{c_1, \dots, c_l\}$ an h-assemblage of type E on Σ of genus $h \geq 5$. If all the curves in $\mathcal{C}$ are admissible for ϕ, then*

$$\operatorname{Mod}(\Sigma, \phi) = \langle T_c \mid c \in \mathcal{C} \rangle.$$

The orbit space of this action was studied by Randal-Williams in [25]. It is classified by the Arf invariant. More precisely, it follows from work of Johnson [18] that the function $(\phi + 1) \bmod 2$ is a quadratic refinement of the mod 2 intersection form. We can therefore define $\mathcal{A}(\phi)$ to be the Arf invariant of this quadratic form. More concretely, let us denote by $i(\cdot, \cdot)$ the geometric intersection number and take a collection of oriented simple closed curves $\{x_1, y_1, \dots, x_g, y_g\}$ such that $\langle x_i, x_j \rangle = \langle y_i, y_j \rangle = 0$ and $\langle x_i, y_j \rangle = i(x_i, y_j) = \delta_{i,j}$. We then have

$$\mathcal{A}(\phi) = \sum_{i=1}^{g} (\phi(x_i) + 1)(\phi(y_i) + 1) \mod 2.$$

This is of course independent of the choice of the curves $\{x_1, y_1, \dots, x_g, y_g\}$.

Proposition 4.8 [25, Theorem 2.9] *Let $g \geq 2$. The action of the mapping class group on the set of isotopy classes of framings on $\Sigma = \Sigma_{g,1}$ has exactly two orbits, distinguished by the Arf invariant.*

As a consequence, for a given Σ there are exactly two conjugacy classes of framed mapping class groups as subgroups of $\operatorname{Mod}(\Sigma)$.

Remark 4.9 (caveat) In this section we only stated results for surfaces with connected boundary, in terms of *absolute* framings. For general surfaces, the whole theory is still valid, but needs to be formulated for *relative* framings, ie only allowing isotopies that are trivial on the boundary. In this more general context, the framed mapping class group is the stabilizer of the *relative* isotopy class of a framing, and one needs to also take into account the action on arcs, getting so-called generalized winding number functions. The orbit space is now classified by a generalized Arf invariant together with the values of the

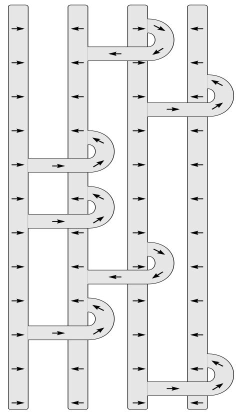


Figure 10: The framing on Σ_β for $\beta = \sigma_3\sigma_1\sigma_2\sigma_1^2\sigma_3\sigma_2$. On the vertical disks it is horizontal with alternating directions, on the twisted bands it is parallel to the core.

framing on the different boundary components. However, if the boundary is connected the absolute and relative theories are equivalent and we can use this slightly simpler formulation.

4.1 A framing for positive braids

Let β be a nonsplit positive braid and Σ_β its fibre surface. We can construct a framing ϕ_β on Σ_β as in Figure 10. An explicit and straightforward computation now shows that every standard curve γ_i on Σ_β is admissible for ϕ_β. Therefore, the monodromy group of β is contained in the framed mapping class group of ϕ_β:

$$MG(\beta) \leqslant \mathrm{Mod}(\Sigma_\beta, \phi_\beta).$$

We will prove that, at least for positive braids whose closure is a knot of big enough genus, the monodromy group is equal to this framed mapping class group. Therefore, in view of the previous discussion, we now want to compute the Arf invariant of ϕ_β.

Proposition 4.10 *Let β be a positive braid whose closure is a knot K. Then*

$$\mathcal{A}(\phi_\beta) = \mathcal{A}(K),$$

where $\mathcal{A}(K)$ is the classical Arf invariant of K.

To prove Proposition 4.10, we will need to discuss a bit more in detail the Arf invariant. Let V be a finite dimensional vector space over $\mathbb{Z}_2$ equipped with a nonsingular, symmetric bilinear pairing $\langle\cdot,\cdot\rangle\colon V \times V \to \mathbb{Z}_2$. Recall that a *quadratic refinement* of the bilinear pairing $\langle\cdot,\cdot\rangle$ is a function $q\colon V \to \mathbb{Z}_2$ such that for all $x, y \in V$

$$q(x+y) = q(x) + q(y) + \langle x, y\rangle.$$

To such a mod 2 quadratic form it is classically associated the Arf invariant $\mathcal{A}(q) \in \mathbb{Z}_2$.

In our context, we will take $V = H_1(\Sigma_\beta, \mathbb{Z}_2)$ and $\langle\cdot,\cdot\rangle$ the mod 2 intersection form. As we have already mentioned, the framing ϕ_β induces a quadratic refinement of the intersection form, whose Arf invariant is $\mathcal{A}(\phi_\beta)$. On the other hand, if the closure of β is a knot K, it is known that the Seifert form also induces such a quadratic refinement. More precisely, if $S\colon H_1(\Sigma_\beta)\times H_1(\Sigma_\beta)\to\mathbb{Z}$ denotes the Seifert form, we can define $q\colon V\to\mathbb{Z}_2$ by $q(x) = S(x,x) \mod 2$. It is a classical result that the Arf invariant of this quadratic form is indeed an invariant of K, that we denote by $\mathcal{A}(K)$.

Proof of Proposition 4.10 Let β be a positive braid whose closure is a knot K and Σ_β its fibre surface, equipped with the framing ϕ_β. The family of curves γ_i form a basis of $V = H_1(\Sigma_\beta, \mathbb{Z}_2)$. Since by construction all the γ_i are admissible for ϕ_β, for every i we have the equality

$$\phi_\beta(\gamma_i) + 1 = 1 = S(\gamma_i,\gamma_i) \mod 2.$$

Since $\{\gamma_i\}$ is a basis, it now follows from the defining equation of a quadratic refinement that for every $x\in V$

$$\phi_\beta(x) + 1 = q(x) \mod 2.$$

Therefore the two quadratic forms $(\phi_\beta + 1)$ mod 2 and q coincide, so their Arf invariants also do. $\square$

5 Proof of the main theorem

In this section we will give the proof of Theorem 1.2, stating that, up to finitely many exceptions, the monodromy group of a positive braid not of type A_n and whose closure is a knot is a framed mapping class group. In the previous section we have constructed a framing ϕ_β on the fibre surface Σ_β and seen that $MG(\beta) \leqslant \mathrm{Mod}(\Sigma_\beta, \phi_\beta)$, so we only need to deal with the opposite inclusion. This will be done by applying Proposition 4.7. As a first step, we have to find appropriate subsurfaces supporting an E-arboreal spanning configuration. For this, we will separately consider the case of braids on 3-strands (Proposition 5.1), on at least 11 strands (Proposition 5.2) and finally with an intermediate number of strands (Proposition 5.6).

Proposition 5.1 *Let β be a prime positive 3-braid of genus $g \geq 5$ which is not of type A_n or D_n. Then, excepting finitely many braids, up to positive braid isotopy its linking graph contains an induced subtree which is an E-arboreal spanning configuration on a subsurface of genus $g \geq 5$.*

Proof Let β be a positive 3-braid which is not of type A_n. Up to elementary conjugation and braid relation we can assume that $\beta = \sigma_1^{a_1}\sigma_2^{b_1}\cdots\sigma_1^{a_m}\sigma_2^{b_m}$, with $a_i \geq 2$ and $b_i \geq 1$ for all $i\in\{1,\dots,m\}$. First of all, notice that if we can find a suitable subtree for a braid $\sigma_1^{a_1}\sigma_2^{b_1}\cdots\sigma_1^{a_m}\sigma_2^{b_m}$, the result will also hold for any braid $\sigma_1^{a'_1}\sigma_2^{b'_1}\cdots\sigma_1^{a'_m}\sigma_2^{b'_m}$ for $a'_i \geq a_i$ and $b'_i \geq b_i$. We will now prove the result by case distinction over m.

***m* ≥ 5** Every braid with $m\geq 5$ has genus $g\geq 5$ so it is clearly enough to prove the result for $m = 5$. If one of the b_i is at least 2, we can assume that $\beta = \sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2^2$. In the left of Figure 11 we

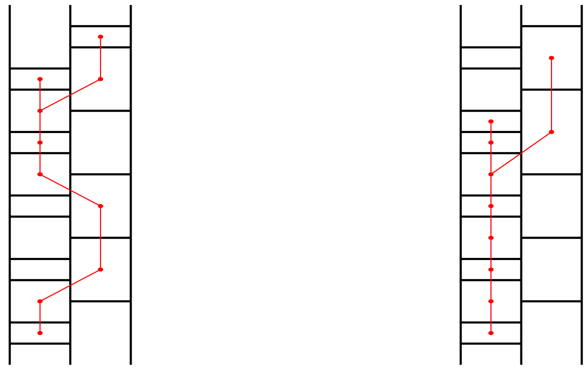


Figure 11: Subtrees of $\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2^2$ and $\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^3\sigma_2\sigma_1^2\sigma_2$.

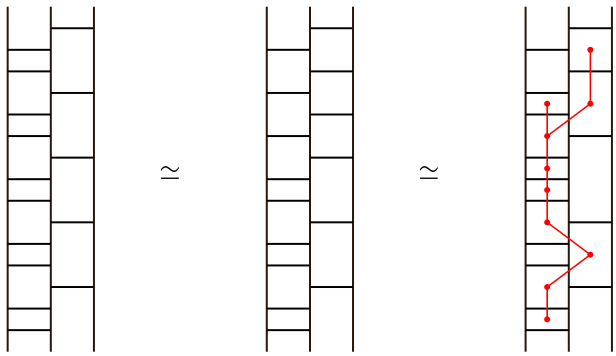


Figure 12: The braid $\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2$.

now see an induced subtree of the linking graph with the required properties. Similarly if one of the a_i is at least 3 we can assume that $\beta = \sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^3\sigma_2\sigma_1^2\sigma_2$, and we find the induced subtree of the right of Figure 11.

We are now only left with the braid $\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2$. Here we do not directly find an appropriate subtree, but Figure 12 shows a sequence of braid relations that makes it visible.

$\boldsymbol{m=4}$ We will treat several cases. Let us first assume that there is an i such that $b_i \geq 2$. If there are $i \neq j$ such that $b_i, b_j \geq 2$, then up to cyclic ordering we only have to deal with the two cases depicted in the left of Figure 13, where we see the sought subtrees. Similarly, if there is only one b_i greater than 2 but there is one a_j bigger than 3 we will find one of the trees in the right of Figure 13. Finally, if all the a_j are equal to 2 and there is only one b_i greater than 2, it is enough to consider the braid $\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2^2$, for which we can find the subtree after applying some braid relations as in Figure 14.

We are now left with $b_i = 1$ for all i. Notice that in that case there need to be at least one $a_i \geq 3$, otherwise the braid has genus less than 5. If there are two nonconsecutive a_i and a_j greater than 3, it is enough to consider the braid $\sigma_1^3\sigma_2\sigma_1^2\sigma_2\sigma_1^3\sigma_2\sigma_1^2\sigma_2$, for which we find an appropriate subtree in the left of Figure 15.

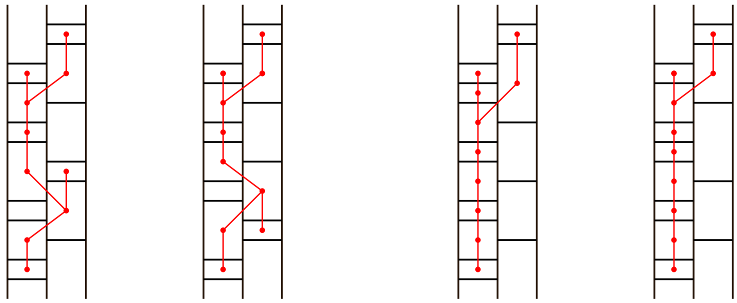


Figure 13: The cases when $m = 4$ and $b_i, b_j \geq 2$ (left) or $b_i \geq 2$ and $a_j \geq 3$ (right).

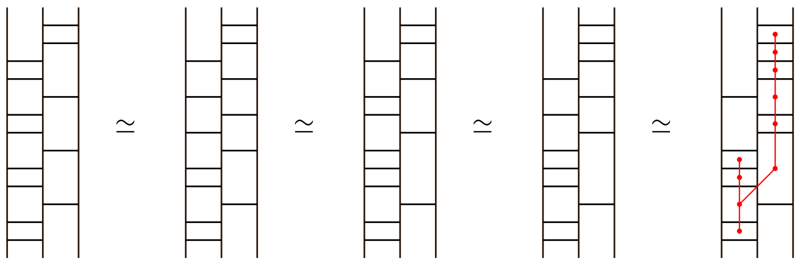


Figure 14: The braid $\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2^2$.

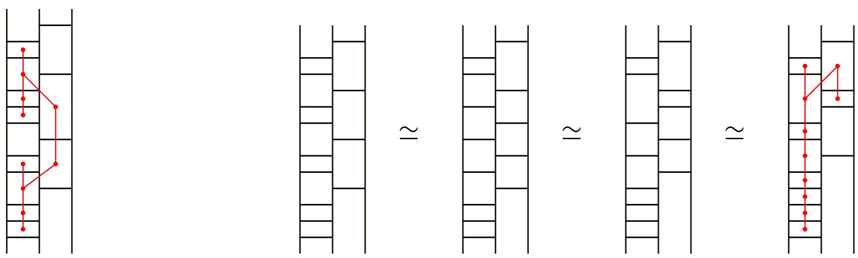


Figure 15: The cases $m = 4$ and $b_i = 1$ for all i.

If not, up to cyclic ordering there must be two consecutive $a_i = a_{i+1} = 2$, in which case we can apply a sequence of braid relations as we did in the right of Figure 15 and find our subtree.

$m = 3$ This will be the lengthier case, since there are many low genus braids that require special treatment. Let $\beta = \sigma_1^{a_1}\sigma_2^{b_1}\sigma_1^{a_2}\sigma_2^{b_2}\sigma_1^{a_3}\sigma_2^{b_3}$ be a braid of genus $g \geq 5$, then a simple argument implies that $\sum a_i + \sum b_i \geq 12$. If $\sum b_i \geq 8$, it is enough to consider the braids in Figure 16. Similarly, when $\sum a_i \geq 11$ it is enough to consider the case when all the b_i are equal to 1, and up to elementary conjugation we can assume that $a_3 \geq 3$. In this case, by taking all the vertices in the left column and only the topmost of the right column we will always end up finding a tripod tree $T(1, k, 9-k)$ for $k \geq 2$, which all correspond to subsurfaces of genus 5; see Figure 17 for some examples.

We are now left with the low genus cases.

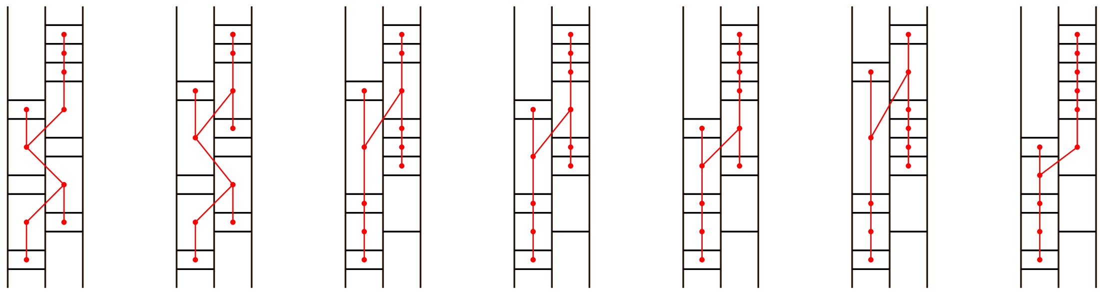


Figure 16: When $m = 3$ and $\sum b_i = 8$.

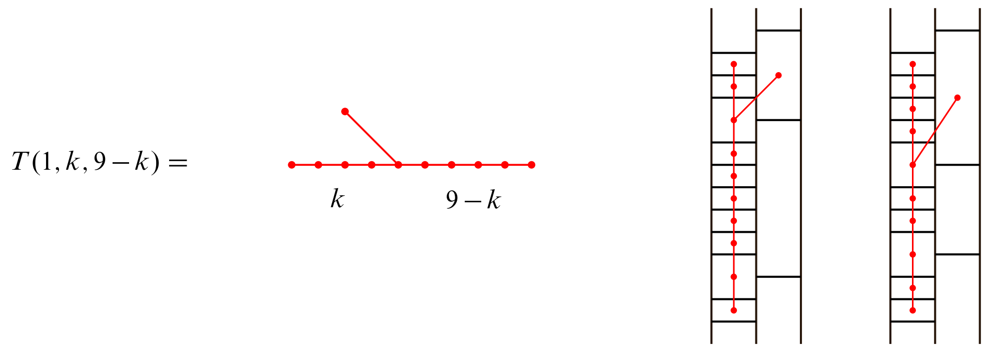


Figure 17: The tripod trees for $m = 3$ and $\sum a_i = 11$.

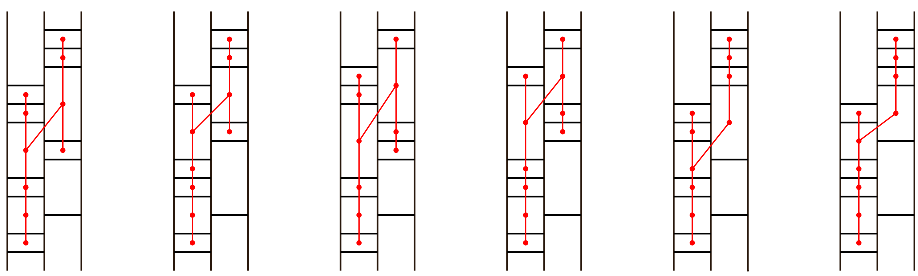


Figure 18: When $\sum a_i = 7$ and $\sum b_i = 6$, with $b_1 = 1$.

- $\boldsymbol{\sum a_i = 6}$ If $\sum b_i = 6$ we always get a link with 3 components and genus 4. If $\sum b_i = 7$ and there is at least one of the b_i equal to one, up to elementary conjugation we can assume that $\beta = \sigma_1^2\sigma_2\sigma_1^2\sigma_2^{b_2}\sigma_1^2\sigma_2^{b_3}$ with $b_2 + b_3 = 6$. Using that $\sigma_1^2\sigma_2\sigma_1^2$ commutes with σ_2 we get $\sigma_2^{b_2}\sigma_1^2\sigma_2\sigma_1^4\sigma_2^{b_3}$, which is conjugate to $\sigma_1^2\sigma_2\sigma_1^4\sigma_2^6$, whose intersection graph is a tree with the required properties. We are now left with $b_i \geq 2$ for all i. Up to elementary conjugation there is only one such braid, $\sigma_1^2\sigma_2^2\sigma_1^2\sigma_2^2\sigma_1^2\sigma_2^3$. Here there are no possible braid relations to apply and it is not possible to find a subtree of big enough genus.
- $\boldsymbol{\sum a_i = 7}$ Let us first assume that $\sum b_i = 6$. If there is at least one b_i equal to one, we can directly find our subtrees. In Figure 18 we see some of the cases. The omitted ones are symmetric

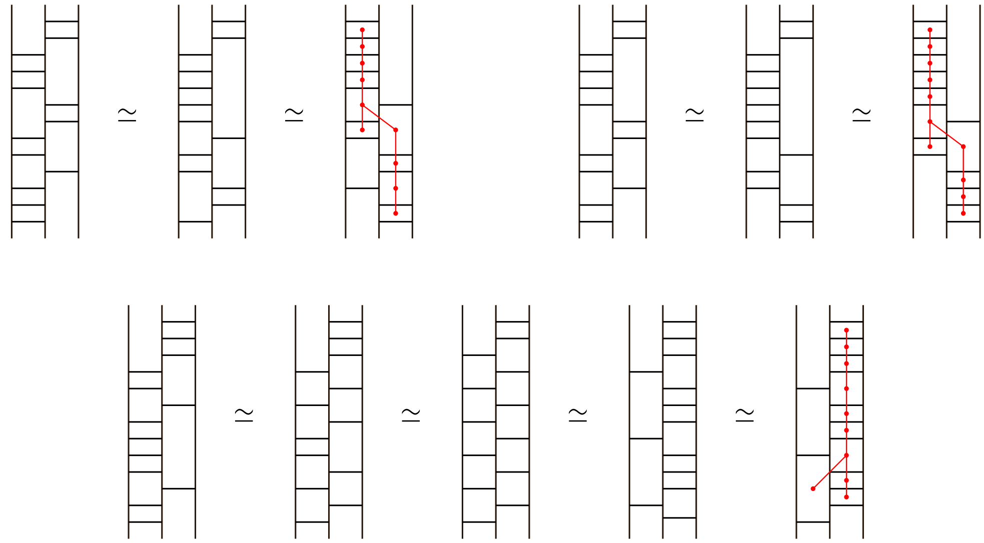


Figure 19: When $\sum a_i = 8$ and $\sum b_i = 5$.

and will give the same subtrees. Notice that this will also cover all the braids with $\sum a_i \geq 7$ and $\sum b_i \geq 7$. If $b_i = 2$ for all i, up to elementary conjugation there is only the braid $\sigma_1^3\sigma_2^2\sigma_1^2\sigma_2^2\sigma_1^2\sigma_2^2$, for which again we cannot find any subtree of big enough genus.

If $\sum b_i = 5$, up to conjugation we have $\beta = \sigma_1^3\sigma_2^{b_1}\sigma_1^2\sigma_2^{b_2}\sigma_1^2\sigma_2^{b_3}$. If $b_2 = 1$, using that $\sigma_1^2\sigma_2\sigma_1^2$ commutes with σ_2 we get the braid $\sigma_1^5\sigma_2\sigma_1^2\sigma_2^4$, whose intersection graph is a tree with the required properties. We are left with the three braids $\sigma_1^3\sigma_2\sigma_1^2\sigma_2^2\sigma_1^2\sigma_2^2$, $\sigma_1^3\sigma_2^2\sigma_1^2\sigma_2^2\sigma_1^2\sigma_2$ and $\sigma_1^3\sigma_2\sigma_1^2\sigma_2^3\sigma_1^2\sigma_2$. For the first, up to elementary conjugation and applying the commutativity relation as before we have

$$\sigma_1^3\sigma_2\sigma_1^2\sigma_2^2\sigma_1^2\sigma_2^2 \simeq \sigma_1^2\sigma_2^2\sigma_1^3\sigma_2\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2\sigma_1\sigma_2^2\sigma_1^2\sigma_2\sigma_1^2 = \sigma_1^4\sigma_2^2\sigma_1\sigma_2^3\sigma_1^2 \simeq \sigma_1^6\sigma_2^2\sigma_1\sigma_2^3$$

and we get a suitable tree. The second braid is symmetric and will lead to the same intersection tree. For the last, we similarly get

$$\sigma_1^3\sigma_2\sigma_1^2\sigma_2^3\sigma_1^2\sigma_2 = \sigma_1\sigma_2^3\sigma_1^2\sigma_2\sigma_1^4\sigma_2 \simeq \sigma_2\sigma_1\sigma_2^3\sigma_1^2\sigma_2\sigma_1^4 = \sigma_1^3\sigma_2\sigma_1^3\sigma_2\sigma_1^4 \simeq \sigma_1^7\sigma_2\sigma_1^3\sigma_2.$$

- $\boldsymbol{\sum a_i = 8}$ If $\sum b_i \geq 6$, then either we are already done by the case $\sum a_i = 7$ (if one of the b_i is equal to one) or it is symmetric to the case $\sum b_i \geq 8$. If $\sum b_i = 5$ after applying some positive braid isotopy we can always find an appropriate subtree, with the lone exception of $\beta = \sigma_1^3\sigma_2\sigma_1^3\sigma_2^2\sigma_1^2\sigma_2^2$, for which we couldn't find any. In Figure 19 we see some of the cases, the remaining ones being braid isotopic to those. Finally, if $\sum b_i = 4$, we only get links of 3 components and genus 4 except for the braid $\beta = \sigma_1^3\sigma_2\sigma_1^3\sigma_2\sigma_1^2\sigma_2^2$ (and the symmetric $\beta = \sigma_1^2\sigma_2\sigma_1^3\sigma_2\sigma_1^3\sigma_2^2$), for which we see the tree in Figure 20.
- $\boldsymbol{\sum a_i = 9}$ If $\sum b_i \geq 4$, it is enough to consider $\beta = \sigma_1^{a_1}\sigma_2\sigma_1^{a_2}\sigma_2\sigma_1^{a_3}\sigma_2^2$. By taking all the vertices of the linking graph excepted the lowermost of the right column, according to the value of a_3

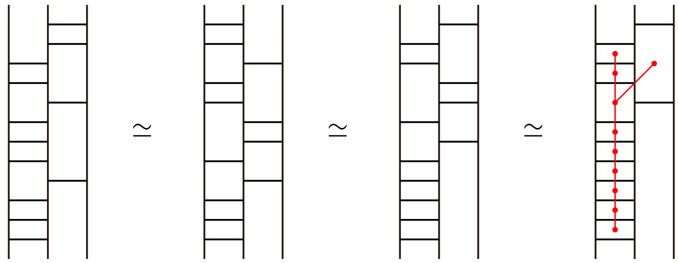


Figure 20: The braid $\sigma_1^3\sigma_2\sigma_1^3\sigma_2\sigma_1^2\sigma_2^2$.

we will get one of the tripod trees $T(1,2,6)$, $T(2,2,5)$ and $T(3,2,4)$, which all correspond to surfaces of genus 5. If $\sum b_i = 3$ and there is one even a_i, we only have to consider the three braids $\sigma_1^3\sigma_2\sigma_1^2\sigma_2\sigma_1^4\sigma_2$, $\sigma_1^3\sigma_2\sigma_1^4\sigma_2\sigma_1^2\sigma_2$ and $\sigma_1^5\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2$. The first two are symmetric, and using that $\sigma_1^2\sigma_2\sigma_1^2$ commutes with σ_2 we see that the first one is braid equivalent to the last, for which we furthermore have $\sigma_1^5\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2 = \sigma_1^7\sigma_2\sigma_1^2\sigma_2^2$, whose intersection graph is a tree. Finally, if all the a_i are odd, we get a link of genus 4.

- $\boldsymbol{\sum a_i = 10}$ The only case left is when $\sum b_i = 3$. If one of the a_i is odd we can suppose that a_3 is odd, in which case by taking all the bricks excepted the lowermost of the right column we will get a tripod tree $T(1,2,6)$ or $T(1,4,4)$, which both correspond to subsurfaces of genus 5. If all the a_i are even, up to elementary conjugation we only have the braids $\sigma_1^4\sigma_2\sigma_1^4\sigma_2\sigma_1^2\sigma_2$ and $\sigma_1^6\sigma_2\sigma_1^2\sigma_2\sigma_1^2\sigma_2$. Those are actually related by braid relations and elementary conjugations, and the very same argument used for $\sum a_i = 9$ and $\sum b_i = 3$ will yield the required tree.

$\boldsymbol{m = 2}$ For a braid $\beta = \sigma_1^{a_1}\sigma_2^{b_1}\sigma_1^{a_2}\sigma_2^{b_2}$ of genus at least 5 the intersection graph is always a tree with at least 10 crossings. Furthermore, by direct inspection we see that those trees will always contain E_6 unless they are of type D_n.

$\boldsymbol{m = 1}$ In this case we only get nonprime braids.

To sum up, the result holds for all braids excepted $\sigma_1^2\sigma_2^2\sigma_1^2\sigma_2^2\sigma_1^2\sigma_2^3$, its symmetric $\sigma_1^3\sigma_2^2\sigma_1^2\sigma_2^2\sigma_1^2\sigma_2^2$ (which gives the same link with opposite orientation) and $\sigma_1^3\sigma_2\sigma_1^3\sigma_2^2\sigma_1^2\sigma_2^2$ (which gives an invertible link). □

We will now consider braids with big positive braid index.

Proposition 5.2 *Let β be a prime positive braid on $N \geq 11$ strands and whose closure is a knot not of type A_n. Then, up to positive braid isotopy and excepted finitely many braids, its linking graph contains an induced subtree which is an E-arboreal spanning configuration on a subsurface of genus $g \geq 5$.*

The strategy to prove Proposition 5.2 is very simple: we will try to explicitly construct the required subtree and see that, each time our construction fails, either the closure is not a knot or we can reduce the

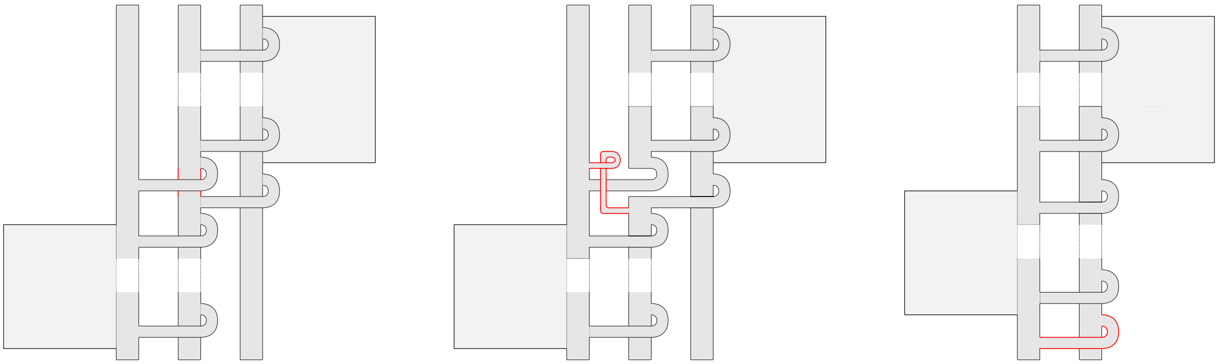


Figure 21: An isotopy that reduces the number of strands.

number of strands. The finitely many exceptions come from Proposition 5.1 and Proposition 5.6, in case we can reduce our braid to one of the exceptions therein. We will therefore heavily rely on the following two lemmas.

Lemma 5.3 *Let $\beta \in B_N^+$ be a prime positive braid on $N \geq 3$ strands. If for some i the linking graph of the subword induced by all the generators σ_i and σ_{i+1} is a path, then there exists a positive braid $\beta' \in B_{N-1}^+$ such that $\hat{\beta} = \hat{\beta}'$ and $MG(\beta) = MG(\beta')$.*

Proof Up to elementary conjugation and symmetry, we can assume that the subword induced by σ_i and σ_{i+1} is of the form $\sigma_i^a \sigma_{i+1} \sigma_i \sigma_{i+1}^b$. Moreover, we can suppose that all the generators σ_j for $j < i$ appear before the last occurrence of σ_i and all the generators σ_j for $j > i+1$ appear after the first occurrence of σ_{i+1}. In Figure 21 we see an isotopy between the fibre surface Σ_β and the fibre surface $\Sigma_{\beta'}$ of a new braid β' with one strand less: the portion of the $(i+1)^{\text{th}}$ disk lying between the first occurrence of σ_{i+1} and the last occurrence of σ_i (in red in the leftmost picture) is slid along the last σ_i, becoming a band between the i^{th} and $(i+1)^{\text{th}}$ disk (central image); this band is then slid along the back of the two disks to be brought in the lowermost position. A direct computation now shows that $MG(\beta) = MG(\beta')$. □

Lemma 5.4 *Let*

$$A = \{\sigma_1^a \sigma_2 \sigma_3^b \sigma_2 \sigma_1^c \sigma_2 \sigma_3^d \sigma_2 \sigma_1^e \mid a, b, c, d, e \in \mathbb{N}\},$$
$$B = \{\beta_1 \sigma_2 \sigma_3 \beta_2 \sigma_3 \sigma_2 \beta_3 \mid \beta_1, \beta_3 \in \langle \sigma_3, \sigma_4 \rangle,\ \beta_2 \in \langle \sigma_1, \sigma_2 \rangle\},$$
$$C = \{\beta_1 \sigma_2 \beta_2 \sigma_2 \sigma_3 \beta_3 \sigma_3 \beta_4 \mid \beta_1, \beta_4 \in \langle \sigma_1, \sigma_4 \rangle,\ \beta_2 \in \langle \sigma_3, \sigma_4 \rangle,\ \beta_3 \in \langle \sigma_1 \sigma_2 \rangle\}.$$

If $\beta \in A \cup B \cup C$, then the closure of β has at least two components.

Proof In Figure 22 we see some schematic drawings of the linking diagrams of braids from the three families, in which one component of the closure is highlighted. □

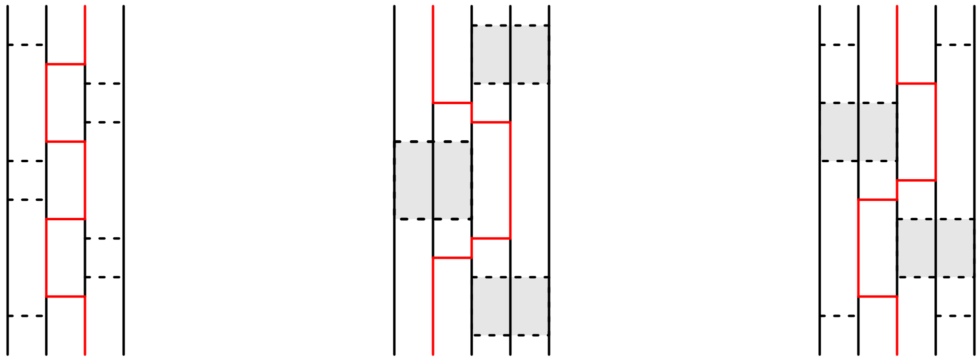


Figure 22: Some positive braids with disconnected closure.

Notice that, even though for sake of simplicity we only stated Lemma 5.4 for braids with few strands, the result clearly also applies in case some columns of the brick diagram of a braid on more strands exactly look as in Figure 22 (or are symmetric to those).

To construct the trees required in Proposition 5.2, we will also need the following lemma from [20].

Lemma 5.5 [20, Lemma 7] *Let β be a prime positive braid and v be a vertex of its linking graph. Then there is an induced path in the linking graph connecting v to any other column of the brick diagram.*

We will briefly recall the algorithm for constructing such a path, since this will be used in what follows. Let us say that we want to connect v to a column to its right. Start at v and move up or down its column until reaching the closest brick linked to the right (potentially, already v). Now, move to the right and repeat the procedure. If at the moment of moving to the right there are several possibilities, choose the brick which is the closest to a brick in the same column linked again to its right. It is easy to see that those choices prevent the creation of cycles, so that the result will be a path.

Proof of Proposition 5.2 Let β be a prime positive braid on $N \geq 11$ strands. By Lemma 5.3 we can assume that, for every pair of adjacent columns in the brick diagram, the linking graph restricted to those columns is not a path. Let us furthermore repeatedly apply all the possible braid relations of the form $\sigma_i\sigma_{i+1}\sigma_i \rightsquigarrow \sigma_{i+1}\sigma_i\sigma_{i+1}$, until no subword $\sigma_i\sigma_{i+1}\sigma_i$ is left in β. Our strategy goes as follows: We will start considering an induced path connecting the leftmost column to the rightmost, constructed with the previous algorithm, and try to add to it one single vertex, in order to get a tripod tree containing E_6. Since $b \geq 11$, the tripod tree will have at least 11 vertices and hence correspond to a subsurface of genus at least 5. So, let us fix one such path and look at the third column of the brick diagram. If we can add a brick of this column to the path and get an (induced) tripod tree we are done. There are two reasons why this might not be possible: either because there are no leftover bricks in the third column or because every available brick is linked to more than one brick of the path and adding it would generate a cycle. We will now analyse those cases in detail. By symmetry, we can assume that in the third column our path arrives from the left to a brick v_3, potentially moves *down* to a brick w_3 and then continues to the right.

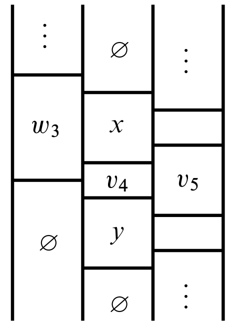


Figure 23: We only show the columns 3–5. The path goes through w_3, v_4, either x or y and v_5.

If there are no leftover bricks in the third column, then by the construction rule of our paths we know that w_3 is the only brick of column 3 linked to the right. We can now apply elementary conjugations on the right-hand side of the diagram in order to have all the generators σ_i for $i \geq 4$ appear before the last occurrence of σ_3, and perform again all the possible braid moves $\sigma_i\sigma_{i+1}\sigma_i \rightsquigarrow \sigma_{i+1}\sigma_i\sigma_{i+1}$. Those transformations will not affect the first 3 columns and the part of the path therein. We now get that the subbraid generated by σ_3 and σ_4 is $\sigma_4^a\sigma_3\sigma_4^b\sigma_3^c$, with $c \geq 1$ and $a, b \geq 2$ by Lemma 5.3. Let us denote by v_4 the only brick of column 4 linked to w_3, and let us attach a path connecting v_4 to the rightmost column.

If at least one of the bricks immediately above or below v_4 is not linked to the portion of the path in the fifth column (in particular, if v_4 is itself linked to the right), it can safely be added to get a tripod tree. We directly see that we are left with the case of Figure 23. Notice that, up to modifying the path in the fourth and fifth columns, we can always choose whether it passes by x or y. Now, if there is a brick x' above x, either it is not linked to the path in the fifth column, in what case we can directly connect it to x, or it is, in what case we can change our path to $w_3 \to v_4 \to x \to x' \to \{\textit{path in the fifth column}\}$ (thus avoiding v_5) and connect y to v_4. Similarly, we can assume that there is no brick below y.

Let us now consider the fifth column. Notice that there must be at least one brick immediately above and one immediately below v_5 that are not linked to the fourth column, otherwise we could apply one of the forbidden braid relations. By applying the same reasoning as before, we conclude that we can always obtain a tripod tree, unless there are no other bricks in the column. In the latter case, however, the closure of the braid is not a knot by Lemma 5.4 (compare with the leftmost diagram of Figure 22).

We can now suppose that there are some leftover bricks in the third column, but adding any of them to our path creates a cycle. The idea is analogous to what we just did: we will try to locally "reconstruct" the linking graph, successively exclude all the cases where we can find the required tripod and see that in the end we are left with one of the links from Lemma 5.4. However, the analysis gets much more delicate and will need lengthy case distinctions to cover the various ways adjacent columns can be connected. First of all, in the third column there could be bricks left both above and below the path, only above or only below.

I If there are bricks above v_3 and below w_3, we will be in one of the two cases of Figure 24.

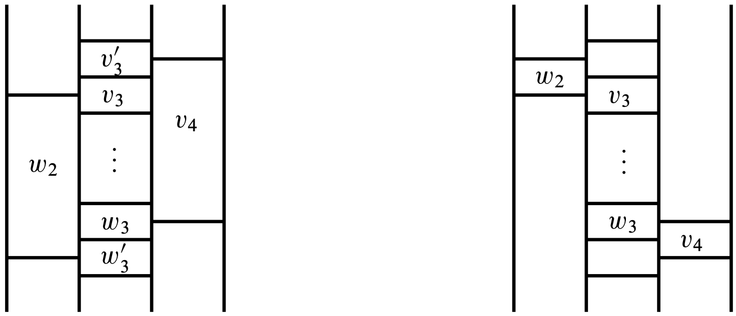


Figure 24: In both cases the path arrives from w_2, moves to v_3, then goes down to w_3 and finally to v_4.

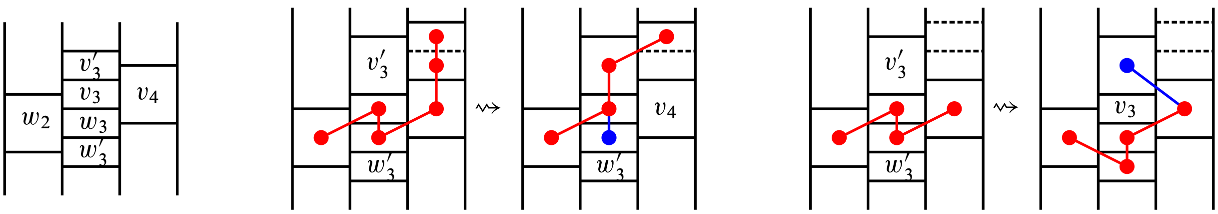


Figure 25: Diagrams for Case I.A.a; in each case, we drew on the left the original path, on the right the modified path with in blue the isolated vertex of the tripod.

- I.A In the left-hand case of Figure 24, recalling that the path was constructed with the algorithm of Lemma 5.5, we know that either v_3 and w_3 are adjacent or they coincide. We will analyse those cases in great detail, since they serve as example of the kind of reasoning applied also to the rest of the proof.
  - I.A.a If v_3 and w_3 are distinct and adjacent, as in the left of Figure 25, again by the construction rule of our paths we know that w_3' is not linked to the right at all and v_3' is not linked to the path to the left. Now, if v_3' is linked to the path to the right above v_4, we could change our path to $w_2 \to v_3 \to v_3' \to \{\textit{path in the fourth column}\}$, thus avoiding v_4, and connect w_3 to v_3 to get a tripod (see centre of Figure 25). Otherwise, we can instead consider $w_2 \to w_3' \to w_3 \to v_4 \to \{\textit{path}\}$ and connect v_3' to v_4 (right of Figure 25).
  - I.A.b If $v_3 = w_3$, then we know that w_2 has to be linked to the first column, otherwise we could perform one of the forbidden braid relations. We will further distinguish according to how w_2 is linked to the first column.
    - I.A.b.1 Let us suppose first that w_2 is linked to a brick v_1 below it, as in the left-hand side of Figure 26. Notice that the brick denoted by w_2' needs to exist because of the condition on the possible braid relations. Hence, we can assume that in the first column there are at most two bricks, both linked with w_2, and that the brick immediately below w_2' (if any) is linked with v_1, otherwise we could immediately find an appropriate tripod, as shown in Figure 26.

      We are therefore left we the diagram on the left-hand side of Figure 27. If there is a linking between the second and third columns above v_3, we could modify our path by

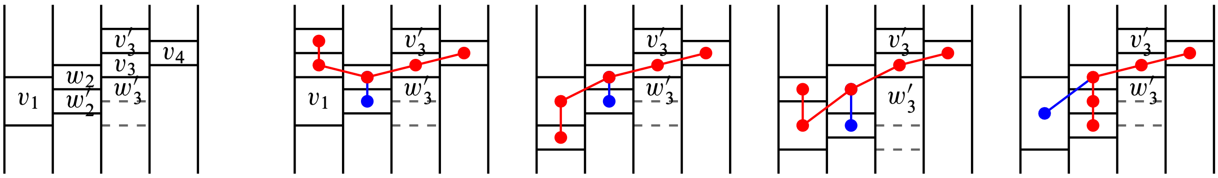


Figure 26: First diagrams for Case I.A.b.1. In the third column, the brick w_3' is linked to w_2 and may or may not be linked to w_2'.

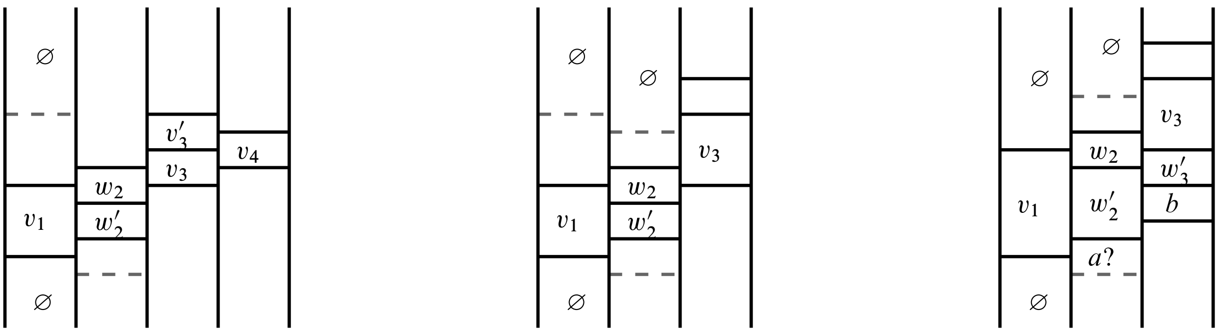


Figure 27: Further diagrams for Case I.A.b.1. The dashed lines show where the following brick (if existing) would be. In the third column, there is still a brick w_3' below v_3 as in Figure 24, which is linked to w_2 and may or may not be linked to w_2'.

starting from v_1 and w_2, then moving upwards in the second column until we reach the first connection with the third column above v_3 and finally going down on the third column until the first connection to the original path in the fourth column (which occurs at the latest at v_3'). This will give us a path avoiding v_3. We can now safely connect w_3' to w_2 and get a tripod. If not, up to elementary conjugations on the first two columns, we can suppose that there are no bricks in the second column above v_3, as in the central picture of Figure 27. In this case, we can assume that above w_2 there is at most one brick. Now, if in the first column there are two bricks, again by elementary conjugation we are back to the case where there is a brick below v_1 and we are done. We are hence left with just one brick in the first column, as in the right-hand side of Figure 27. Notice that in this case the brick w_3' is forced to be linked to w_2', otherwise the closure of the braid is not a knot by the second case of Lemma 5.4. This in turn forces the existence of the brick denoted by b below w_3', otherwise we could apply a forbidden braid relation. If there is a brick a below w_2', we can consider $v_1 \to a \to w_2' \to w_3' \to v_3 \to \{path\}$ and connect b to w_3'. On the other hand, if there are no bricks below w_2' we see that either the closure of the braid is not a knot, if there is a brick above w_2 (third case of Lemma 5.4), or we can reduce the number of strands with Lemma 5.3.

I.A.b.2 We can now suppose that w_2 is linked to a brick v_1 above it, but is not linked with any brick of the first column below it, as in the leftmost image of Figure 28. If there are at

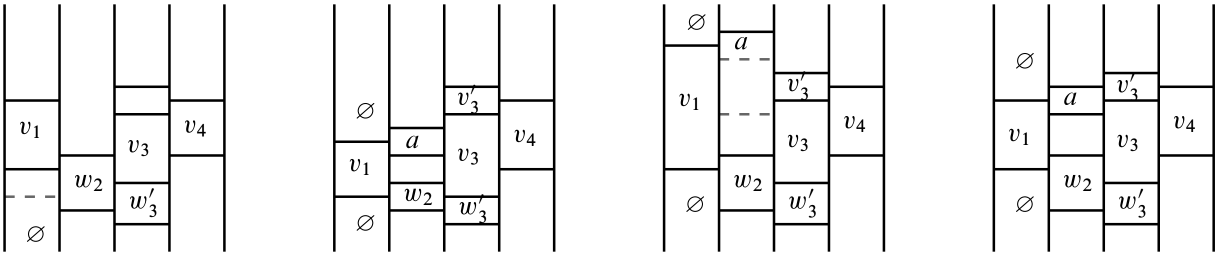


Figure 28: Diagrams for case I.A.b.2. In the second column, there is at least one brick above w_2 but below v_1.

least two bricks below w_2 we immediately find a tripod. If there is exactly one brick below w_2, we can furthermore assume that v_1 is the only brick in the first column. Let us now consider how v_1 is connected with the second column. If it is only linked to w_2, by applying an elementary conjugation we are back Case I.A.b.1, where v_1 was below w_2. Notice that the existence of a brick below w_2 ensures that the condition about the possible braid relations is still satisfied after the conjugation. If v_1 is linked to another brick of the second column above w_2, called a, and a is below v_3', as in the second image of Figure 28, we immediately see that either we find a suitable tripod or the closure is not a knot, depending on how many bricks there are in the second column between w_2 and a (there is at least one by the condition on braid relations; if it is unique, we fall in the second case of Lemma 5.4, else we find a tripod). Finally, if a is above v_3' or linked to it, as in the two right-hand side images of Figure 28, we know that there is a brick between a and w_2 linked to v_3 (potentially, this could be a). We can now consider $v_1 \to a \to \{\textit{second column}\} \to v_3 \to \{\textit{path}\}$, thus avoiding w_2, and connect w_3' to v_3.

The only case left now is when there are no bricks below w_2. Again, if v_1 is linked to another brick a of the second column above w_2 the exact same argument as before applies. If v_1 is linked only to w_2, this time we cannot simply apply an elementary conjugation to reduce to a previously treated case. However, if there are no bricks above v_1 (resp. below v_1) we could apply Lemma 5.3, whilst if there are bricks in the first column both above and below v_1 it is immediate to conclude that either we find a tripod or the closure is not a knot, as in the first case of Lemma 5.4.

I.B In the right-hand case of Figure 24, we know that w_2 needs to be linked to a brick v_1 in the first column. Again, we will separately consider whether v_1 is above or below w_2.

I.B.a Suppose first that w_2 is linked to a brick v_1 above it, as depicted in the left of Figure 29. Note that the brick denoted by v_2 must exist, otherwise we could perform a forbidden braid relation. By excluding all the cases where one can immediately find a tripod, we are left with at most two bricks in the first column, both linked to w_2, and we know that the brick above v_2 (if any) is linked to v_1; see Figure 29.

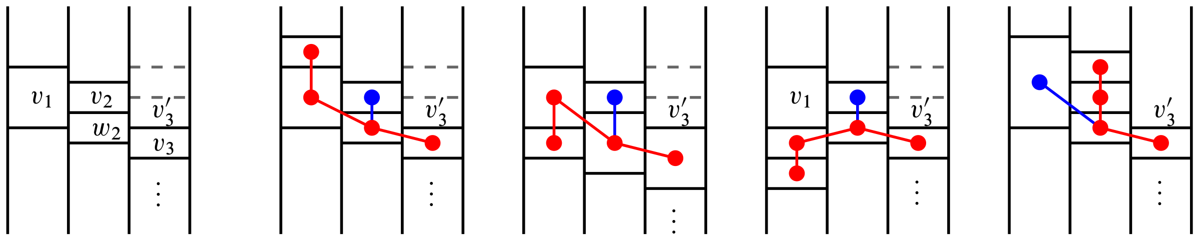


Figure 29: First diagrams for Case I.B.a. The dashed lines show where the brick v_3' could end.

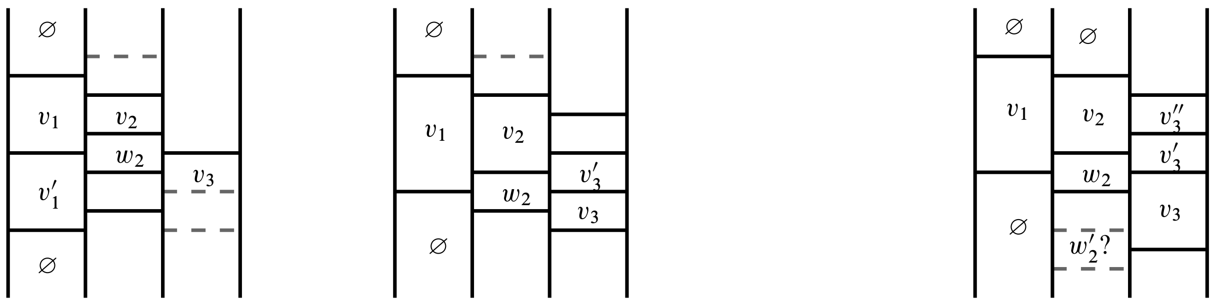


Figure 30: Additional diagrams for Case I.B.a.

We hence can reduce the study to one of the cases in the left-hand side of Figure 30. If there are two bricks in the first column, we either have a brick above v_2, in which case we can find a tripod by simply starting our path from v_1' and adding two bricks above w_2, or we can apply an elementary conjugation to the first column to get a brick below v_1', which again immediately gives a tripod. If in the first column there is just one brick, we know that v_2 needs to be linked to the third column, otherwise the closure is not a knot by Lemma 5.4 (second case). Thus, we can now suppose that there are no bricks above v_2, otherwise we immediately find a tripod, so we are left with the diagram on the right-hand side of Figure 30. Notice that now by Lemma 5.3 there needs to be at least one brick below w_2, otherwise we can reduce the number of strands. If none of the bricks below w_2 is linked to v_3, we see that according to the number of those bricks we either get a tripod or the closure is not a knot by (a symmetry of) the third case in Lemma 5.4. Hence we can suppose that there is a brick w_2' below w_2 linked to v_3. If w_2' is connected to the original path in the third column below v_3, we can instead consider $v_1 \to w_2 \to \cdots \to w_2' \to \{path\}$ and get a tripod by connecting to w_2 the bricks v_3' and v_3''. If not, we can simply take our original path starting from v_3 and connect to it w_2', v_3' and v_3''.

I.B.b Suppose now that w_2 is only linked to a brick v_1 below it, as in the leftmost image of Figure 31; note that, as depicted, there must exist one brick immediately below w_2 not linked to v_1, otherwise we could perform a braid relation. First, we immediately see that there can be at most one brick above w_2, and if this brick exists then v_1 is the only brick of the first column, otherwise we easily find a tripod. After excluding the additional easy cases shown in Figure 31,

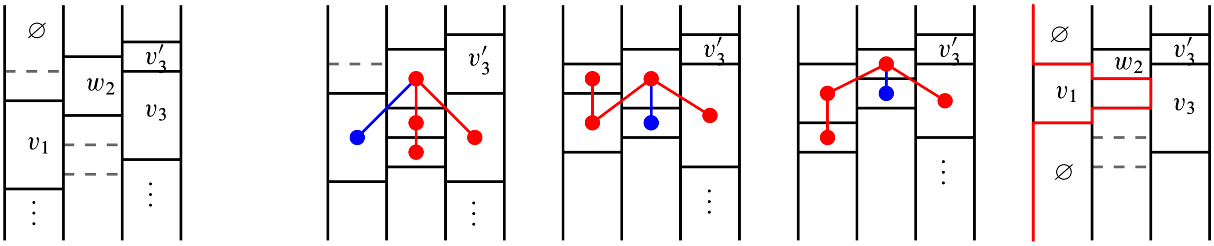


Figure 31: First diagrams for Case I.B.b.

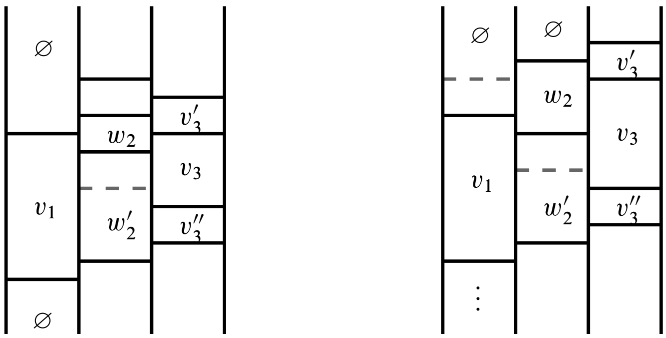


Figure 32: Additional diagrams for Case I.B.b.

we are left with the diagrams of Figure 32: that is, there must exist a brick w_2' below w_2 which is linked to v_3 but above v_1.

First, if w_2' is linked to the path in the third column below v_3, we can take $v_1 \to w_2 \to \cdots \to w_2' \to$ *path* and add to it a brick in the third column (which will be at most v_3''). Otherwise, if w_2' is not connected to the path and there is a brick w_2'' below it, we can simply take our original path from v_3 and add to it v_3', w_2' and w_2''. Finally, let's assume that there are no bricks below w_2'. If there is a brick above w_2 we can apply an elementary conjugation to the first column and get back to the previous case. If not, Lemma 5.3 forces the existence of bricks above and below v_1, in which case either we get a tripod or the closure is not a knot, as in the first case of Lemma 5.4.

II Let us now consider the case where there is at least one free brick v_3' above v_3, but none below w_3. First of all, if after w_3 our path moves to a brick v_4 of the fourth column which is below it, we are basically in the same situation as Case I.B, and the precise same arguments apply. We can hence suppose that the path moves upwards in the fourth column. We will now treat different cases according to how v_3' is linked to the neighbouring columns.

II.A If v_3' is not linked to the right, we know that it needs to be linked to a brick w_2 in the second column, which in turns needs to be linked to a brick v_1 in the first column.

II.A.a Let us suppose first that v_1 is above w_2, as in the leftmost image of Figure 33. Note that we are in a situation similar to Case I.B.a, with the only difference that now the bricks above v_3' could

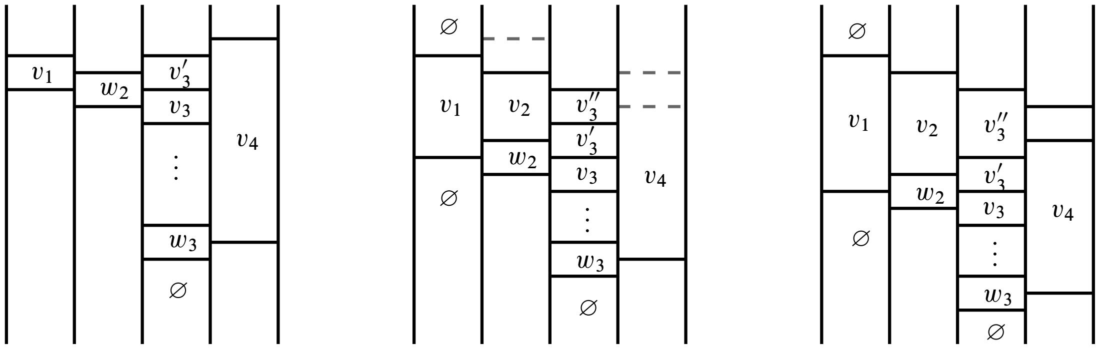


Figure 33: Diagrams for Case II.A.

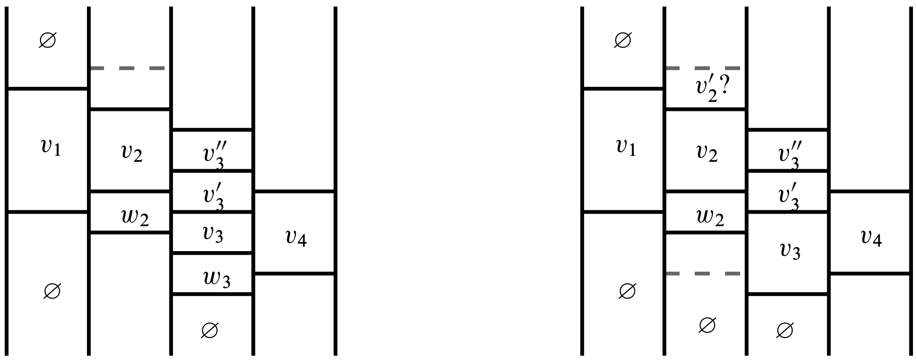


Figure 34: Diagrams for Case II.B.a.

potentially be linked to the path in the fourth column; in particular, all the arguments therein still apply to the current situation, as long as they do not involve the bricks above v_3'. Hence, by Case I.B.a, we can suppose that there is only one brick in the first column, as in the central image of Figure 33. Furthermore, if the brick v_3'' is not linked to its right, all the arguments from Case I.B.a still apply. We are then left with the rightmost diagram of Figure 33. Now, if v_3'' is not linked to the path above v_4 it can directly be added as additional vertex, otherwise we can instead consider the path $v_1 \to w_2 \to v_3' \to v_3'' \to \{path\}$ and add a brick to this new path in the fourth column.

II.A.b If v_1 is below w_2, we are in a situation analogous to Case I.B.b, and in fact all the arguments therein still apply to the current setting, as we never made use of the bricks of the third column above v_3'.

II.B If v_3' is linked to the right (to v_4) and to the left (to a brick w_2), by the construction rules of the path we know that either v_3 and w_3 are adjacent or they coincide, and by the assumption on the braid relations w_2 is linked to a brick v_1 in the first column.

II.B.a If v_1 is above w_2, after repeating the arguments of Case I.B.a we can suppose that there is only one brick in the first column, so we are left with the two diagrams of Figure 34.

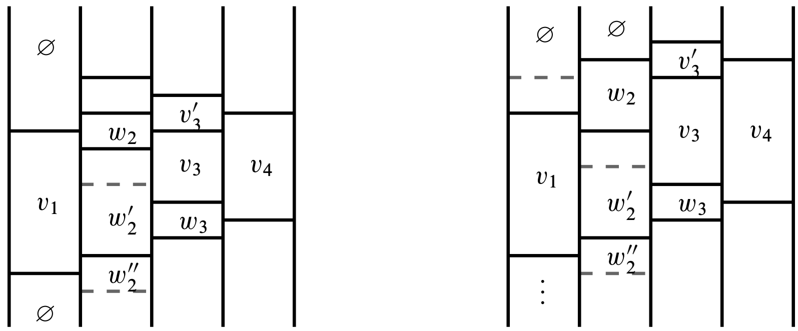


Figure 35: Diagrams for Case II.B.b.

II.B.a.1 Let us first consider the case where v_3 and w_3 are distinct and adjacent, as on the left of Figure 34. If v_3' is not linked to the path above v_4, we can simply consider $v_1 \to w_2 \to v_3' \to v_4 \to \{path\}$ and add v_3'' (notice that this would also work if v_3 and w_3 did coincide). If v_3' is linked to the path in the fourth column above v_4, take instead $v_2 \to v_3' \to \{path\}$ and add v_3 and w_3.

II.B.a.2 Suppose now that v_3 and w_3 coincide, as on the right of Figure 34. In this case, notice that no brick below w_2 can be linked to v_3 (otherwise we could perform a forbidden braid relation), and that therefore if there are at least two bricks below w_2 we immediately get a tripod. It follows that there needs to be a brick v_2' above v_2, otherwise either we can apply Lemma 5.3 (if there are no bricks below w_2) or the closure is not a knot, as in the third case of Lemma 5.4 (if there is exactly one brick below w_2). Now, if v_3' is not linked to the path in the fourth column above v_4, we can find the same tripod as in Case II.B.a.1. If v_3' is linked to the path above v_4, we can instead consider $v_2' \to v_2 \to v_3' \to \{path\}$ and add v_3.

II.B.b Finally, if v_1 is below w_2, after repeating the arguments of Case I.B.b we are left with one of the diagrams of Figure 35. Note that the case where v_3 and w_3 coincide is excluded by the condition on the braid relations. Furthermore, again by what was done in Case I.B.b, we know that we can assume the existence of a brick w_2'' below w_2'. Hence, if v_3' is not connected to the path above v_4 we can take $v_1 \to w_2 \to v_3' \to v_4 \to \{path\}$ and add w_3, if v_3' is connected to the path above v_4 we can instead take $w_2'' \to w_2' \to v_3 \to v_3' \to \{path\}$ and add w_3.

II.C If v_3' is not linked to the left, then it must be linked to the right to v_4. It follows that either v_3 and w_3 are adjacent or they coincide, as in Figure 36. In both cases, if v_3' is connected to the path above v_4, we can simply let our path pass by v_3' instead of w_3 (thus skipping v_4) and add a brick in the fourth column (which will be at most v_4').

Suppose now that v_3' is not connected to the path above v_4 and v_3, w_3 are distinct. If w_3 is linked to the left we are in the situation at the left-hand side of Figure 37 and we directly find a tripod by considering $v_1 \to w_2 \to w_3 \to v_4 \to \{path\}$ and adding v_3'. If not, we are in the situation at the right-hand side of Figure 37. Note that this is analogous to Figure 23, and the same arguments discussed there apply to the current setting.

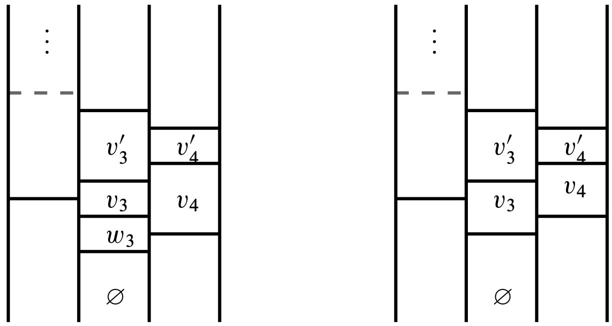


Figure 36: Diagrams for Case II.C; v_3 is linked to the second column, but v_3' is not.

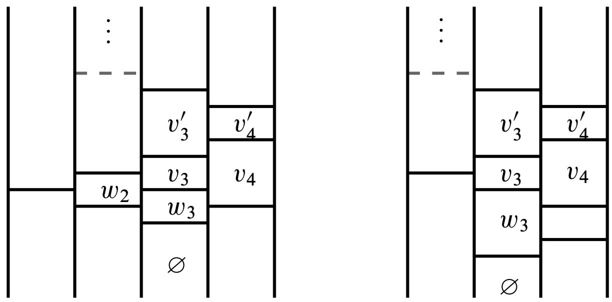


Figure 37: Diagrams for Case II.C. On the left, we know that w_2 needs to be linked to some brick v_1 in the first column.

We are left with the case where v_3 and w_3 coincide and v_3' is not connected to the path above v_4. We will now consider how the third and second column are connected.

II.C.a Let us suppose first that there is a brick v_2 in the second column below v_3. We know that v_2 needs to be linked to a brick in the first column, otherwise we could perform a forbidden braid relation.

II.C.a.1 If there is a brick v_1 in the first column above v_2, we are in one of the situations in the left of Figure 38. In both cases, we can assume that v_1 is the only brick of the first column linked to v_2, otherwise we find a tripod after elementary conjugation, as shown in the right of the figure. Moreover, in the leftmost case we now directly see that either we find a tripod (if there is at least another brick in the first column) or the closure is not a knot by Lemma 5.4.

Let us now focus on the second image of Figure 38. First of all, using Lemma 5.3 we deduce that there must be a brick in the second column above v_1, as shown in the left of Figure 39. Note that we only drew the "extremal" cases; in all the others (having either more bricks below v_2 or more bricks in the first column), one can easily find a tripod. By excluding additional direct cases, we end up with the diagram on the right-hand side of Figure 39: indeed, we can assume that there is no brick below v_2, otherwise by elementary conjugations we would get two bricks above v_2'' and would find a tripod by taking $\{\textit{second column}\} \to \cdots \to v_2' \to v_3 \to \{\textit{path}\}$ and adding v_1. With similar

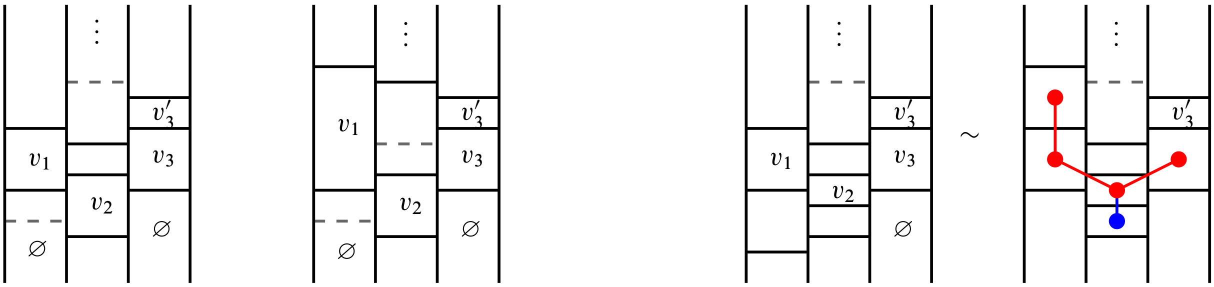


Figure 38: Diagrams for Case II.C.a.1, when v_3 and w_3 coincide and there is a brick v_2 below v_3.

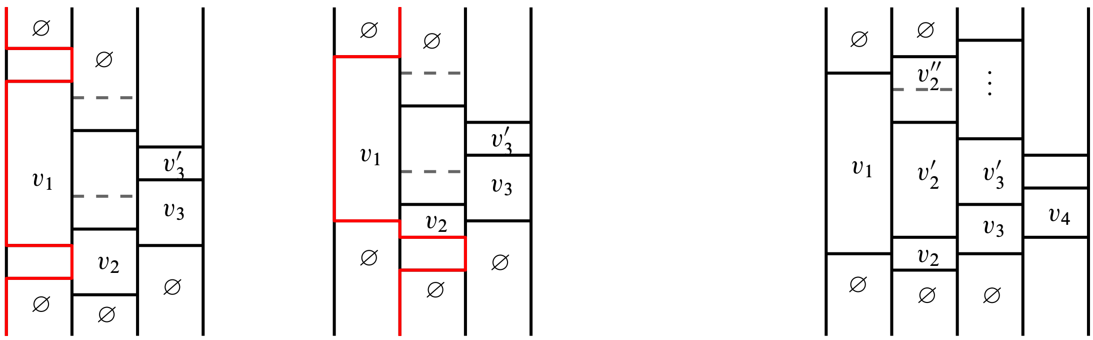


Figure 39: Additional diagrams for Case II.C.a.1.

arguments we can conclude there are no bricks in the second column above v_2'' and v_1 is the only brick of the first column. Finally, we now see that there needs to be a brick in the third column above v_2'', otherwise the closure is not a knot by the second case of Lemma 5.4. If there are at least two bricks of the third column above v_2', we get a tripod by taking $\{\textit{third column}\} \to \cdots \to v_3' \to v_4 \to \{\textit{path}\}$ and adding v_2' and v_2. Otherwise, we can consider $v_1 \to v_2'' \to \cdots \to v_2' \to v_3 \to \{\textit{path}\}$ and add the other brick in the third column linked to v_2', which now we know will not be linked to any other brick of the second column (it is also useful to remember that, as stated at the beginning of Case II.C.a, v_3', and hence all the bricks of the third column above it, is not connected to the path in the fourth column above v_4).

II.C.a.2 If there are no bricks in the first column above v_2, but v_2 is linked to a brick v_1 below it, as in the left of Figure 40, we can directly conclude that, depending on the number of bricks in the first column, either the closure is not a knot by Lemma 5.4 or we find an appropriate tripod.

II.C.b Suppose now that there are no bricks in the second column below v_3, which is therefore only linked to a brick v_2 above it.

II.C.b.1 If in the second column there are bricks both above and below v_2, noticing that if there are at least four bricks in the second column we are done, we are only left with the cases

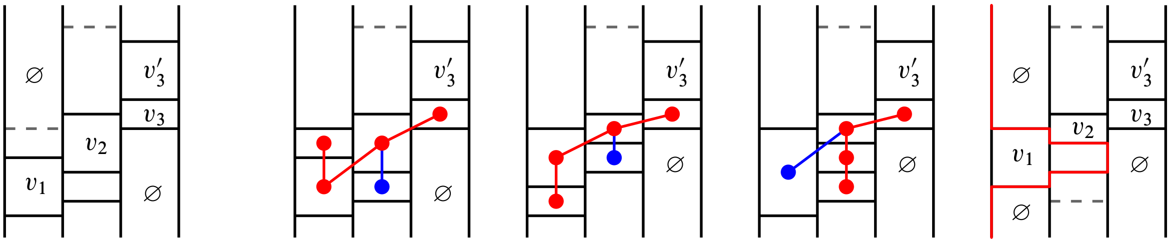


Figure 40: Diagrams for Case II.C.a.2.

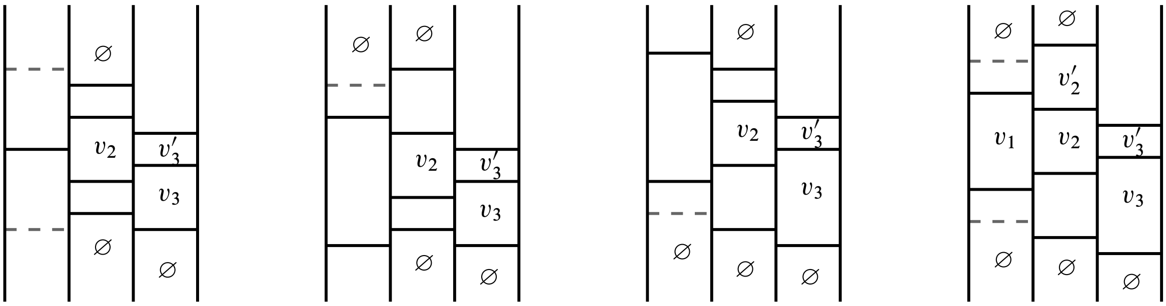


Figure 41: Diagrams for Case II.C.b.1, when v_3 and w_3 coincide and there is no brick in the second column below v_3.

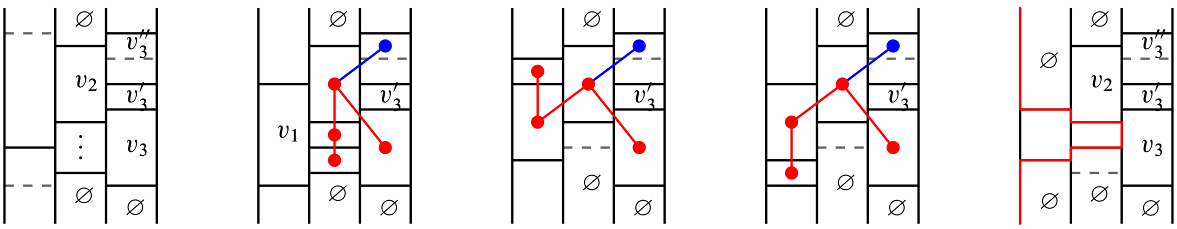


Figure 42: Diagrams for Case II.C.b.2.

of Figure 41. For the leftmost diagram, if there is only one brick in the first column the result is not a knot by Lemma 5.4, otherwise up to elementary conjugation we get a tripod. In the two central diagrams we directly find a tripod. In the rightmost diagram, if v_2' is not linked to the third column the closure is not a knot by the first case of Lemma 5.4, otherwise in the third column there is in particular a brick v_3'' linked to v_2 from above, and we get a tripod by taking $v_1 \to v_2' \to v_2 \to v_3 \to \{path\}$ and adding v_3'' (again, we use that v_3' is not linked to the path above v_4, hence v_3'' is also not).

II.C.b.2 If in the second column there are only bricks below v_2, by minimality of the number of strands there must be a brick v_3'' in the third column above v_2. If v_2 is not linked to the first column, as in the left of Figure 42, we can simply take our original path starting from the first column and add to it v_3''. If v_2 is linked to the left, notice that by the condition on

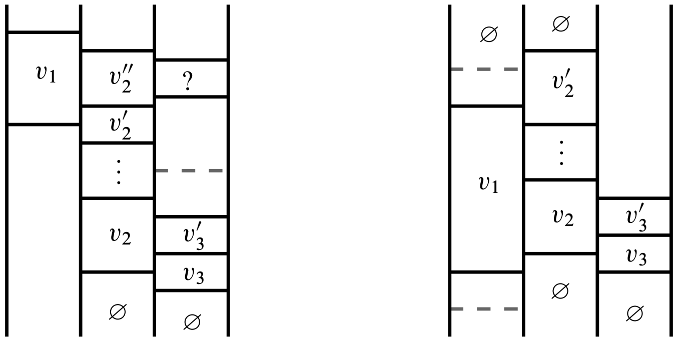


Figure 43: Diagrams for Case II.C.b.3.

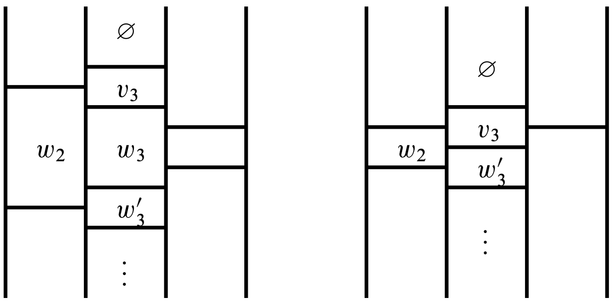


Figure 44: The two main possibilities for Case III.A.

braid relations it can only be linked to a brick v_1 from below; in Figure 42 we show that we always get a tripod or a link with at least two components.

II.C.b.3 Finally, if in the second column there are only bricks above v_2, let us consider v_2' the first brick of the second column linked to a brick v_1 of the first column (starting from v_2 upwards, potentially $v_2' = v_2$). If there is still a brick v_2'' above it, up to elementary conjugation on the first column we can assume that v_1 is above v_2', as in the left of Figure 43. We now directly see that we can suppose there is only one brick in the first column and that according to whether v_2'' is linked to its right or not, we either get a tripod or a link with more than one component by Lemma 5.4. If there are no more bricks above v_2', we are left with the diagram at the right-hand side of Figure 43. By Lemma 5.3, we know that there must be bricks above and below v_1 and we conclude with an usual argument, according to the number of those bricks.

III We finally have to treat the case where there is a free brick w_3' below w_3, but no brick above v_3. Once more, we distinguish according to how w_3' is connected to the path.

III.A Let us first suppose that w_3' is linked to a brick w_2 of the original path in the second column (which, by construction, will also be linked to v_3). Then either v_3 and w_3 are adjacent or they coincide, as in Figure 44.

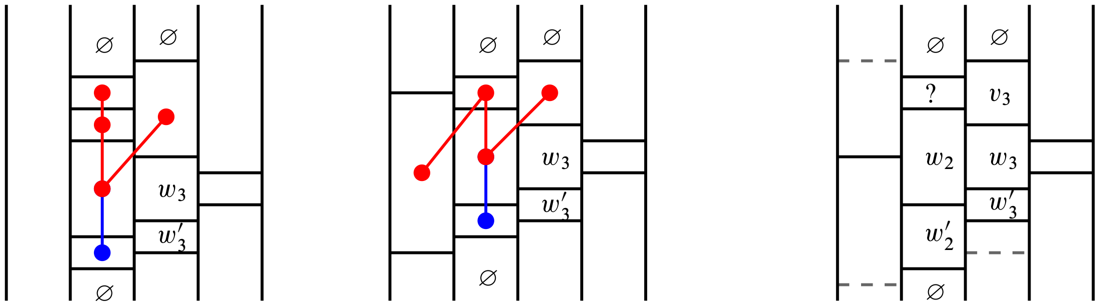


Figure 45: First diagrams for Case III.A.a, when w'_3 is linked to the second column from below.

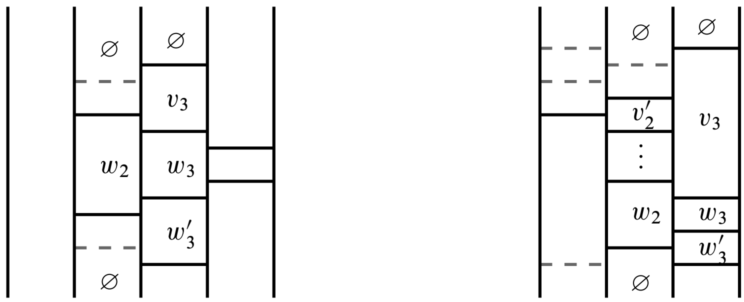


Figure 46: Final diagrams for Case III.A.a.

III.A.a If v_3 and w_3 are distinct, by construction we furthermore know that v_3 and w'_3 are not linked to any brick of the fourth column. If v_3 is linked to a brick v_2 of the second column above w_2, we know that in turns v_2 needs to be linked to the first column. In this case, we could simply connect the first column to v_3 via v_2 (thus skipping w_2), continue with our original path and add to it w'_3 to get a tripod. Similarly, suppose that w'_3 is linked to some brick w'_2 in the second column below w_2. If there is a connection between the first and second columns below w_2, the previous argument still applies: we can connect w'_3 to the first column bypassing w_2, continue with our original path from w_3 and connect v_3 as isolated leaf of the tripod. Otherwise, all the bricks in the second column below w_2 are "free" and can be added to our path. In particular, if there are at least two of them we are done. Moreover, as shown in the left of Figure 45, we also directly find a tripod if there are at least two bricks above w_2 or if w_2 is not connected to the first column. We are then now left with the rightmost diagram of Figure 45. Here it is clear that if there are at least two bricks in the first column we find a tripod (potentially after one elementary conjugation), otherwise Lemma 5.3 forces the existence of a brick above w_2, in which case the closure is not a knot by Lemma 5.4.

We can therefore now suppose that there is also no brick of the second column below w'_3, as depicted in the left of Figure 46. If in the second column there are bricks both above and below w_2, we are basically in the situation of Case II.C.b.1 (with the appropriate changes in the third column) and the same arguments apply. If there are only bricks above w_2, considering v'_2 the

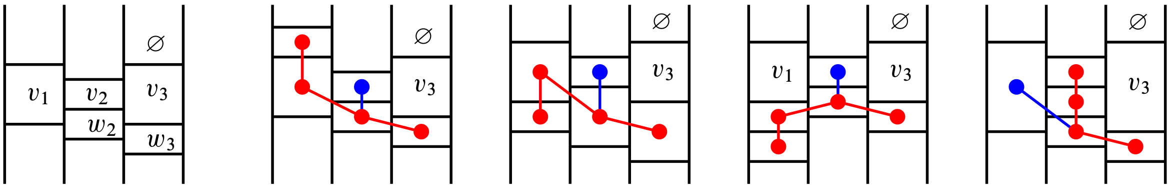


Figure 47: First diagrams for Case III.A.b.1.

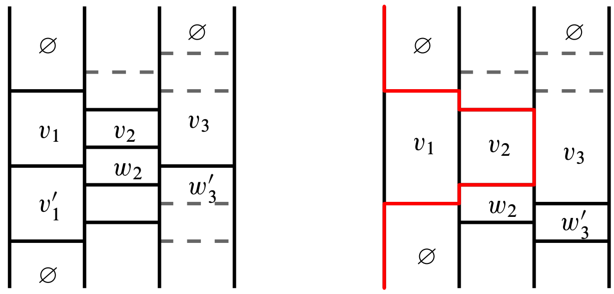


Figure 48: Final diagrams for Case III.A.b.1. Recall that there is no brick in the third column above v_3.

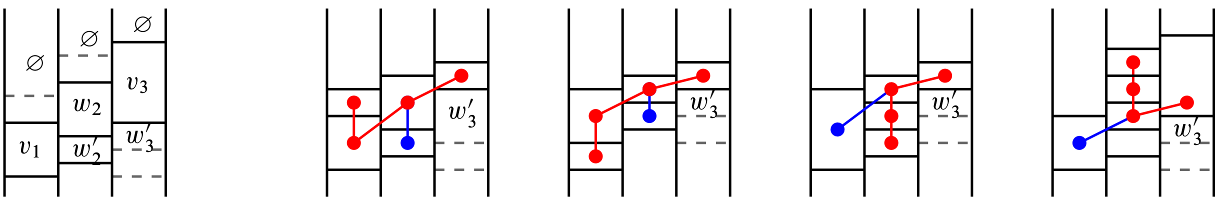


Figure 49: First diagrams for Case III.A.b.2.

first brick of the second column linked to a brick v_1 of the first column, we get a diagram as in the right of Figure 46. Note that this is analogous to Case II.C.b.3 and we conclude similarly. The case where there are only bricks below w_2 is symmetric.

III.A.b If v_3 and w_3 coincide, as in the right-hand side of Figure 44, we know that w_2 needs to be linked to a brick v_1 in the first column.

III.A.b.1 If v_1 is above w_2, as in the left of Figure 47, we are in a situation very similar to Case I.B.a. First, after removing all the cases where one can directly find a tripod, we can suppose that there are at most two bricks in the first column, both linked to w_2, and we know that the brick immediately above v_2 (if any) is linked to v_1. We are then left with diagrams as in Figure 48. In the right-hand side we directly see that the closure is not a knot, while the left-hand side can be solved as in Case I.B.a (compare with Figure 30).

III.A.b.2 If w_2 is not linked to any brick in the first column from above, as in the left of Figure 49, after removing some easy cases shown in Figure 49, we can suppose that there is at most one brick in the first column, and we are left with a diagram as in the left-hand side of Figure 50. Note that this is similar to Case I.A.b.1; compare with the rightmost diagram of Figure 27.

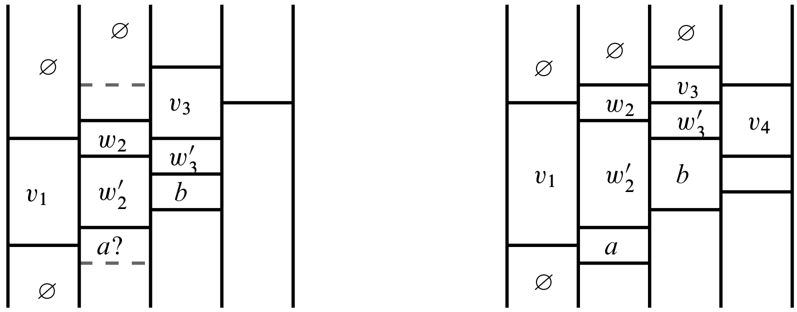


Figure 50: Additional diagrams for Case III.A.b.2.

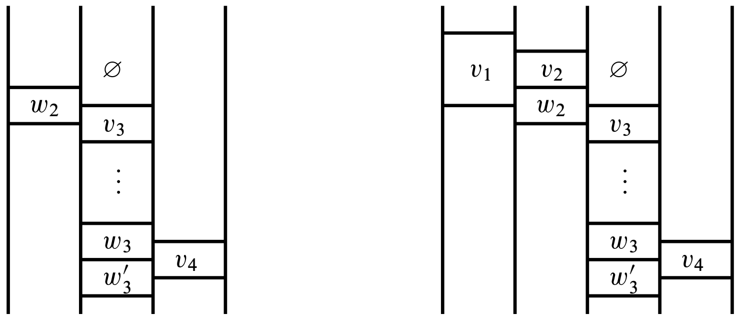


Figure 51: First diagrams for Case III.B.a.

Now, if b is not linked to the path in the fourth column, as was the case in Case I.A.b.1, or the brick denoted by a does not exist, the same argument discussed therein still works. If b is linked to the path in the fourth column from below, one can consider $v_1 \to a \to w_2' \to w_3' \to b \to \{path\}$ and add v_3 to get a tripod. Similarly if w_3' is linked to the path in the fourth column from below. Finally, if b is linked to the path in the fourth column from above but w_3' is not, we are in the case drawn in the right-hand side of Figure 50. If in the third column there are no bricks below a, one can simply perform an elementary conjugation on the second column to get a brick v_2 above w_2, take the original path starting from v_3 and add to it $w_3' \to w_2'$ and v_2 to obtain a tripod. Finally, if in the third column there is a brick below a, in particular w_2' is linked to a brick b' of the third column below b. One can hence take $v_1 \to a \to w_2' \to b' \to \cdots \to b \to v_4 \to \{path\}$ (or potentially skipping w_2' if b' is also linked to a) and connect v_3 to v_4.

III.B We now suppose that w_3' is not linked to the original path in the second column (and therefore has to be linked to the path in the fourth column). By construction, we know that v_3 is linked to some brick in the second column.

III.B.a Assume first that v_3 is linked to a brick w_2 above it, as in the left of Figure 51. Note that the situation is similar to the one analysed in Case I.B, and many of the arguments discussed therein will apply to the current case. First of all, we know that w_2 will be linked to a brick of the first

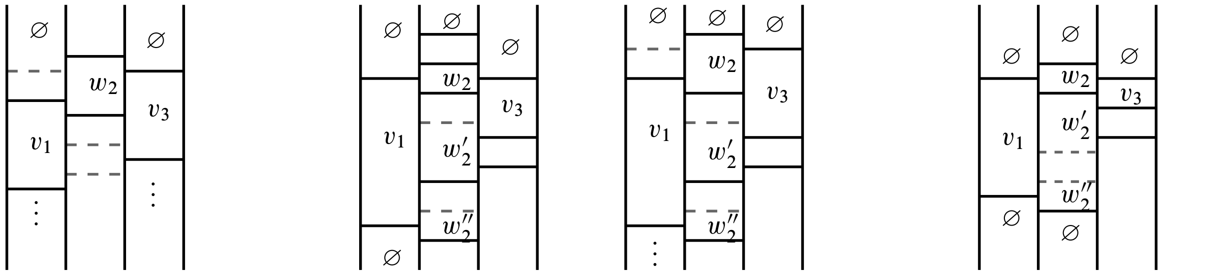


Figure 52: Diagrams for Case III.B.a.

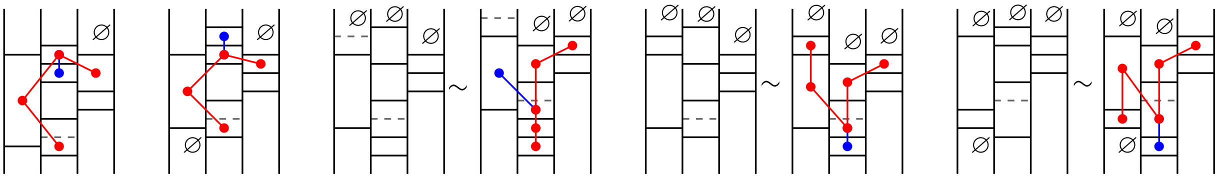


Figure 53: Additional diagrams for Case III.B.a.

column. If it is linked to a brick v_1 above it, as in the right of Figure 51, we conclude directly as in Case I.B.a (compare also with the left diagram of Figure 47 and the discussion of Case III.A.b.1).

We can now assume that w_2 is only linked to a brick v_1 below it, as in the left of Figure 52. By Case I.B.b, we are only left with the two diagrams in the centre of Figure 52, and we furthermore can assume that the brick w_2' is not linked to the original path in the third column below v_3 (so no other brick of the second column below w_2' is) and that, as drawn, there is at least one brick w_2'' in the second column below v_1 (compare with Figure 32 and the discussion preceding it).

After removing all the cases where one can directly find a tripod, as shown in Figure 53, we are left with the rightmost diagram of Figure 52.

But now we observe that there needs to be a brick in the third column below w_2'', otherwise the closure is not a knot by Lemma 5.4. In particular, w_2' is linked to a brick of the third column below v_3. Recalling that w_2' is not linked to the original path in the third column below v_3, it follows that either w_2' is linked to w_3' or to some brick below it (in the notation of Figure 51).

III.B.a.1 Let us first assume that w_2' is linked to w_3', as in Figure 54. Notice that in that case by construction v_3 is not linked to the path in the fourth column. If w_3' is only linked to the second column in w_2', as in the left of Figure 54, we can find can simply take $v_1 \to w_2'' \to \cdots \to w_2' \to w_3'' \to \{\textit{path in the fourth column}\}$ and connect v_3 to w_2'. Otherwise, we know that there exists at least one brick in the third column below w_3', as in the right of Figure 54. Consider now how w_3' is connected to the path in the fourth column: if it is only connected to v_4, all the bricks of the third column below w_3' are free to use and we can take $w_2 \to w_2' \to w_3' \to v_4 \to \{\textit{path}\}$ and connect the brick below w_3' as leaf of the tripod; if w_3' is connected to the path in the fourth column from below, via a brick w_4, take instead $v_1 \to w_2'' \to \cdots \to w_3' \to w_4'' \to \{\textit{path}\}$ and connect w_3 as a leaf.

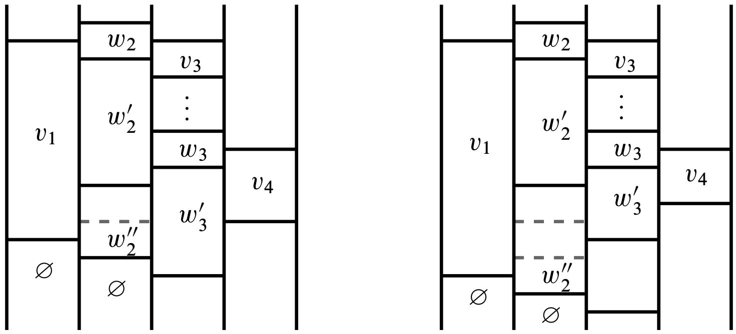


Figure 54: Diagrams for Case III.B.a.1.

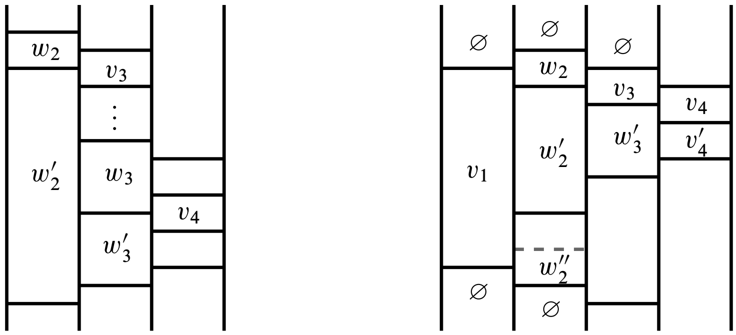


Figure 55: Diagrams for Case III.B.a.2.

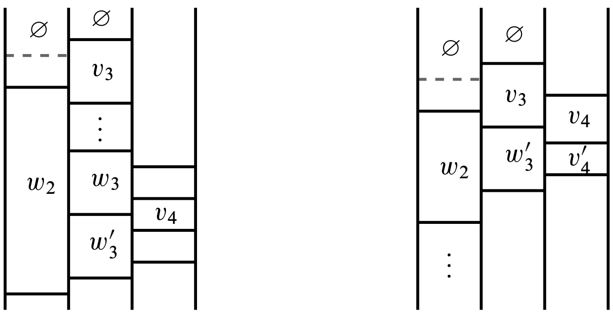


Figure 56: Diagrams for Case III.B.b.

III.B.a.2 If w_2' is linked to a brick w_3'' below w_3', we have one of the diagrams of Figure 55. If v_3 and w_3 are distinct, as in the left of Figure 55, we recognize the diagram of Figure 23, and the argument discussed there applies to the current setting. If v_3 and w_3 coincide, we have the diagram on the right of Figure 55. Once again, we consider how w_3' is connected to the path in the fourth column. If w_3' is linked to the path in the fourth column under v_4, we can simply take $v_1 \to w_2 \to v_3 \to w_3' \to \{path\}$ and add a brick of the fourth column (which will be at most v_4'). Finally, if w_3' is not linked to the path in the fourth column below v_4, then all the bricks in the third column under w_3' also are not, and can hence be freely used. If there is still at least one brick in the third column under w_3'', we can take $w_2 \to w_2' \to w_3'' \to \cdots \to w_3' \to v_4 \to \{path\}$ and add a brick below w_3'' to get a tripod. If w_3'' is the last brick of the third column, in particular it is not linked to any of the bricks below w_2', so we can take $v_1 \to w_2'' \to \cdots \to w_2' \to v_3 \to \{path\}$ and connect w_3'' to w_2'.

III.B.b We can now suppose that v_3 is only linked to a brick w_2 of the second column below it. In particular, our original path was passing by w_2, which is therefore not linked to w_3'. We now get the diagrams of Figure 56. In the left-hand side, where v_3 and w_3 are distinct, we end up with a diagram similar to Figure 23 and the exact same arguments apply. Suppose now that v_3 and w_3 coincide, as in the right-hand side of Figure 56. If w_3' is linked to the path in the fourth column under v_4, we can simply take $w_2 \to v_3 \to w_3' \to \{path\}$ and add a brick of the fourth column (which will be at most v_4'). Otherwise, we are in a situation perfectly symmetric to Case II.C.b, in particular as in Figures 41, 42 and 43, and again the same arguments apply. □

We still have to consider the braids of intermediate positive braid index. One could probably study those by hand, in a similar way to Propositions 5.1 and 5.2, but the computations would quickly get too complicated. Instead, we will treat them by directly applying Proposition 5.1, at the cost of losing some low genus cases.

Proposition 5.6 *Let β be a prime positive braid on $4 \le N \le 10$ strands whose closure is a knot not of type A_n. Suppose that β has genus $g(\beta) > 4(N-1)$. Then there exists a family of curves on Σ_β that is an E-arboreal spanning configuration on a subsurface of genus at least 5.*

The curves appearing in Proposition 5.6 will not necessarily be vertices of the intersection graph, but we might need to do some "change of basis", ie modify some of the curves by applying appropriate Dehn twists. This will change the intersection pattern of the curves in question, but not the subsurface they span nor the subgroup that the corresponding Dehn twists generate in $\mathrm{Mod}(\Sigma_\beta)$.

Proof Let β be such a positive braid. Since $g(\beta) > 4(N-1)$, there exists $1 \le i \le N-2$ such that the subword induced by all the generators σ_i and σ_{i+1} has first Betti number at least 12, when seen as a 3-braid. Let us denote this subword by $\beta_{i,i+1}$. By Proposition 5.1, either $\beta_{i,i+1}$ is positively isotopic to a 3-braid $\beta'_{i,i+1}$ containing the required spanning configuration, or it is of type A_n or D_n (the other finitely many exceptions have first Betti number 11).

In the first case, the required positive braid isotopy might not be realizable when $\beta_{i,i+1}$ is seen as a subword of β. However, since at the level of curves the effect of braid relations and elementary conjugations is obtained by Dehn twists, we can still find a family of curves in $\Sigma_{\beta_{i,i+1}} \subset \Sigma_\beta$ whose intersection pattern is equal to the linking graph of $\beta'_{i,i+1}$, and the result follows.

If $\beta_{i,i+1}$ is of type A_n, since there are only three strands one can directly verify that up to elementary conjugation its linking graph is a path. We can therefore apply Lemma 5.3 to β and reduce it to a braid with less strands.

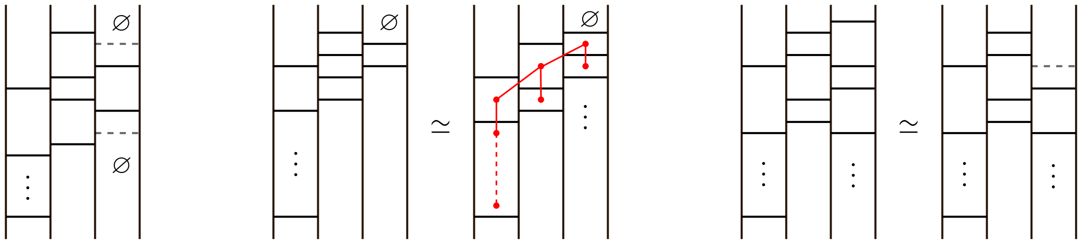


Figure 57: The columns $i, i+1$ and $i+2$ of a braid β such that $\beta_{i,i+1} = \sigma_i^{n-3}\sigma_{i+1}^2\sigma_i\sigma_{i+1}^2$.

If $\beta_{i,i+1}$ is of type D_n, up to elementary conjugation and symmetry it is of one of three forms: $\sigma_i^{n-3}\sigma_{i+1}^2\sigma_i\sigma_{i+1}^2$, $\sigma_i^{n-2}\sigma_{i+1}\sigma_i^2\sigma_{i+1}$ or $\sigma_i^a\sigma_{i+1}\sigma_i\sigma_{i+1}^b\sigma_i\sigma_{i+1}$ with $a+b=n-2$. This follows from a direct computation, or can be seen by applying the classification of checkerboard graphs of type D_n contained in Lucas Fernández Vilanova's PhD thesis [12]. In all the cases one can see that, if the closure is connected, we can always add a brick in a neighbouring column and find the required subtree. We will do it for $\beta_{i,i+1} = \sigma_i^{n-3}\sigma_{i+1}^2\sigma_i\sigma_{i+1}^2$, the others are analogous. In this case, we know that $i < N-2$, otherwise the closure is not a knot by Lemma 5.4. Since β is prime, its intersection graph is connected, so at least one of the three bricks in the $(i+1)^{\text{th}}$ column needs to be linked to its right. After removing the cases where one directly finds an appropriate subtree, we are left with one of the three cases of Figure 57. The first one is excluded since the closure is not a knot; in the second one we can find a subtree after braid relation, as shown in the figure; for the third one, up to elementary conjugation we can suppose that there are no generators σ_{i+2} above the last occurrence of σ_{i+1}. Now we see that if there are at least two bricks in the $(i+2)^{\text{th}}$ column we are done, otherwise either the closure is not a knot (if $i+2 = N-1$) or we can still add one brick further to the right and again find the required subtree. □

Everything is now ready to prove our main theorem.

Proof of Theorem 1.2 Let β be a prime positive braid not of type A_n and whose closure is a knot. We want to prove that $MG(\beta) = \mathrm{Mod}(\Sigma_\beta, \phi_\beta)$ by using Proposition 4.7. Let $V = \{\gamma_1, \dots, \gamma_{2g}\}$ be the family of standard curves on Σ_β corresponding to the vertices of the linking graph of β. In Propositions 5.1, 5.2 and 5.6 we have constructed the starting E-arboreal spanning configuration of genus $h \geq 5$ for all but finitely many such prime positive braids. In general, this is obtained by taking a subfamily of curves $V_0' \subset V$ and potentially modifying some of them by applying Dehn twists around other curves of V_0', obtaining a family V_0 of curves in Σ_β. In particular, the subsurface spanned by V_0 is the same as the subsurface spanned by V_0'. It is now clear that the remaining curves of $V \setminus V_0'$ can be attached in an order that respects the definition of h-assemblage, so that

$$\mathrm{Mod}(\Sigma_\beta, \phi_\beta) = \langle T_c \mid c \in V_0 \cup (V \setminus V_0') \rangle = \langle T_c \mid c \in V \rangle = MG(\beta). \qquad \square$$

Remark 5.7 In fact, our proof of Theorem 1.2 also applies to many links. Indeed, the requirement of the closure of β being a knot was uniquely used to exclude links as in Lemma 5.4: all these have one unknotted component whose total linking number with the other components is precisely 2. In particular, the proof works without problems for links whose components are all knotted or whose pairwise linking numbers are all big enough.

Interestingly, this is essentially always the case in the special class of links of singularities, if we exclude the special families A_n and D_n. In what follows, the reader can refer to [7] for the background material on plane curve singularities. If f_1 and f_2 are irreducible singularities with associated knots K_1 and K_2, then the link of $f = f_1 f_2$ is $L(f) = K_1 \cup K_2$, and the linking number $lk(K_1, K_2)$ equals the intersection multiplicity of the two branches. It follows that in the link of a singularity all linking numbers are strictly positive. Now, let f be a singularity whose link has a component which is unknotted and has total linking number with the other components equal to 2, as in Lemma 5.4. By the previous discussion, f has at most three branches. Suppose first that $f = f_1 f_2$ has only two branches, and $L(f) = K_1 \cup K_2$. Since one component is the unknot and the multiplicity of a singularity equals the braid index of the associated link by [28], we can assume that $f_2 = y + x\tilde{f}(x, y)$. Let now m be the multiplicity and $y = g(x^{1/m})$ the Puiseux series of f_1; we obtain $2 = lk(K_1, K_2) = \operatorname{ord}\big(g(t) + t^m \tilde{f}(t^m, g(t))\big) \geq m$, from which we conclude that K_1 has braid index at most 2. Finally, since the link of a reducible singularity is determined by the components and the pairwise linking numbers, and all the possible pairs of a positive 2-braid and an unknot with linking number 2 are realized by singularities of type A_n or D_n, it follows that f belongs to one of those two families. Similarly, if f has three branches one can conclude that all the components of $L(f)$ are unknotted, so that the link is determined by the triple of linking numbers (where two of the linking numbers are now equal to 1). Since all such triples are realized by singularities of type D_n, f must belong to this family. Therefore, up to finitely many low genus exceptions, we completely recover the main result of [24], saying that the geometric monodromy group of a singularity not of type A_n and D_n is a framed mapping class group.

Remark 5.8 In contrast to the case of singularities, it does not seem possible to extend the proof to all positive braid links. Even excluding the two exceptional families A_n and D_n, there are other infinite families, both with bounded and unbounded braid index, that most likely do not contain an E_6. For example, we could not find such subtrees for the braids $\beta_n = \sigma_1\sigma_2^2\sigma_1\sigma_2^{n-4}\sigma_3\sigma_2^2\sigma_3 \in B_4^+$, whose linking graph is the extended Dynkin diagram $\widetilde{D}_n$, nor for $\beta_N = (\sigma_1 \cdots \sigma_N\sigma_N \cdots \sigma_1)^2 \in B_{N+1}^+$. We do not know whether the corresponding monodromy groups are equal to the whole framed mapping class group.

References

[1] **N A'Campo**, *Sur la monodromie des singularités isolées d'hypersurfaces complexes*, Invent. Math. 20 (1973) 147–169 MR Zbl

[2] **N A'Campo**, *Le groupe de monodromie du déploiement des singularités isolées de courbes planes, I*, Math. Ann. 213 (1975) 1–32 MR Zbl

[3] **N A'Campo**, *Generic immersions of curves, knots, monodromy and Gordian number*, Inst. Hautes Études Sci. Publ. Math. 88 (1998) 151–169 MR Zbl

[4] **N A'Campo**, *Real deformations and complex topology of plane curve singularities*, Ann. Fac. Sci. Toulouse Math. 8 (1999) 5–23 MR Zbl

[5] **S Baader**, **L Lewark**, **L Liechti**, *Checkerboard graph monodromies*, Enseign. Math. 64 (2018) 65–88 MR Zbl

[6] **S Baader**, **M Lönne**, *Secondary braid groups*, preprint (2020) arXiv 2001.09098

[7] **E Brieskorn**, **H Knörrer**, *Plane algebraic curves*, Birkhäuser, Basel (1986) MR Zbl

[8] **A Calderon**, **N Salter**, *Framed mapping class groups and the monodromy of strata of abelian differentials*, J. Eur. Math. Soc. 25 (2023) 4719–4790 MR Zbl

[9] **O Couture**, **B Perron**, *Representative braids for links associated to plane immersed curves*, J. Knot Theory Ramifications 9 (2000) 1–30 MR Zbl

[10] **P R Cromwell**, *Positive braids are visually prime*, Proc. Lond. Math. Soc. 67 (1993) 384–424 MR Zbl

[11] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Math. Ser. 49, Princeton Univ. Press (2012) MR Zbl

[12] **L R Fernández Vilanova**, *Positive Hopf plumbed links with maximal signature*, PhD thesis, Universität Bern (2021) Available at `https://boristheses.unibe.ch/3661/`

[13] **H Goda**, **M Hirasawa**, **Y Yamada**, *Lissajous curves as A'Campo divides, torus knots and their fiber surfaces*, Tokyo J. Math. 25 (2002) 485–491 MR Zbl

[14] **S M Gusein-Zade**, *Dynkin diagrams for singularities of functions of two variables*, Funkcional. Anal. i Priložen. 8 (1974) 23–30 MR Zbl In Russian; translated in Funct. Anal. Appl. 8 (1974) 295–300

[15] **S M Gusein-Zade**, *Intersection matrices for certain singularities of functions of two variables*, Funkcional. Anal. i Priložen. 8 (1974) 11–15 MR Zbl In Russian; translated in Funct. Anal. Appl. 8 (1974) 10–13

[16] **M Hirasawa**, *Visualization of A'Campo's fibered links and unknotting operation*, Topology Appl. 121 (2002) 287–304 MR Zbl

[17] **M Ishikawa**, *Plumbing constructions of connected divides and the Milnor fibers of plane curve singularities*, Indag. Math. 13 (2002) 499–514 MR Zbl

[18] **D Johnson**, *Spin structures and quadratic forms on surfaces*, J. Lond. Math. Soc. 22 (1980) 365–373 MR Zbl

[19] **C Labruère**, *Generalized braid groups and mapping class groups*, J. Knot Theory Ramifications 6 (1997) 715–726 MR Zbl

[20] **L Liechti**, *On the genus defect of positive braid knots*, Algebr. Geom. Topol. 20 (2020) 403–428 MR Zbl

[21] **M Lönne**, *Fundamental group of discriminant complements of Brieskorn–Pham polynomials*, C. R. Math. Acad. Sci. Paris 345 (2007) 93–96 MR Zbl

[22] **J Milnor**, *Singular points of complex hypersurfaces*, Ann. of Math. Stud. 61, Princeton Univ. Press (1968) MR Zbl

[23] **B Perron**, **J P Vannier**, *Groupe de monodromie géométrique des singularités simples*, Math. Ann. 306 (1996) 231–245 MR Zbl

[24] **P Portilla Cuadrado**, **N Salter**, *Vanishing cycles, plane curve singularities and framed mapping class groups*, Geom. Topol. 25 (2021) 3179–3228 MR Zbl

[25] **O Randal-Williams**, *Homology of the moduli spaces and mapping class groups of framed, r–Spin and Pin surfaces*, J. Topol. 7 (2014) 155–186 MR Zbl

[26] **L Ryffel**, *Curves intersecting in a circuit pattern*, Topology Appl. 332 (2023) art. id. 108522 MR Zbl

[27] **B Wajnryb**, *Artin groups and geometric monodromy*, Invent. Math. 138 (1999) 563–571 MR Zbl

[28] **R F Williams**, *The braid index of an algebraic link*, from "Braids", Contemp. Math. 78, Amer. Math. Soc., Providence, RI (1988) 697–703 MR Zbl

Section de mathématiques, Université de Genève
Genève, Switzerland

livio.ferretti@unige.ch

Received: 25 May 2022 Revised: 4 May 2023

Algebraic & Geometric Topology **25**:1 (2025) 207–243
DOI: 10.2140/agt.2025.25.207
Published: 5 March 2025

Highly twisted diagrams

NIR LAZAROVICH
YOAV MORIAH
TALI PINSKY

We prove that knots and links that have a 3-highly twisted irreducible diagram with more than two twist regions are hyperbolic. Furthermore, this result is sharp. The result is obtained using combinatorial techniques, using a new approach involving the Euler characteristic. By using geometric techniques, Futer and Purcell proved hyperbolicity under the assumption that the diagram is 6-highly twisted.

57K10, 57K32, 57K99

1 Introduction

The prevailing feeling among low dimensional topologists is that "complicated" links $\mathcal{L}$ in $\mathbb{S}^3$ are hyperbolic, ie the open manifold $\mathbb{S}^3 \setminus \mathcal{L}$ can be endowed with a complete, finite-volume, hyperbolic metric of sectional curvature -1. Being hyperbolic is a property of the manifold with far reaching consequences. However, proving that a specific link $\mathcal{L}$ is hyperbolic turns out to be nontrivial. This is especially true if the link $\mathcal{L}$ is "heavy duty", ie it has a very large crossing number. See for example Minsky and Moriah's work [12].

Our main theorem is:

Theorem A *Let $D(\mathcal{L})$ be a connected, prime, twist-reduced, 3-highly twisted diagram of a link $\mathcal{L}$ with at least two twist regions. Then $\mathcal{L}$ is hyperbolic.*

For the definitions see Section 2. Intuitively, being 3-highly twisted means that any crossing of the diagram is part of a sequence of at least 3 crossings of the same strings. For example, the diagram of the link in Figure 2 is 3-highly twisted. Clearly not all links have a diagram that satisfies the conditions of the theorem. However the subset of links that do is a "large" subset in a sense that can be made precise, see the discussion in [10] by Lustig and Moriah.

The assumption of being 3-highly twisted makes Theorem A *sharp* since there are nonhyperbolic links with 2-highly twisted link diagrams, as Figure 1 shows.

The question of when can one decide if the complement of a link in $\mathbb{S}^3$ is a hyperbolic manifold from a projection diagram has been of interest for a long time. The first result in this direction is by Hatcher and

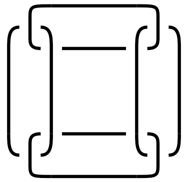


Figure 1: A nonhyperbolic link with a 2-highly twisted diagram.

Thurston [7] who proved that complements of 2-bridge knots which have at least two twist regions (they are not torus knots or links) are hyperbolic. The second is Menasco's result [11] that a nonsplit prime alternating link which is not a torus link is hyperbolic. Later Futer and Purcell proved in [4], among other results, that every link with a connected, prime, twist-reduced, 6-highly twisted diagram which has at least two twist regions is hyperbolic. Two other relevant results are by Giambrone [5] and by Futer, Kalfagianni and Purcell [3], in which the condition that the diagram is 6-highly twisted is replaced by conditions related to its "semiadequacy" (as defined there).

Theorem A, which is proved using combinatorial techniques, weakens the conditions imposed by Futer and Purcell [4] on $D(\mathcal{L})$ from 6-highly twisted to 3-highly twisted. Their result is obtained by applying geometric bounds, using Lackenby's 6-surgery theorem, see [8], to the corresponding fully augmented links. The fact that combinatorial techniques can be used to improve on geometric bounds is not surprising and was repeatedly demonstrated in the study of three manifolds, for example in work by Culler, Gordon, Luecke and Shalen [2], Gordon and Luecke [6] and Li [9].

We believe that the methods used in this paper are interesting in themselves, and could be used in studying other problems. For example, as a corollary to Theorem A we obtain a simple method to construct essential surfaces in complements of links with highly twisted diagrams. This is stated in Section 7 as Theorem B.

Although there are nonhyperbolic links with 2-highly twisted diagrams, we expect that the 3-highly twisted condition can be weakened to generalize Theorem A to a larger class of links which includes alternating links (cf [11]).

Outline of the proof

By Thurston [14], it suffices to show that the link complement has incompressible boundary, and is irreducible, atoroidal and unannular. These can be formulated as the nonexistence of an essential surface S of nonnegative Euler characteristic. Given an essential surface S in the link complement, consider its curves of intersection $\mathcal{C}$ with the projection plane P. To each curve $c \in \mathcal{C}$ we assign its "contribution" to the Euler characteristic $\chi_+(c)$ and show that $0 \le \chi(S) = \sum_{c \in \mathcal{C}} \chi_+(c)$ (see Lemma 4.2). In general there might be curves $c \in \mathcal{C}$ with $\chi_+(c) > 0$. Given such a curve, the 3-highly twisted condition forces the existence of neighboring curves with negative χ_+. This allows us to redistribute the Euler characteristic among the curves by defining for each $c \in \mathcal{C}$ a modified Euler characteristic $\chi'(c)$ so that $0 \le \chi(S) = \sum_{c \in \mathcal{C}} \chi'(c)$ and $\chi'(c) \le 0$ (see Lemmas 5.15 and 5.16). This shows that $\chi(S) \le 0$. Moreover, we get that $\chi'(c) = 0$ for all $c \in \mathcal{C}$. A case by case analysis then shows that all curves must be of a particular form (see Definition 6.3 and Propositions 6.7 and 6.13) from which it follows that S must be a boundary parallel torus.

Acknowledgments

Lazarovich was supported by the Israel Science Foundation (grant 1562/19). The authors thank the anonymous referee for helpful comments.

2 Preliminaries

2.1 Bubbles and twist regions

Let $\mathcal{L} \subset \mathbb{S}^3$ be a link. The projection of a link L in the isotopy class $\mathcal{L}$ onto a plane P together with the crossing data is a *link diagram* of L and is denoted by $D(L)$. Let ε be sufficiently small so that the closed ε-balls around the crossings of $D(L)$ are disjoint. Let $\mathcal{B}_1, \dots, \mathcal{B}_r$ be ε-balls around the r crossings of the diagram. The boundaries $B_i = \partial \mathcal{B}_i$, $1 \le i \le r$, are the *bubbles* of the diagram. The link $\mathcal{L}$ is isotopic to a link L which is embedded in $P \cup \bigcup_i B_i$. Note that P divides each bubble into two hemispheres denoted by B_i^+ and B_i^-. Denote the two 2-spheres

$$P^{\pm} = \Big(P \setminus \bigcup_i \mathcal{B}_i\Big) \cup \bigcup_i B_i^{\pm}.$$

Each of P^+, P^- bounds a 3-ball $H^{\pm}$ in $\mathbb{S}^3 \setminus L$.

A *twist region* T is a disk in P which contains a maximal (with respect to inclusion) chain of bigons in $D(L)$ describing a trivial integer 2-tangle. See Figure 2 for an example of a diagram with twist regions. We will assume that a twist region contains the projection of the bubbles around the crossings in T. We

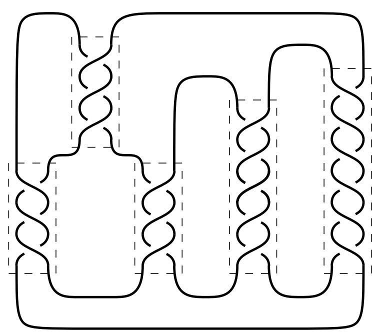


Figure 2: A 3-highly twisted link diagram. The twist regions are the dashed rectangles.

will often abuse terminology, and use twist regions to refer to the regions in $P^{\pm}$ which project to twist regions. Correspondingly, we treat the bubbles around the crossings of T as being contained in T. A *twist box* is the tangle $(\boldsymbol{T}, t)$ where $\boldsymbol{T}$ is the product $T \times [-2\varepsilon, 2\varepsilon]$ for a twist region T, and t is the tangle $\boldsymbol{T} \cap L$. The diagram $D(L)$ uniquely decomposes into disjoint twist regions.

Definition 2.1 Let $D(L)$ be a link diagram.

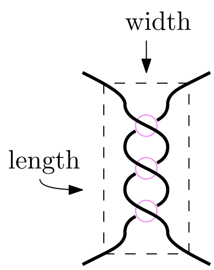
(1) The diagram $D(L)$ is *prime* if any simple closed curve in P intersecting $D(L)$ transversely in two points bounds a subdiagram with no crossings.

(2) A *twist-reduction subdiagram* is a subdiagram of $D(L)$ enclosed by a simple closed curve γ in P which intersects $D(L)$ transversely in four points composed of two pairs each of which is adjacent to a crossing of $D(L)$ but which is not a chain of bigons describing an integer 2-tangle. The diagram $D(L)$ is *twist-reduced* if it contains no twist-reduction subdiagram.

(3) For $k \in \mathbb{N}$, the diagram is *k-highly twisted* if every twist region has at least k crossings.

Note that every diagram can be made twist-reduced by performing flypes on twist-reduction subdiagrams.

Definition 2.2 A twist region T intersects the link diagram in four points, dividing its boundary ∂T into four segments. If the twist region has at least two crossings, then a pair of opposite segments of ∂T can be called the *length edge* or *width edge* of T as in



3 Surfaces in link complements

3.1 Normal position

We are interested in studying compact surfaces S properly embedded in $\mathbb{S}^3 \setminus \mathcal{N}(L)$. If $\partial S \neq \varnothing$ we extend S by shrinking the neighborhood $\mathcal{N}(L)$ radially. This determines a map $\iota\colon (S, \partial S) \to (\mathbb{S}^3, L)$, whose image we denote by S as well, which is an embedding on the interior of S.

Lemma 3.1 *Let $S \subset \mathbb{S}^3 \setminus \mathcal{N}(L)$ be a proper surface with no meridional boundary components, and let $(\boldsymbol{T}, t)$ be a twist box. Then, up to isotopy, $S \cap \boldsymbol{T}$ is a disjoint union of disks $D \subset (\boldsymbol{T}, t)$ of one of the following three types*:

Type 0 *D separates the two strings of t.*

Type 1 *∂D decomposes as the union of two arcs $\alpha \cup \beta$ such that $\alpha \subset t$ and $\beta \subset \partial \boldsymbol{T}$.*

Type 2 *∂D decomposes as the union of four arcs $\alpha_1 \cup \beta_1 \cup \alpha_2 \cup \beta_2$ where $\alpha_i \subset t_i$ and $\beta_i \subset \partial \boldsymbol{T}$.*

Moreover, the isotopy decreases the number of bubbles that S meets and we may further assume that $\iota|_{\partial L} \colon \partial S \to L$ is a covering map.

Proof If no component of ∂S is a meridian, we may assume that up to isotopy $\iota|_{\partial L} \colon \partial S \to L$ is a covering map.

The twist box $(\boldsymbol{T}, t)$ is a trivial 2-tangle. The complement $\boldsymbol{T} \smallsetminus \mathcal{N}(t)$ can be identified with $U \times [0,1]$, where U is a twice holed disk. Let E be the disk $\alpha \times [0,1]$, where α is the simple arc connecting the two holes of U. Up to a small isotopy, we may assume that S intersects E transversely. Since the bubbles in $\boldsymbol{T}$ are in some neighborhood of E, we may assume that S meets a bubble if it does so in E. The intersection $S \cap E$ comprises of simple closed curves and arcs. All curves and arcs except those connecting $\alpha \times \{0\}$ to $\alpha \times \{1\}$ can be eliminated by an isotopy pushing S off $\boldsymbol{T}$. This isotopy decreases the number of bubbles S meets. The number of bubbles the resulting surface meets equals the number of such arcs times the number of crossings in the corresponding twist region.

Up to isotopy, we may also assume that S intersects $U \times \left\{\frac{1}{2}\right\}$ transversely. Hence, $S \cap \left(U \times \left\{\frac{1}{2}\right\}\right)$ is a collection of simple closed curves and arcs. By pushing S outwards towards the boundary of the disk U, one can assume that each component of $S \cap \left(U \times \left\{\frac{1}{2}\right\}\right)$ is of the following form:

(0) An arc connecting the boundary of the disk U to itself separating the holes, and intersecting α once.

(1) An arc connecting a hole to the boundary of the disk and not intersecting α.

(2) An arc connecting the two holes and not intersecting α.

Thus, $S \cap \left(U \times \left[\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon\right]\right)$ is a collection of disks of the form $\alpha \times \left[\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon\right]$, where α is an arc of type (0), (1) or (2) as stated. By an ambient isotopy, we can stretch the slab $U \times \left[\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon\right]$ to $U \times [0,1] = T$. The number of bubbles the resulting surface meets equals the number of arcs of type (0) times the number of twist in the twist region. The arcs of type (0) are in one-to-one correspondence with the arcs of $S \cap E$. Note that the fact that $\iota \colon \partial S \to L$ is a covering map was not affected by the isotopies above. □

Definition 3.2 A surface $S \subset \mathbb{S}^3 \smallsetminus \mathcal{N}(L)$ is *in normal position* if it intersects the planes $P^{\pm}$ transversely and the map $\partial S \to L$ that is obtained by shrinking $\mathcal{N}(L)$ radially is a covering map onto its image. In particular, S has no meridional boundary components.

Lemma 3.3 *Let $S \subset \mathbb{S}^3 \smallsetminus \mathcal{N}(L)$ be a surface in normal position, and let $(\boldsymbol{T}, t)$ be a twist box. Then, up to isotopy, each component of the intersection $S \cap \boldsymbol{T} \cap P^{\pm}$ looks as in Figure 3.* □

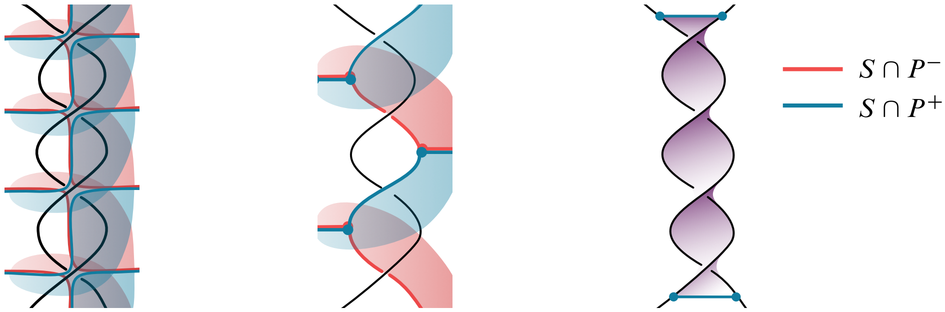


Figure 3: The possible three types of intersection of S with a twist box.

3.2 Curves of intersection

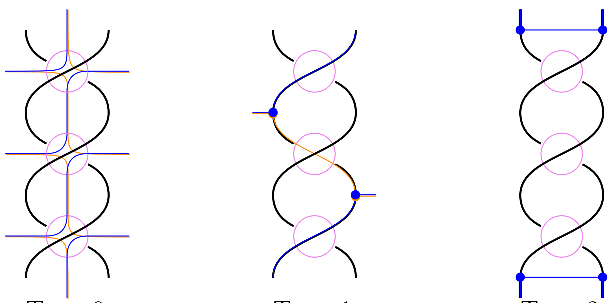
Let $S \subseteq \mathbb{S}^3 \setminus \mathcal{N}(L)$ be a surface in normal position. We would like to study the surface S through its curves of intersection with the planes $P^{\pm}$.

Let $\mathcal{T}$ be the union of all twist boxes of L. Consider the collection of disks $\mathcal{D}$ of Type 2 which occur as intersections $S \cap \mathcal{T}$. We may assume that $\partial\mathcal{D} \subset P \cup L$, and that the subsurface $\widehat{S} = S \setminus \mathcal{D}$ is transversal to $P^{\pm}$.

Recall the map $\iota\colon (S, \partial S) \to (\mathbb{S}^3, L)$. Define $\mathcal{C}^+ = \partial\iota^{-1}(\widehat{S} \cap H^+)$ and $\mathcal{C}^- = \partial\iota^{-1}(\widehat{S} \cap H^-)$. Now define $\mathcal{C} = \mathcal{C}^+ \cup \mathcal{C}^-$. As each of $P^{\pm}$ is a 2-sphere, $\widehat{S} \cap H^{\pm}$ is a collection of subsurfaces of $\widehat{S}$, the boundary of which are simple closed curves $c \subset S$. For $c \in \mathcal{C}^+$, denote by S_c the component of $\widehat{S} \cap H^+$ so that $c \subset \partial S_c$, and respectively for $c \in \mathcal{C}^-$.

We think of curves in $\mathcal{C}^{\pm}$ as curves on $P^{\pm}$, as they are disjoint outside L. Here, and in most of the figures in the remainder of this paper, curves in $\mathcal{C}^+$ are colored blue while curves in $\mathcal{C}^-$ are colored orange.

Assume S is in normal position, and let $\boldsymbol{T}$ be a twist box. The curves of intersection $\mathcal{C}$ of S which meet a connected component of $\iota(S) \cap \boldsymbol{T}$ must meet the corresponding twist region T in one of the following three configurations:



In order to analyze the curves $c \in \mathcal{C}$ we need to consider specific subarcs and points of c which we define next.

Definition 3.4 For a curve $c \in \mathcal{C}$, we define the following arcs and points (they are illustrated in the figures following the definition):

(1) An $\circ$-*joint* ("interior-joint") of c is a subarc of c which is a connected component of $c \cap \iota^{-1}(B)$ for some bubble B. The number of $\circ$-joints of c is denoted $\mathfrak{J}_\circ(c)$.

(2) A ∂-*joint* ("boundary-joint") of c is an endpoint of a connected component of $c \cap \partial S$. The number of ∂-joints of c is denoted $\mathfrak{J}_\partial(c)$

(3) A *joint* of c is an $\circ$-joint or a ∂-joint of c. The number of joints of c is denoted by $\mathfrak{J}(c) = \mathfrak{J}_\circ(c) + \mathfrak{J}_\partial(c)$.

(4) Define $\mathcal{C}_{i,j} = \{c \in \mathcal{C} \mid \mathfrak{J}_\circ(c) = i,\ \mathfrak{J}_\partial(c) = j\}$.

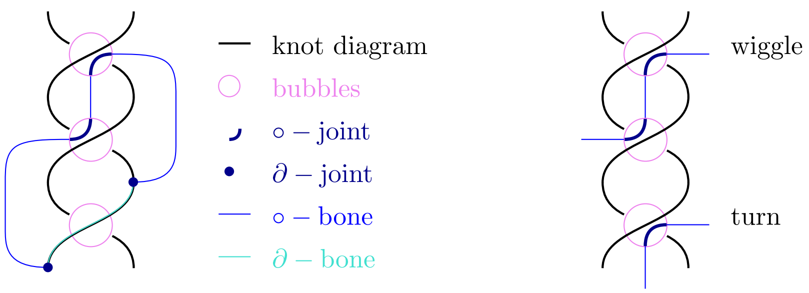
(5) A *bone* of c is a connected component of c minus its joints. Note that all bones are arcs which are mapped by ι to P.

(6) A ∂-*bone* of c is a bone which is contained in ∂S. Note that the endpoints of ∂-bones are ∂-joints. All other bones are $\circ$-*bones*.

(7) A *limb* of c is a subarc $\alpha \subset c$ with endpoints in the interiors of bones. Two limbs are equal if there is an isotopy of limbs (in c) between them. In particular, their endpoints lie in the interior of the same bones. The quantities $\mathfrak{J}(\alpha)$, $\mathfrak{J}_\circ(\alpha)$ and $\mathfrak{J}_\partial(\alpha)$ are defined as for curves.

(8) A *turn* of c is a limb of c that contains exactly one joint, this joint is an $\circ$-joint and the endpoints of the limb are outside twist regions. A curve *turns* at a twist region if it contains a turn in that region.

(9) A *wiggle* of c is a limb of c that contains exactly two joints, these joints are $\circ$-joints through consecutive bubbles of a twist box, and the endpoints of the limb are outside the twist regions. A curve *wiggles* through a twist region if it contains a wiggle in that region.



(10) Let $\mathcal{B}$ be a 3-ball bounded by a bubble B, then the components of $S \cap \mathcal{B}$ are *saddles*. Those are disks whose boundary is $\iota(\alpha_1^+ \cup \alpha_2^+ \cup \alpha_1^- \cup \alpha_2^-)$, where $\alpha_i^\pm$ are $\circ$-joints of curves in $\mathcal{C}^\pm$, respectively. The two $\circ$-joints α_1^+, α_2^+ (and the two $\circ$-joints α_1^-, α_2^-) are said to be *opposite*. See Figure 4.

Remark 3.5 Note that a ∂-joint connects a ∂-bone and $\circ$-bone, while a $\circ$-joint connects two $\circ$-bones. The ∂-joints and ∂-bones are contained in the boundary of S, while $\circ$-joints and $\circ$-bones are contained in the interior of S.

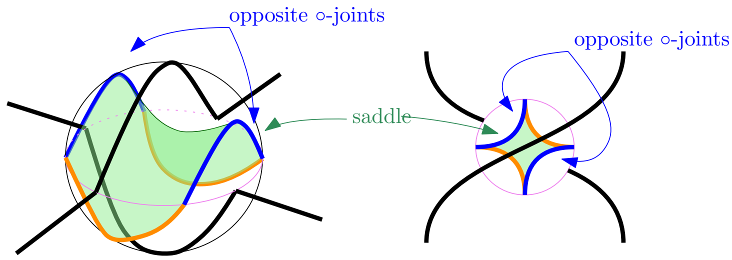
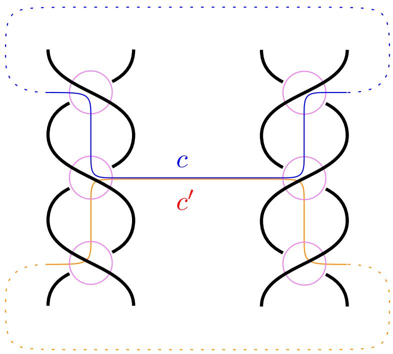


Figure 4: A saddle of S in a bubble viewed from an angle and from the top.

Definition 3.6 Two curves (or limbs of curves) $c, c' \in \mathcal{C}$ are *abutting* if they share an $\circ$-bone and $c \neq c'$. Necessarily, if $c \in \mathcal{C}^+$ then $c' \in \mathcal{C}^-$ and vice versa.

The following figure shows an example of two abutting curves c and c':



3.3 Taut surfaces

Definition 3.7 Given an incompressible surface $S \subset \mathbb{S}^3 \setminus \mathcal{N}(L)$ we define a *lexicographic complexity* of S as

$$\mathrm{Com}(S) = \Big(\sum_{c \in \mathcal{C}} \mathfrak{J}_\circ(c), \sum_{c \in \mathcal{C}} \mathfrak{J}_\partial(c), |\mathcal{C}| \Big). \tag{3-1}$$

Recall that a properly embedded surface S in a 3-manifold M is called *essential* if it is either a 2-sphere which does not bound a 3-ball, or it is incompressible, boundary incompressible and not boundary parallel.

Definition 3.8 Let $S \subset \mathbb{S}^3 \setminus \mathcal{N}(L)$ be an essential surface in normal position. The surface S is *taut* if either

(i) S is an essential 2-sphere, and S minimizes complexity among all essential 2-spheres, or

(ii) S is not a 2-sphere, the link L is not split (ie $\mathbb{S}^3 \setminus L$ is irreducible), and S minimizes complexity in its isotopy class.

The next lemma shows that the intersection curves of taut surfaces must have certain properties.

Lemma 3.9 *Assume that the diagram $D(L)$ is connected. Let $S \subset \mathbb{S}^3 \smallsetminus \mathcal{N}(L)$ be a taut surface. Then, for all $c \in \mathcal{C}$ we have:*

(1) *S_c is a disk.*

(2) *$\mathfrak{J}_\partial(c)$ is even.*

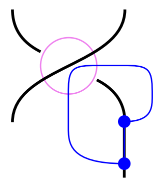
(3) *If $\mathfrak{J}_\partial(c) \le 2$ then $\mathfrak{J}_\circ(c) > 0$.*

(4) *If $\mathfrak{J}_\partial(c) = 0$ then $\mathfrak{J}_\circ(c)$ is even.*

(5) *If a curve c meets a bubble B more than once, then it does so in two opposite $\circ$-joints.*

(6) *If a curve c has two ∂-joints on a connected component of $P^+ \cap L$ (or $P^- \cap L$), then they are the endpoints of a ∂-bone. Moreover, the two $\circ$-bones incident to them are in different regions of $P \smallsetminus D(L)$.*

(7) *The curve c is not a curve in $\mathcal{C}_{1,2}$ bounding on $P^\pm$ exactly one component of $L \cap P^\pm$ as depicted here:*



Proof Let $S \subset \mathbb{S}^3 \smallsetminus \mathcal{N}(L)$ be an essential surface satisfying (i) or (ii). Note that in both cases, compressing along a disk $D \subset \mathbb{S}^3 \smallsetminus \mathcal{N}(L)$ with $D \cap S = \partial D$ results either in two essential spheres, or a surface in the same isotopy class of S. Thus, by the assumption on S, surfaces obtained by such a compression cannot have lower complexity.

(1) Since S is essential, each subsurface S_c must be planar, as otherwise it contains a nontrivial compression disk. If S_c has more than one boundary component then compressing along a disk in H^+ or H^- whose boundary separates boundary components of S_c will result in a surface with fewer intersections with P in contradiction to the choice of S.

(2) By definition, $\mathfrak{J}_\partial(c)$ is the number of endpoints of arcs in $c \smallsetminus L$. Since each arc has two endpoints, $\mathfrak{J}_\partial(c)$ is even.

(3) By (2), $\mathfrak{J}_\partial(c)$ is either two or zero. If $\mathfrak{J}_\partial(c) = 0$ and $\mathfrak{J}_\circ(c) = 0$ then, since $D(L)$ is connected, c bounds a disk on $P \smallsetminus L$. Compressing S along this disk reduces the number of intersections with P.

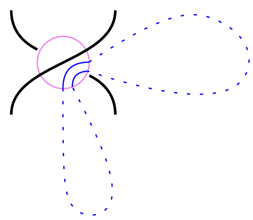
If $\mathfrak{J}_\partial(c) = 2$ and $\mathfrak{J}_\circ(c) = 0$ then c bounds a disk D in P such that $\partial D = \alpha \cup \beta$, where α is a ∂-bone of c and β is an $\circ$-bone of c. However, this is impossible by (2).

(4) The diagram $D(L)$ is a 4-regular graph, and thus it partitions P into regions which can be given a checkerboard coloring, ie can be colored black and white so that two adjacent regions are colored in different colors. Consider the colors of complementary regions of $P \smallsetminus \mathcal{T}$ which the curve c intersects. If $\mathfrak{J}_\partial(c) = 0$ every change of colors, of these regions along c, accounts for one bubble that c meets. Since c is a closed curve the total number of color changes is even, and correspondingly $\mathfrak{J}_\circ(c)$ is even.

(5) Without loss of generality assume that $c \in \mathcal{C}^+$. Each time c meets a bubble B it does so along a ∂-bone or an $\circ$-joint. Since c meets B twice then it does so along either two ∂-bones, an $\circ$-joint and a ∂-bone or two $\circ$-joints. If c meets B in two ∂-bones, then the disk S_c contains an arc connecting the two ∂-bones. This arc together with an arc on $\partial\mathcal{N}(L)$ bounds (by an innermost argument) a compression disk for S. Compressing along this disk reduces the complexity of S.

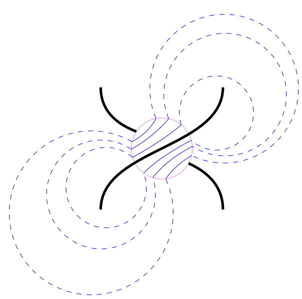
If c meets B in an $\circ$-joint and a ∂-bone, then the disk S_c contains an arc connecting the $\circ$-joint and the ∂-bone. This arc, together with an arc on B bounds a disk. Isotoping S through this disk reduces the complexity of S.

Thus, c meets B in two $\circ$-joints. Assume that c has two $\circ$-joints in B on the same side of L as in



The isotopy of S defined in the proof of Lemma 1(ii) of [11] (for the case $R \cap L = \varnothing$ in his notation) reduces our complexity since the number $\sum_{c\in\mathcal{C}} \mathfrak{J}_\circ(c)$ strictly decreases.

By Lemma 1(ii) of [11], c does not have It follows that c has at most two $\circ$-joints in the same bubble, and they are separated by L in B^+. The number of components of $S \cap B^+$ separating each of the $\circ$-joints of c from L is the same: Each component of $S \cap B$ separating an $\circ$-joint of c and L belongs to a curve c' in $\mathcal{C}^+$. As curves in $\mathcal{C}^+$ do not intersect, in order to close up, c' has to return to B on the other side of L between L and the other $\circ$-joint of c in B. This is depicted here:

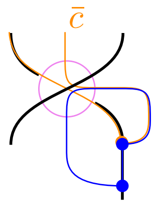


The intersection of S with the ball bounded by B, is a finite collection of stacked saddles, it follows that the $\circ$-joints of c belong to the same saddle.

(6) If c has two ∂-joints on the same component of $P^+ \cap L$. Let α in $P^+ \cap L$ and β in S_c be arcs connecting the two ∂-joints. Unless α is a ∂-bone of c, by compressing along the disk bounded by $\alpha \cup \beta$ (using an innermost such disk) complexity is reduced because $\sum_{c\in\mathcal{C}} \mathfrak{J}_\partial(c)$ strictly decreases while $\sum_{c\in\mathcal{C}} \mathfrak{J}_\circ(c)$ does not increase.

If c contains a ∂-bone α such that the two adjacent $\circ$-joints are in the same region of $P \smallsetminus D(L)$ then, by pushing S through P in a neighborhood of α, we reduce the number of intersection points by 2, in contradiction to the minimal complexity of S.

(7) Assume in contradiction that c is such a curve, ie as the blue curve here:



By (6), any curve contained in c has to be of a similar form. Assume c is an innermost such curve. The curve $\bar{c}$ abutting c, as depicted in the following figure, meets the bubble twice, at a ∂-bone and an $\circ$-joint, in contradiction to (5). □

Remark 3.10 Note that if S is taut then $\mathcal{C}_{0,0} = \mathcal{C}_{0,2} = \mathcal{C}_{i,2k+1} = \mathcal{C}_{2k+1,0} = \varnothing$ for all $i, k \in \mathbb{N} \cup \{0\}$.

4 Euler characteristic and curves of intersection

From now on we assume that the surface S is taut.

4.1 Distributing Euler characteristic among curves

For each curve $c \in \mathcal{C}$ we will define the *contribution* of c, and show that the Euler characteristic of S can be computed by summing up the contributions of curves $c \in \mathcal{C}$.

Definition 4.1 The *contribution* $\chi_+(c)$ of a curve $c \in \mathcal{C}$ is defined by

$$\chi_+(c) = 1 - \tfrac{1}{4}\mathfrak{J}(c).$$

Lemma 4.2 *If $S \subset \mathbb{S}^3 \smallsetminus \mathcal{N}(L)$ is taut then $\chi(S) = \sum_{c \in \mathcal{C}} \chi_+(c)$.*

Proof The union of the collection of all the curves $c \in \mathcal{C}$ on S is an embedded graph $\widehat{X}$ of S. The vertices $\widehat{X}^0$ of the graph $\widehat{X}$ are the ∂-joints and the endpoints of $\circ$-joints. The edges $\widehat{X}^1$ of the graph $\widehat{X}$ are the bones and the $\circ$-joints. The graph $\widehat{X}$ partitions S into disk regions of three types:

(1) subsurfaces $S_c \subseteq \widehat{S} \cap H^{\pm}$ for $c \in \mathcal{C}^{\pm}$,

(2) saddles $R \subseteq \widehat{S} \cap \mathcal{B}$ where $\mathcal{B}$ is a 3-ball bounded by a bubble, or

(3) regions $D \subset S$ corresponding to Type 2 disks.

In case (3), the regions D are disks whose boundary consists of two arcs on L and two edges of $\widehat{X}$. By collapsing each such disk D to one of the edges in $\widehat{X}$ we get a homotopic surface. By abuse of notation, we call it S, and call the corresponding graph X obtained from $\widehat{X}$ by gluing together pairs of $\circ$-bones. Note that in the new surface, $\partial S \subset X$ and consists of circles comprised of ∂-bones and ∂-joints. Moreover, along every $\circ$-bone there are two abutting curves. It follows that

$$\begin{aligned}\chi(S) &= \chi(X) + \sum_{S' \subseteq \widehat{S} \cap H^{\pm}} \chi(S') + \sum_{R \subseteq \widehat{S} \cap B} \chi(R) \\ &= |X^0| - |X^1| + \sum_{S' \subseteq \widehat{S} \cap H^{\pm}} \chi(S') + \sum_{R \subseteq \widehat{S} \cap B} \chi(R).\end{aligned} \tag{4-1}$$

We compute how each $c \in \mathcal{C}$ contributes to each of the summands in (4-1):

The vertices of X Every curve $c \in \mathcal{C}$ passes through $2\mathfrak{J}_\circ(c)$ vertices of X^0 in the interior of S (those are the endpoints of $\circ$-joints it passes). Furthermore, it goes through $\mathfrak{J}_\partial(c)$ vertices of X^0 in ∂S. Each of these vertices belongs to two (abutting) curves $c \in \mathcal{C}$. Hence,

$$|X^0| = \sum_{c \in \mathcal{C}} \big(\mathfrak{J}_\circ(c) + \tfrac{1}{2}\mathfrak{J}_\partial(c)\big) \tag{4-2}$$

The edges of X Every curve $c \in \mathcal{C}$ passes through $2\mathfrak{J}_\circ(c) + \mathfrak{J}_\partial(c)$ edges in X^1. Note that every $\circ$-joint edge and every ∂-bone edge belongs to exactly one curve in $\mathcal{C}$, while each $\circ$-bone edge belongs to two curves in $\mathcal{C}$. Each $\circ$-joint edge appears in exactly one curve c and is counted once in $\mathfrak{J}_\circ(c)$. Hence, the number of $\circ$-joint edges is $\sum_{c \in \mathcal{C}} \mathfrak{J}_\circ(c)$. Similarly each ∂-bone edge appears in exactly one curve c and is counted twice in $\mathfrak{J}_\partial(c)$. Hence, the number of ∂-bone edges is $\sum_{c \in \mathcal{C}} \frac{1}{2}\mathfrak{J}_\partial(c)$. Finally, each $\circ$-bone edge in c accounts for two vertices in X^0. So the number of $\circ$-bone edges is equal to

$$\tfrac{1}{2}|X^0| = \tfrac{1}{2}\bigg(\sum_{c \in \mathcal{C}} \big(\mathfrak{J}_\circ(c) + \tfrac{1}{2}\mathfrak{J}_\partial(c)\big)\bigg).$$

Adding these contributions together gives

$$|X^1| = \sum_{c \in \mathcal{C}} \big(\tfrac{3}{2}\mathfrak{J}_\circ(c) + \tfrac{3}{4}\mathfrak{J}_\partial(c)\big). \tag{4-3}$$

Regions $S' \subset \widehat{S} \cap H^{\pm}$ To every curve $c \in \mathcal{C}$ there is a disk $S_c \subseteq \widehat{S} \cap H^{\pm}$. Thus,

$$\sum_{S' \subseteq \widehat{S} \cap H^{\pm}} \chi(S') = \sum_{c \in \mathcal{C}} 1. \tag{4-4}$$

Saddle regions $R \subset \widehat{S} \cap \mathcal{B}$ Each $\circ$-joint of a curve $c \in \mathcal{C}$ belongs to the boundary such a region. And so each curve passes through the boundary of $\mathfrak{J}_\circ(c)$ such regions. As each saddle region has four $\circ$-joints in its boundary, we have

$$\sum_{R \subseteq \widehat{S} \cap B} \chi(R) = \sum_{c \in \mathcal{C}} \tfrac{1}{4}\mathfrak{J}_\circ(c). \tag{4-5}$$

Summing over all of the above we get,

$$\begin{aligned}\chi(S) &= |X^0| - |X^1| + \sum_{S' \subseteq \widehat{S} \cap H^{\pm}} \chi(S') + \sum_{R \subseteq \widehat{S} \cap B} \chi(R) \\ &= \sum_{c \in \mathcal{C}} \big(1 - \tfrac{1}{4}(\mathfrak{J}_{\circ}(c) + \mathfrak{J}_{\partial}(c))\big) \\ &= \sum_{c \in \mathcal{C}} \chi_+(c). \end{aligned}$$ □

5 Redistribution of Euler characteristic

Standing assumption Throughout the rest of the paper, we assume that the diagram $D(L)$ is connected, prime, twist-reduced, 3-highly twisted and contains at least two twist regions.

In this section we redistribute the positive contribution of the Euler characteristic of curves, χ_+, so that after the redistribution each curve's contribution is nonpositive.

We first characterize the curves of intersection that have a positive χ_+. The characterization is done in the following lemma:

Lemma 5.1 *Let $c \in \mathcal{C}$ so that $\chi_+(c) > 0$ (ie $\mathfrak{J}(c) < 4$) then $c \in \mathcal{C}_{2,0}$ or $c \in \mathcal{C}_{1,2}$ and it is one of the six forms of Figure 5 (up to isotopy).*

Note that the curves c in cases (i) and (ii) are in $\mathcal{C}_{2,0}$. The curves in cases (iii), (iv), (v) and (vi) belong to $\mathcal{C}_{1,2}$.

Proof It follows from Definition 4.1, Lemma 3.9 and Remark 3.10 that if $\chi_+(c) > 0$ then $c \in \mathcal{C}_{2,0}$ or $c \in \mathcal{C}_{1,2}$. Since the diagram is prime, a curve $c \in \mathcal{C}_{2,0}$ must contain two turns at different twist boxes. Hence c is as depicted in figures (i) or (ii).

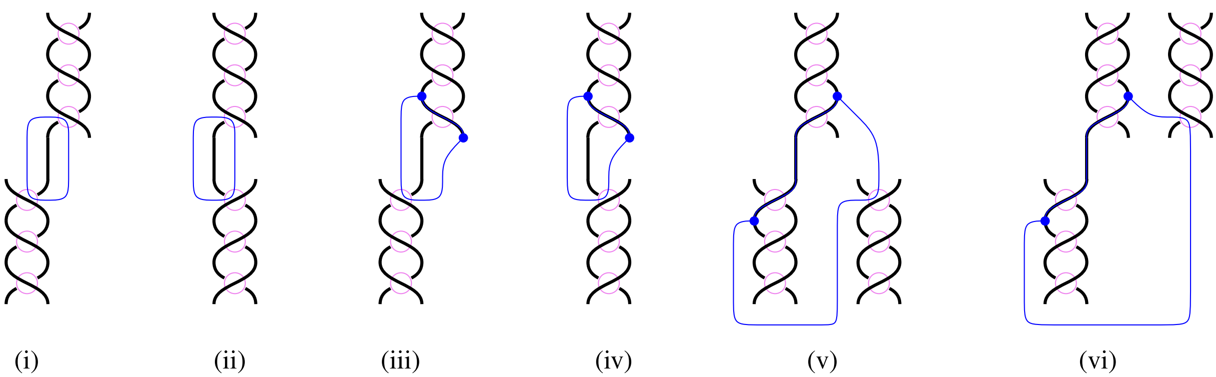


Figure 5: The six possibilities for a curve in $c \in \mathcal{C}_{2,0} \cup \mathcal{C}_{1,2}$.

Let $c \in \mathcal{C}_{1,2}$ and let α denote a limb which is a small extension of the unique ∂-bone in c. The endpoints of α must be in regions of $P \smallsetminus D(L)$ of different color since the complementary limb of α in c is a turn. Since S is taut, Lemma 3.9(7) implies that α must pass over at least one crossing of $D(L)$. Since the endpoints of α are in regions of different colors, α cannot connect the two regions adjacent to the two length edges of a twist region. Thus, α enters a twist region T through its length edge, and exists on L. It can meet one or two twist regions. If α meets one twist region, then it must be as in figures (iii) or (iv). Otherwise, up to isotopy, it must be as in figures (v) or (vi). □

Definition 5.2 Denote by $\mathcal{C}_{>0}$ (resp. $\mathcal{C}_{\leq 0}$) the set of curves $c \in \mathcal{C}$ such that $\chi_+(c) > 0$ (resp. ≤ 0). The lemma above shows that $\mathcal{C}_{>0} = \mathcal{C}_{2,0} \cup \mathcal{C}_{1,2}$. The *type* of a curve in $\mathcal{C}_{>0}$ corresponds to the types of curves as depicted in Figure 5. For example, a curve of $\mathcal{C}_{2,0}$ is of type (i) or (ii).

Next, we will describe a distinguished set, denoted by $\mathcal{K}$, of limbs of curves in $\mathcal{C}$ to which we will "reallocate" some of the positive contribution of curves in $\mathcal{C}_{>0}$. We begin with a definition:

Definition 5.3 An *extremal* bubble is a first or last bubble of a twist region. A curve wiggles *extremally* if it wiggles through an extremal bubble. Assume that a curve or an arc β wiggles through a twist region extremally, then the $\circ$-bone $\alpha \subset \beta$ which leaves the twist region from the extremal bubble of the wiggle is called a *core* of β.

Definition 5.4 A *vertebra* is an $\circ$-bone τ in a curve c connecting two turns of c in two twist regions T, T' so that τ meets the length edge of T and the width edge of T'.

A *rib* is a closed curve $c \in \mathcal{C}_{4,0}$ which consists of exactly two turns and an extremal wiggle.

Remark 5.5 Note that the two $\circ$-bones of a curve $c \in \mathcal{C}_{2,0}$ of type (i) are vertebrae.

Lemma 5.6 *Assume that c_0 is not a rib, and that c_0 contains a vertebra τ_0. Then, there exists a finite sequence of curves $c_0, c_1, \dots, c_n$, limbs $\kappa_1, \dots, \kappa_n$ and bones $\tau_0, \dots, \tau_{n-1}$ such that:*

(1) *For $1 \leq i < n$, c_i is a rib and τ_i is the $\circ$-bone connecting its two turns.*
(2) *For $1 \leq i \leq n$, κ_i is a limb of c_i with a unique core τ_{i-1} and $\mathfrak{J}(\kappa_i) = 3$. In particular, the curve c_i abuts the curve c_{i-1} along the core τ_{i-1}.*
(3) *The curve c_n has $\chi_+(c_n) < 0$.*

Moreover, given κ_n one can uniquely determine the $\circ$-bone τ_0 and hence the curves c_i, arcs κ_i and bones τ_i as above.

Definition 5.7 We will refer to the curves c_i (resp. limbs κ_i) in the lemma as the *layers curves* (resp. *layer limbs*) of τ_0, and to the curve c_n (resp. arc κ_n) as the *terminal layer curve* (resp. *terminal layer limb*) of τ_0.

Proof of Lemma 5.6 We produce a sequence of curves, limbs and bones, satisfying the assumptions above, which terminates at the first curve c_n such that $\chi_+(c_n) < 0$.

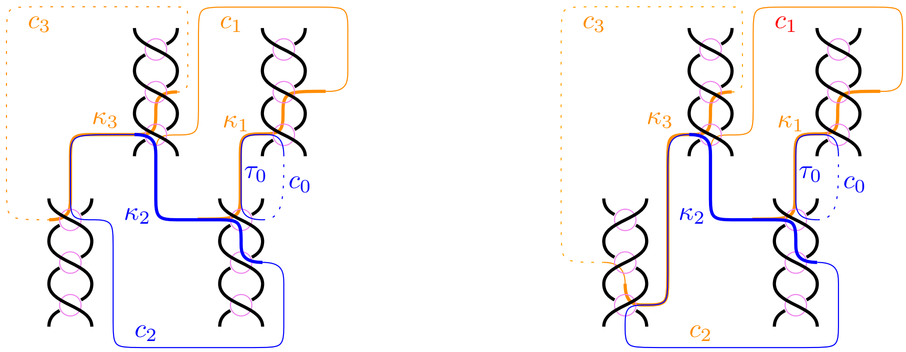


Figure 6: Example 5.8.

Let c_0 and τ_0 be as in the statement of the lemma. Let c_1 be the curve abutting c_0 along τ_0. The bone τ_0 connects an extremal wiggle and a turn of c_1, hence it is a core of c_1. Let κ_1 be the limb of c_1 containing τ_0 and the adjacent wiggle and turn. If $\chi_+(c_1) < 0$, then stop the process. Otherwise, $\mathfrak{J}(c_1) = 4$. Hence, by Lemma 3.9, the curve c_1 is a rib, ie it consists of a wiggle and two turns and has a unique core. Let α_1, α_1' be the $\circ$-joint of the two turns of c_1, and assume that $\alpha_1 \subset \kappa_1$. Let τ_1 be the $\circ$-bone of c_1 connecting α_1, α_1'. The bone τ_1 meets the length edge of the twist region containing α_1.

Assume first that τ_1 is not a vertebra of c_1, ie it meets the length edge of the twist region containing α_1'. Then, the curve c_2 abutting c_1 along τ_1 contains two wiggles in two different twist regions. It follows that $\chi_+(c_2) < 0$ as otherwise c_2 bounds a twist reduction subdiagram. Let κ_2 be the limb in c_2, abutting τ_1, consisting of a wiggle through the bubble of α_1, and one more $\circ$-joint at the bubble containing α_1'. The bone τ_1 is the unique core of the limb κ_2, and the process stops ($n = 2$).

If τ_1 is a vertebra, then we iterate the process. That is, we consider the curve c_2 abutting c_1 along τ_1, and the limb κ_2 of c_2 containing τ_1 and its adjacent wiggle and turn.

Since there are finitely many curves and limbs, the process either terminates or is periodic. It cannot be periodic because the initial curve c_0 is not a rib, but note that all the curves c_i for $i < n$ are ribs.

Finally, given κ_n, the curve c_{n-1} is the curve abutting c_n along the unique core of κ_n. The curve c_{n-1} has a unique core, which is also the core of a unique arc κ_{n-1} (with $\mathfrak{J}(\kappa_{n-1}) = 3$). Repeating this process, we can retrace the sequence all the way to τ_0. □

Example 5.8 In Figure 6 we see two examples of outputs of the process in Lemma 5.6. Starting with the curve c_0 which is not a rib, and the vertebra τ_0 of c_0, we get the curves c_0, c_1, c_2, c_3, limbs $\kappa_1, \kappa_2, \kappa_3$ and $\circ$-bones τ_0, τ_1, τ_2. The limbs κ_i are shown in bold the figure. The $\circ$-bones τ_0, τ_1, τ_2 are the cores of $\kappa_1, \kappa_2, \kappa_3$ respectively.

In Figure 6, left, note that the $\circ$-bone τ_2 is a vertebra. If $\chi_+(c_3) < 0$ then the process stops at c_3 and κ_3 is its terminal limb. Otherwise, c_3 is again a rib, and the process continues.

In Figure 6, right, the $\circ$-bone connecting the two turns of c_2 is not a vertebra, as it meets the length edge of both twist regions. Therefore the limb κ_3, which is shown in bold in the figure, contains a wiggle (through the bubble in which both c_1 and c_2 turn) and "half a wiggle" (through the other bubble in which c_2 turns). As the proof shows, the curve c_3 necessarily satisfies $\chi_+(c_3) < 0$, and thus it is the terminal curve of the process.

Definition 5.9 (definition of $\mathcal{K}$) (1) Each $c \in \mathcal{C}_{2,0}$ of type (i) has two $\circ$-bones. Each $\circ$-bone τ of c is a vertebra and hence by Lemma 5.6 determines a terminal layer limb κ_n. Let $\mathcal{K}_{3,0}$ be the set of all terminal layer limbs associated with all $\circ$-bones of curves in $c \in \mathcal{C}_{2,0}$.

(2) Each $c \in \mathcal{C}_{2,0}$ of type (ii) determines an arc κ which abuts c and wiggles through the two twist regions. Let $\mathcal{K}_{4,0}$ be the collection of all arcs κ obtained in this way.

(3) Each $c \in \mathcal{C}_{2,1}$, determines an arc κ which abuts c, wiggles through the twist region in which c turns, and contains one of the ∂-joints of c. Let $\mathcal{K}_{2,1}$ be the collection of all arcs κ obtained in this way.

Finally, let $\mathcal{K}$ be the set $\mathcal{K}_{3,0} \cup \mathcal{K}_{4,0} \cup \mathcal{K}_{2,1}$.

Lemma 5.10 *Given $\kappa \in \mathcal{K}$ one can uniquely determine the curve $c \in \mathcal{C}_{>0}$ that determines it. Conversely, to each curve in $\mathcal{C}_{2,0}$ of type* (i) *there are two curves of $\mathcal{K}_{3,0}$ corresponding to the two choices of $\circ$-bones of c. To each curve in $\mathcal{C}_{1,2}$ and each curve in $\mathcal{C}_{2,0}$ of type* (ii) *there is a unique arc in $\mathcal{K}_{2,1}$ and $\mathcal{K}_{4,0}$ respectively.*

In particular,

$$\tfrac{1}{2}|\mathcal{K}_{3,0}| + |\mathcal{K}_{4,0}| = |\mathcal{C}_{2,0}| \quad \textit{and} \quad |\mathcal{K}_{2,1}| = |\mathcal{C}_{1,2}|.$$

Proof If $\kappa \in \mathcal{K}_{4,0}$ (resp. $\kappa \in \mathcal{K}_{2,1}$) then the curve $c \in \mathcal{C}_{2,0}$ of type (ii) (resp. $c \in \mathcal{C}_{1,2}$) that determines it is the curve abutting κ. If $\kappa \in \mathcal{K}_{3,0}$ then by Lemma 5.6 there is a unique $\circ$-bone τ of a curve $\mathcal{C}_{2,0}$ of type (i) that determines it. The converse statements follow. □

Lemma 5.11 *The arcs in $\mathcal{K}$ which belong to curves in $\mathcal{C}^+$* (*resp. $\mathcal{C}^-$*) *are pairwise disjoint.*

Proof Assume that the two arcs $\kappa, \kappa' \in \mathcal{K}$ meet. Since, by Lemma 5.10, every arc in $\mathcal{K}$ is determined by its core, it suffices to show that κ and κ' have the same core.

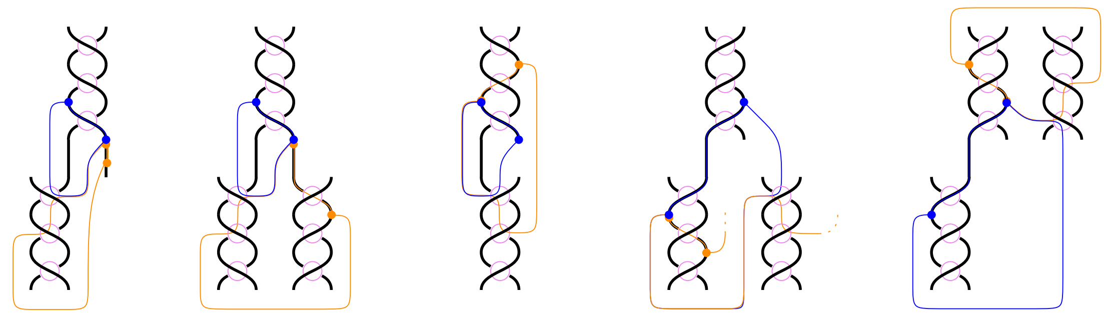
By assumption κ, κ' share a joint, and hence are limbs of the same curve c. If this joint is a ∂-joint then $\kappa, \kappa' \in \mathcal{K}_{2,1}$ and their core is the unique $\circ$-bone incident to their shared ∂-joint. If the joint is an $\circ$-joint that is part of a wiggle of c, then, since the diagram is 3-highly twisted, there is a unique core emanating from the extremal bubble of this wiggle, which is shared by both κ and κ'. Finally, if the joint is an $\circ$-joint that is part of a turn of c, then $\kappa, \kappa' \in \mathcal{K}_{3,0}$. The cores of κ, κ' must be the unique $\circ$-bone which emanates from the width edge of the twist region in which the turn occurs because it is also the vertebra of the previous layer curve. □

Lemma 5.12 *For all $\kappa \in \mathcal{K}$ we have $\mathfrak{J}(c) \geq \mathfrak{J}(\kappa) + 2$, where c is the unique curve in $\mathcal{C}$ containing κ.*

Proof Let $\kappa \in \mathcal{K}$ and let $c \in \mathcal{C}$ be the curve containing κ. Assume for contradiction that $\mathfrak{J}(c) < \mathfrak{J}(\kappa) + 2$.

If $\kappa \in \mathcal{K}_{4,0}$ then $\mathfrak{J}(c) < \mathfrak{J}(\kappa) + 2$ implies that $c \in \mathcal{C}_{4,0}$. If this is the case, then since c wiggles through two twist regions, the projection of c to P gives a twist-reduction subdiagram in contradiction to the assumption on the diagram.

If $\kappa \in \mathcal{K}_{2,1}$, then $\mathfrak{J}(c) < \mathfrak{J}(\kappa) + 2$ implies that $c \in \mathcal{C}_{2,2}$. If this is the case, then κ abuts some curve $c_0 \in \mathcal{C}_{1,2}$ which could be one of the figures (iii)–(vi) in Lemma 5.1. For case (iii), there are two possible configurations for the closed curve $c \in \mathcal{C}_{2,2}$ containing κ, while for each of the cases (iv)–(vi) there is only one possible such curve. Thus, all possible cases for the curve c are shown here:



(iii) (iii′) (iv) (v) (vi)

In case (iii) the surface is not taut and in all other cases, the curve c bounds a twist-reduction subdiagram.

If $\kappa \in \mathcal{K}_{3,0}$, then $\mathfrak{J}(c) < \mathfrak{J}(\kappa) + 2$ implies that $c \in \mathcal{C}_{4,0}$, however this is in contradiction to the definition of $\mathcal{K}_{3,0}$ given by Lemma 5.6. □

Remark 5.13 It follows from the proof of Lemma 5.6 that if c contains a limb $\kappa \in \mathcal{K}_{3,0}$ such that the core of τ is not a vertebra of its abutting curve then $\mathfrak{J}(c) \geq \mathfrak{J}(k) + 3$: Indeed, if $c = c_n$, $\kappa = \kappa_n$ and κ abuts c_{n-1} along a bone which is not a vertebra of c_{n-1} then c must have an additional $\circ$-joint which is not contained in κ.

Definition 5.14 Let $c \in \mathcal{C}$. If $c \in \mathcal{C}_{>0}$ define $\chi'(c) = 0$. Otherwise, let $n_{3,0}$ (resp. $n_{4,0}$; $n_{2,1}$) be the number of limbs $\kappa \in \mathcal{K}_{3,0}$ (resp. $\mathcal{K}_{4,0}$; $\mathcal{K}_{2,1}$) in c. We associate to c the quantity

$$\chi'(c) = \chi_+(c) + \tfrac{1}{4} n_{3,0} + \tfrac{1}{2} n_{4,0} + \tfrac{1}{4} n_{2,1}.$$

The next lemma shows that χ' is a redistribution of the Euler characteristic of S among curves in $\mathcal{C}_{\leq 0}$.

Lemma 5.15
$$\chi(S) = \sum_{c \in \mathcal{C}} \chi'(c).$$

Proof By Lemma 4.2, $\chi(S) = \sum_{c \in \mathcal{C}} \chi_+(c)$. Since $\chi'(c) = 0$ for $c \in \mathcal{C}_{>0}$,

$$\sum_{c \in \mathcal{C}} \chi'(c) = \sum_{c \in \mathcal{C}_{\leq 0}} \chi'(c).$$

It thus remains to prove that
$$\sum_{c\in\mathcal{C}} \chi_+(c) = \sum_{c\in\mathcal{C}_{\le 0}} \chi'(c).$$
Subtracting $\sum_{c\in\mathcal{C}_{\le 0}} \chi_+(c)$ from both sides and recalling that $\mathcal{C}_{\le 0} = \mathcal{C} \smallsetminus \mathcal{C}_{>0}$, we have to show that
$$\sum_{c\in\mathcal{C}_{>0}} \chi_+(c) = \sum_{c'\in\mathcal{C}_{\le 0}} (\chi'(c') - \chi_+(c')).$$
The left hand side is simply $\frac{1}{2}|\mathcal{C}_{2,0}| + \frac{1}{4}|\mathcal{C}_{1,2}|$ since $\mathcal{C}_{>0} = \mathcal{C}_{2,0} \cup \mathcal{C}_{1,2}$ and
$$\chi_+(c) = \begin{cases} \frac{1}{2} & \text{if } c \in \mathcal{C}_{2,0}, \\ \frac{1}{4} & \text{if } c \in \mathcal{C}_{1,2}. \end{cases}$$
By the definition of χ', the right hand side gives $\frac{1}{4}|\mathcal{K}_{3,0}| + \frac{1}{2}|\mathcal{K}_{4,0}| + \frac{1}{4}|\mathcal{K}_{2,1}|$. The proof is now complete by Lemma 5.10. □

The next lemma shows that indeed χ' is nonpositive.

Lemma 5.16 $\chi'(c) \le 0$ *for all* $c \in \mathcal{C}$.

Proof If $c \in \mathcal{C}_{>0}$ then $\chi'(c) = 0$. Let $c \in \mathcal{C}_{\le 0}$ and let $n_{3,0}, n_{4,0}, n_{2,1}$ be as in Definition 5.14 of $\chi'(c)$. By Lemma 5.11, the limbs of $\mathcal{K}$ in c are disjoint and therefore
$$\mathfrak{I}(c) \ge \sum_{c\supset\kappa\in\mathcal{K}} \mathfrak{I}(\kappa). \tag{5-1}$$
Hence,
$$\begin{aligned} \chi'(c) &= \chi_+(c) + \tfrac{1}{4}n_{3,0} + \tfrac{1}{2}n_{4,0} + \tfrac{1}{4}n_{2,1} = \big(1 - \tfrac{1}{4}\mathfrak{I}(c)\big) + \tfrac{1}{4}n_{3,0} + \tfrac{1}{2}n_{4,0} + \tfrac{1}{4}n_{2,1} \\ &\le 1 - \sum_{c\supset\kappa\in\mathcal{K}} \tfrac{1}{4}\mathfrak{I}(\kappa) + \tfrac{1}{4}n_{3,0} + \tfrac{1}{2}n_{4,0} + \tfrac{1}{4}n_{2,1} \\ &= 1 + \sum_{c\supset\kappa\in\mathcal{K}_{3,0}} \big(\tfrac{1}{4} - \tfrac{1}{4}\mathfrak{I}(\kappa)\big) + \sum_{c\supset\kappa\in\mathcal{K}_{4,0}} \big(\tfrac{1}{2} - \tfrac{1}{4}\mathfrak{I}(\kappa)\big) + \sum_{c\supset\kappa\in\mathcal{K}_{2,1}} \big(\tfrac{1}{4} - \tfrac{1}{4}\mathfrak{I}(\kappa)\big) \\ &= 1 + \sum_{c\supset\kappa\in\mathcal{K}_{3,0}} \big(\tfrac{1}{4} - \tfrac{1}{4}\cdot 3\big) + \sum_{c\supset\kappa\in\mathcal{K}_{4,0}} \big(\tfrac{1}{2} - \tfrac{1}{4}\cdot 4\big) + \sum_{c\supset\kappa\in\mathcal{K}_{2,1}} \big(\tfrac{1}{4} - \tfrac{1}{4}\cdot 3\big) \\ &= 1 - \tfrac{1}{2}(n_{3,0} + n_{4,0} + n_{2,1}). \end{aligned} \tag{5-2}$$
Now the argument is divided into cases depending on the sum $n = n_{3,0} + n_{4,0} + n_{2,1}$.

Case 0 ($n = 0$) We have $\chi'(c) = \chi_+(c)$. But since $c \in \mathcal{C}_{\le 0}$ we have $\chi_+(c) \le 0$ and we are done.

Case 1 ($n = 1$) That is, c contains a single subarc $\kappa \in \mathcal{K}_{3,0} \cup \mathcal{K}_{4,0} \cup \mathcal{K}_{2,1}$. By Lemma 5.12, $\mathfrak{I}(c) \ge \mathfrak{I}(\kappa) + 2$. If $\kappa \in \mathcal{K}_{4,0}$ we get $\mathfrak{I}(c) \ge 6$ and thus $\chi'(c) = 1 - \frac{1}{4}\mathfrak{I}(c) + \frac{1}{2} \le 0$. Similarly, if $\kappa \in \mathcal{K}_{3,0} \cup \mathcal{K}_{2,1}$ we get $\mathfrak{I}(c) \ge 5$ and thus $\chi'(c) = 1 - \frac{1}{4}\mathfrak{I}(c) + \frac{1}{4} \le 0$.

Case 2 ($n \ge 2$) In this case we are done by inequality (5-2). □

Corollary 5.17 *The link $\mathcal{L}$ is nonsplit nor the unknot.*

name of set	$\mathfrak{J}(\cdot)$	$\mathfrak{J}_\circ(\cdot)$	$\mathfrak{J}_\partial(\cdot)$	the set's composition/classification
$\mathcal{C}_{2,0}$	2	2	0	type (i) or (ii)
$\mathcal{C}_{1,2}$	3	1	2	type (iii), (iv) or (v)
$\mathcal{C}_{4,0}$	4	4	0	4 turns or 1 wiggle and 2 turns or 2 wiggles
$\mathcal{C}_{2,2}$	4	2	2	2 $\circ$-joints and 1 ∂-bone
$\mathcal{C}_{0,4}$	4	0	4	2 ∂-bones
$\mathcal{K}_{3,0}+0,2$	5	3	2	1 limb of $\mathcal{K}_{0,3}$ + ∂-bone
$\mathcal{K}_{2,1}+1,1$	5	3	2	1 limb of $\mathcal{K}_{2,1}$ + 1 $\circ$-joint + 1 ∂-joint
$\mathcal{K}_{4,0}+2,0$	6	6	0	1 limb of $\mathcal{K}_{0,4}$ + 2 $\circ$-joints
$\mathcal{K}_{4,0}+0,2$	6	4	2	1 limb of $\mathcal{K}_{0,4}$ + ∂-bone
$\mathcal{K}_{3,0}+\mathcal{K}_{3,0}$	6	6	0	2 limbs of $\mathcal{K}_{3,0}$ = 2 wiggles + 2 turns
$\mathcal{K}_{2,1}+\mathcal{K}_{2,1}$	6	4	2	2 limbs of $\mathcal{K}_{2,1}$ = 2 wiggles + ∂-bone
$\mathcal{K}_{4,0}+\mathcal{K}_{4,0}$	8	8	0	2 limbs of $\mathcal{K}_{4,0}$ = 4 wiggles

Table 1: Classification of all curves with $\chi'(c)=0$.

Proof Assume, in contradiction, that $\mathbb{S}^3 \smallsetminus \mathcal{N}(\mathcal{L})$ has an essential sphere or a disk bounding a component of L. By Lemma 3.9 we may assume that S is taut. Let $\mathcal{C}$ be its curves of intersection With $P^{\pm}$. By Lemma 5.15,

$$0<\chi(S)=\sum_{c\in\mathcal{C}}\chi'(c).$$

However, this contradicts Lemma 5.16 which states that $\chi'(c)\le 0$ for all $c\in\mathcal{C}$. □

Lemma 5.18 *Let S be a taut surface with $\chi(S)=0$, then any curve $c\in\mathcal{C}$ is one of the following*:

(1) $c\in\mathcal{C}_{2,0}\cup\mathcal{C}_{1,2}$.

(2) $\mathfrak{J}(c)=4$, *ie* $c\in\mathcal{C}_{4,0}\cup\mathcal{C}_{2,2}\cup\mathcal{C}_{0,4}$.

(3) c *contains a limb* $\kappa\in\mathcal{K}$ *and has* $\mathfrak{J}(c)=\mathfrak{J}(\kappa)+2$.

(4) c *is the union of two limbs* $\kappa_1,\kappa_2\in\mathcal{K}$.

Moreover, *the cores of arcs in* $\mathcal{K}_{3,0}$ *are vertebrae of their abutting curves.*

Proof By Lemma 5.15 and 5.16 each curve $c\in\mathcal{C}$ must have $\chi'(c)=0$. It follow from the definition of χ' that the above are the only cases in which $\chi'(c)=0$.

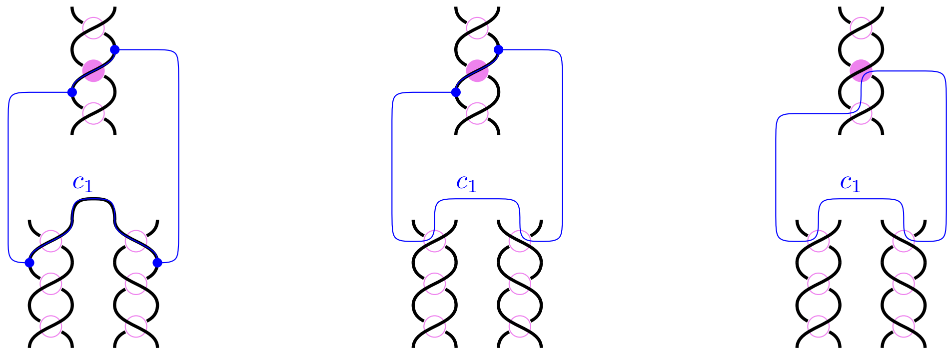
By Remark 5.13 and Case 1 of the proof of Lemma 5.16, if c contains an arc $\kappa\in\mathcal{K}_{3,0}$ whose core is not a vertebra of the abutting curve c' then $\chi'(c)<0$. □

In Table 1 we summarize the possible sets of curves with $\chi'=0$ and assign them names.

6 Atoroidal and unannular

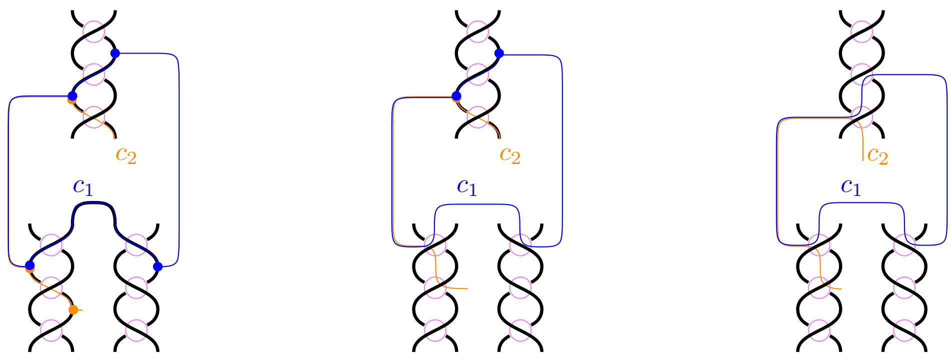
In this section we prove that if the link L has a diagram which satisfies the conditions of Theorem A, then its complement does not contain essential annuli or tori (in particular $\chi(S) = 0$). Let S denote such a 2-torus or annulus. After an isotopy, if need be, we may assume that S is taut. We first need the following technical lemmas.

Claim 6.1 *The following three configurations for curves $c \in \mathcal{C}$ are impossible*:



where the bubble marked in pink is nonextremal.

Proof We argue simultaneously that the three configurations are impossible. In each of these cases, let c_2 (marked in orange) be the depicted curve abutting c_1. The limb of c_2 that is shown here has three joints:



None of these joints belongs to a limb in $\mathcal{K}$: in all cases, the curve c_1 is not in $\mathcal{C}_{2,0}$ or $\mathcal{C}_{1,2}$ nor a rib with a vertebra. By Lemma 5.18, it follows that $\mathfrak{J}(c_2) = 4$. Thus the only way that c_2 can close up is in Figure 7. Let c_3 be the depicted curve abutting c_1. In the two left figures, one sees, as before, that $\mathfrak{J}(c_3) = 4$. In the figure on the right, c_3 is in $\mathcal{K}_{4,0} + 2, 0$ in the notation of Table 1: Indeed, c_3 cannot have $\mathfrak{J}(c_3) = 4$, as otherwise it bounds a twist reduction subdiagram. Thus, c_3 must contain an arc of $\mathcal{K}$. By elimination of the possibilities in Table 1, c_3 must be in $\mathcal{K}_{4,0} + 2, 0$. Thus, in all cases, c_3 can close only after passing through an additional twist region as shown in Figure 8. Thus, taking into account all

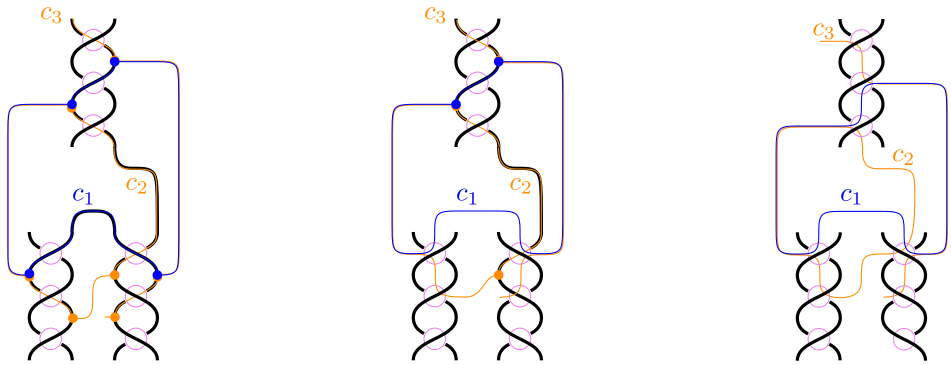


Figure 7: Proof of Claim 6.1. Second figure.

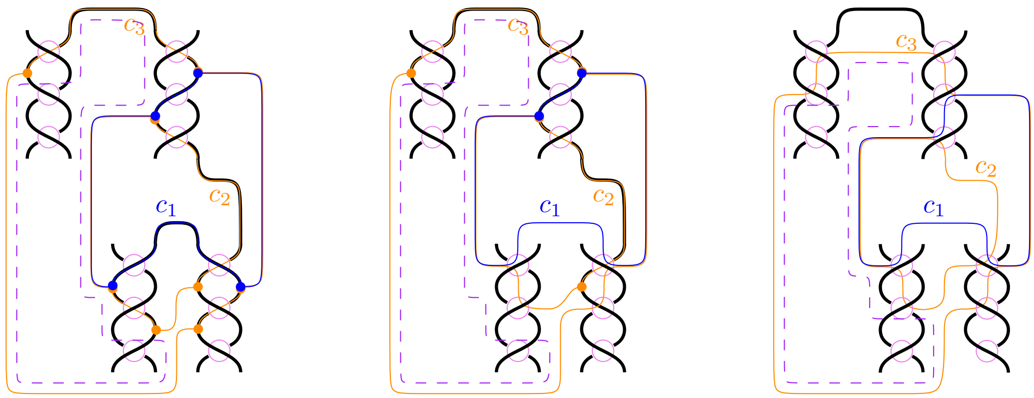


Figure 8: Proof of Claim 6.1. Third figure.

possible configurations of the curves c_2 and c_3 determined by the stated configurations of the c_1 curves, the diagrams are seen to contained a closed curve depicted by the dashed curves in the figures. Each of the dashed curves bounds a twist-reduction subdiagram. This contradicts the assumption that the diagrams are twist reduced, which finishes the proof of the claim. □

Lemma 6.2 *If a curve $c \in \mathcal{C}$ contains a bone τ connecting two turns of c then one of the following holds*:

(1) *the curve $c \in \mathcal{C}_{2,0}$*,

(2) *the curve c is a rib*, *or*

(3) *the $\circ$-bone τ meets the width edge of both twist regions.*

In particular, *c cannot have three consecutive turns.*

Proof Let τ be a bone connecting two turns of c it is therefore an $\circ$-bone. Assume in contradiction that τ and c do not satisfy any of (1)–(3) of the lemma. That is, c is not a curve in $\mathcal{C}_{2,0}$ nor a rib, and τ does not meet the width edge of both twist regions. There are two cases to consider depending on whether τ meets a width edge or not.

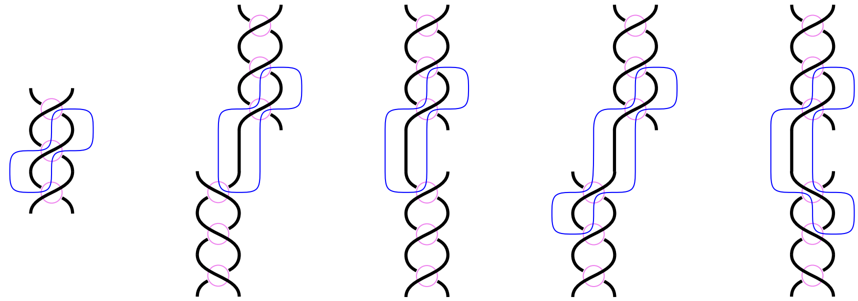


Figure 9: The five configurations of good curves which are not in $\mathcal{C}_{2,0}$.

If τ meets a width edge, then τ is a vertebra, ie it meets the length edge of one twist box and the width of the other. If this occurs set $c_0 = c,\ \tau_0 = \tau$. Since c is assumed not to be a rib, it follows from Lemma 5.6 that there exist curves $c_1, \dots, c_n$, limbs $\kappa_1, \dots, \kappa_n$, and vertebrae $\tau_1, \dots, \tau_{n-1}$ so that the terminal layer c_n has $\chi_+(c_n) < 0$. Moreover, note that the limb κ_n is not in $\mathcal{K}_{3,0}$ as otherwise by the uniqueness property, assured in Lemma 5.6, the "initial" layer curve c_0 must be a curve in $\mathcal{C}_{2,0}$ of type (i). This implies that κ_n does not meet any limb of $\mathcal{K}$. As otherwise, as in the proof of Lemma 5.11, one can prove that κ_n and the limb it meets must be equal. However, by Lemma 5.18, there is no curve, with $\chi' = 0$, which has three $\circ$-joints that do not belong to a limb of $\mathcal{K}$.

If τ does not meet a width edge, then τ meets the length edge of both twist regions. It follows that the curve c' abutting c along τ has two wiggles which are connected by τ. The curve c' cannot be in $\mathcal{C}_{4,0}$ as otherwise it bounds a twist-reduction subdiagram. By Lemma 5.18, one of the wiggles must meet a limb $\kappa \in \mathcal{K}$. Since τ is a core of c', it must be the core κ. It follows that $\kappa \in \mathcal{K}_{4,0}$ and that $c \in \mathcal{C}_{2,0}$ is of type (ii), in contradiction to our assumption.

In both cases, whether τ meets a width edge or not, we arrived at a contradiction. Hence, c must satisfy one of (1)–(3).

Finally, if c has three consecutive turns then c is not a rib nor a curve in $\mathcal{C}_{2,0}$. One of the two bones between the turns of c must meet a length edge and a width edge, contradicting (3). □

Definition 6.3 A curve $c \in \mathcal{C}$ is *good* if it bounds on $P^{\pm}$ exactly one component of $L \cap P^{\pm}$. I.e., it is either in $\mathcal{C}_{2,0}$ as depicted in Figure 5 (i) and (ii), or a curve in one of the forms shown in Figure 9.

We will say that c is good of type (I)–(V), accordingly. Otherwise, c is called *bad*.

Remark 6.4 Under the assumption that the diagram is prime and contains at least two twist regions the twist regions in each of the subfigures (II)–(V) of Figure 9 are distinct, as otherwise there is an arc of L connecting a twist region to itself, resulting in a nonprime subdiagram.

Remark 6.5 If S is a boundary parallel torus, then its intersection curves are good. A key observation is that the converse holds. That is, if all the curves of intersection of S with P are good then S is a

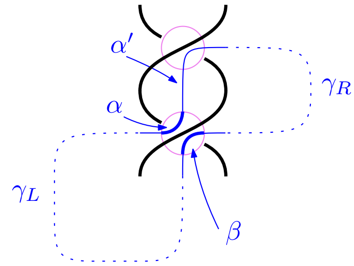


Figure 10: A curve c that meets a bubble twice.

boundary parallel torus: Consider a curve $c \in \mathcal{C}$; c is good and hence bounds on $P^{\pm}$ a unique component ℓ_c of $L \cap P^{\pm}$. When c meets a bubble B, there is a saddle of S bounded by B which meets c and two other curves c_1, c_2, and the component of $L \cap B$ that meets ℓ_c also meets ℓ_{c_1} and ℓ_{c_2}. Thus, the obvious isotopies from disks on S bounded by curves $c \in \mathcal{C}$ to ℓ_c can be glued together to form an isotopy of S to a component of L. Therefore, our goal in the next claims is to show that the curves of intersection of a taut surface S with $\chi(S) = 0$ are good.

Lemma 6.6 *If a curve $c \in \mathcal{C}$ passes through a bubble more than once, then c is good.*

Proof Let c meet the bubble B_1 more than once. By Lemma 3.9(5) it can do so only in two opposite $\circ$-joints, α, β. Therefore, at least one of those $\circ$-joints, say α, is part of a wiggle α' of c. Hence, α' meets an adjacent bubble B_2. Note that $c \smallsetminus (\alpha' \cup \beta)$ consists of two arcs connecting α' and β as depicted in Figure 10. Let γ_R, γ_L be the dotted subarcs of c on the right and left of the figure, respectively. The argument is divided into cases according to Table 1:

(1) The curve c contains three $\circ$-joints, hence it is not in $C_{2,0}$, $C_{2,2}$ or $C_{0,4}$.

(2) If $c \in \mathcal{C}_{4,0}$, then the subarc γ_R of c has no joints while γ_L has one $\circ$-joint. Hence c is good of type (I), (II) or (III).

(3) Since the three $\circ$-joints of c in T are not part of the same limb in $\mathcal{K}_{3,0}$ nor $\mathcal{K}_{4,0}$, then c cannot be in $\mathcal{K}_{3,0} + 0, 2$ or in $\mathcal{K}_{4,0} + 0, 2$ (See Table 1).

(4) The curve c cannot be in $\mathcal{K}_{4,0} + \mathcal{K}_{4,0}$: Otherwise β is part of a wiggle β' of c. The wiggle α' (resp. β') is part of a limb $\alpha'' \in \mathcal{K}_{4,0}$ (resp. $\beta'' \in \mathcal{K}_{4,0}$). Since each limb of $\mathcal{K}_{4,0}$ has a core, the bubble B_2 must be extremal. Then, the subarc γ_R of c contains the other wiggle of α''. The closed curve which is the union of γ_R and an arc on the boundary of the twist region intersects the link diagram twice, and both subdiagrams bounded by it are nontrivial. This contradicts the assumption that the diagram is prime.

(5) A similar argument shows that c cannot be in $\mathcal{K}_{2,1}$+$\mathcal{K}_{2,1}$.

(6) The curve c cannot be in $\mathcal{K}_{2,1} + 1, 1$: Otherwise, α' is the wiggle of some $\alpha'' \in \mathcal{K}_{2,1}$ and β is a turn. Beside α' and β, c has a ∂-bone on the subarc γ_L, and no joints on γ_R. Since $\alpha'' \in \mathcal{K}_{2,1}$ it abuts some $c_0 \in \mathcal{C}_{1,2}$. Hence, $c \cap L$ and $c_0 \cap L$ share endpoints. It follows that the union $(c \cap L) \cup (c_0 \cap L)$ is

a component of L passing over at most two wiggles of the diagram (at $c \cap L$) and under at most two wiggles (at $c_0 \cap L$). This contradicts the assumption that L is 3-highly twisted.

(7) If the curve c is in $\mathcal{K}_{4,0}+2,0$ then c is good of type (V): If β is part of a wiggle β' of c, then at most one of α' and β' is part of limb in $\mathcal{K}_{4,0}$. Without loss of generality, assume α' is a wiggle of a limb $\alpha'' \in \mathcal{K}_{4,0}$. Then, B_2 is extremal, and the subarc γ_R contains the other wiggle of α''. The curve which is the union of γ_R and an arc on the boundary of the twist region intersects the link diagram twice, in contradiction to the assumption that the diagram is prime. If β is a turn, then B_1 is extremal, and the wiggle α' is part of a limb $\alpha'' \in \mathcal{K}_{4,0}$. This limb abuts a curve $c_0 \in \mathcal{C}_{2,0}$, and it follows that c is good of type (V).

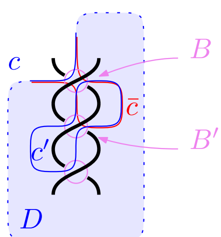
(8) Finally, if the curve c is in $\mathcal{K}_{3,0}+\mathcal{K}_{3,0}$ then c is good of type (IV): The wiggle α' is a wiggle of some limb $\alpha'' \in \mathcal{K}_{3,0}$. If β is part of a wiggle β' of c, then β' is a wiggle of some other limb $\beta'' \in \mathcal{K}_{3,0}$. It follows that each of γ_R, γ_L contains exactly one $\circ$-joint, which is impossible. If β is a turn, then it is the turn of some limb β'' in $\mathcal{K}_{3,0}$. As the core of β'' meets the width of the twist region, its wiggle must be on γ_L. Similarly, the turn of α'' must be on γ_L as well. This implies that γ_R does not contain any joints. If the turn of α'' and the wiggle of β'' are in two different twist regions then the curve abutting (both of) their cores contains three turns. However, this curve is a nonterminal layer curve (in the sense of Lemma 5.6), and those contain at most two turns. Thus, the turn of α'' and the wiggle of β'' are in the same twist region T'. Each of α'' and β'' meets an extremal bubble of T'. Then only option for them to close up is if they meet the same extremal bubble of T'. It follows that c is good of type (IV). □

Proposition 6.7 *All curves in $\mathcal{C}$ are good or in $\mathcal{C}_{0,4}$.*

In the proof of the proposition, we will assume in contradiction that such a curve exists. The proof will follow from the next four lemmas.

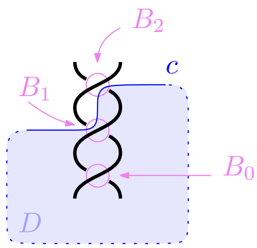
Lemma 6.8 *Assume that there are bad curves which are not in $\mathcal{C}_{0,4}$. Let c be an innermost bad curve in P^+ which is not in $\mathcal{C}_{0,4}$. Let D be the disk bounded by c. Then the curve c does not turn or wiggle through a twist region T which has a bubble contained in D.*

Proof **Assume c turns at T** Let B be the extreme bubble in a twist region T through which c turns. Since the diagram is 3-highly twisted T contains at least two more bubbles let B' be the bubble adjacent to B in T. By assumption B' is contained in the disk D. Consider the curve c' whose $\circ$-joint is opposite to the $\circ$-joint of c in B. Since c is innermost, the curve c' is good. It must be of good type (I) as in this configuration:

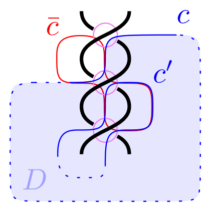


The curve $\bar{c}$ abutting both c and c', passes through the bubble B' twice. Hence, by Lemma 6.6, $\bar{c}$ is good. By the definitions, it cannot be in $\mathcal{C}_{2,0}$ nor good of type (I). If $\bar{c}$ is good of type (II) or (III) it has a turn at a bubble B'', then c passes twice through B'' which by Lemma 6.6 contradicts the assumption that c is bad. If $\bar{c}$ is good of type (IV) or (V) then it meets an extremal bubble B''. The curve c turns at B'' and hence belongs to $C_{2,0}$ which again contradicts the assumption that c is bad.

Assume *c* wiggles through *T* The curve c wiggles through the bubbles B_1, B_2. Let B_0 be a bubble of $T \cap D$ so that B_0, B_1, B_2 are consecutive, as in:



Consider the curve c' whose $\circ$-joint is opposite to the $\circ$-joint of c in B_1. The curve c' is contained in D and wiggles through T passing through the bubbles B_0, B_1:



By assumption c' must be good, and so it wiggles through T and then returns to T, passing through B_0, B_1, B_0 in that order. By Lemma 3.9(5), the two $\circ$-joints of c' in B_0 are opposite sides of the same saddle. Next, consider the curve $\bar{c}$ abutting c' along the two $\circ$-bones of c' connecting B_0 and B_1. The curve $\bar{c}$ passes through B_1 twice. Hence, it is good by Lemma 6.6. It must be of type (I) and in addition passes through B_2. It follows that c abuts $\bar{c}$ along the two $\circ$-bones of $\bar{c}$ connecting B_1 and B_2. Hence, c passes through B_2 twice, and by Lemma 6.6 c is good, contradicting the assumption. □

Lemma 6.9 *Assume that there are bad curves which are not in $\mathcal{C}_{0,4}$. Let c be an innermost bad curve in P^+ which is not in $\mathcal{C}_{0,4}$. Let D be the disk bounded by c. Then the curve c does not wiggle through a twist region.*

Proof Assume that c wiggles through a twist region T. By Lemma 6.8, the disk D does not contain a bubble of T. The curve c wiggles extremely through the twist region by passing through two bubbles B_0, B_1, where B_0 is extremal. By Lemma 6.6, c meets the bubble B_0 once.

The curve c' turning at B_0 is good by choice of c. Therefore, $c' \in \mathcal{C}_{4,0}$ is good of type (II) or (III) or $c' \in \mathcal{C}_{2,0}$ of type (i) or (ii) (as in Lemma 5.1); see Figure 11.

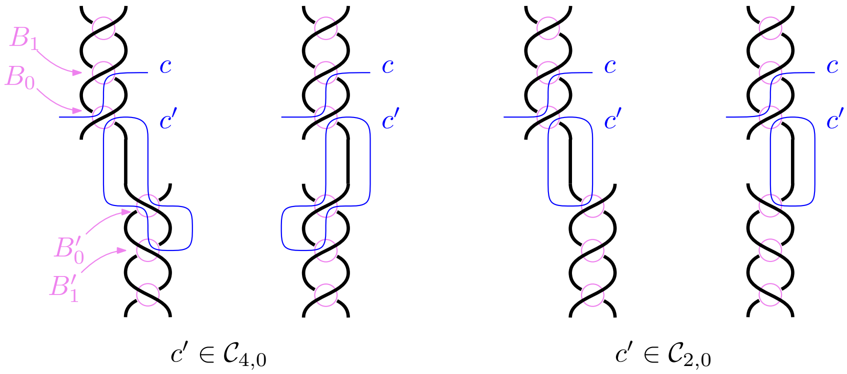
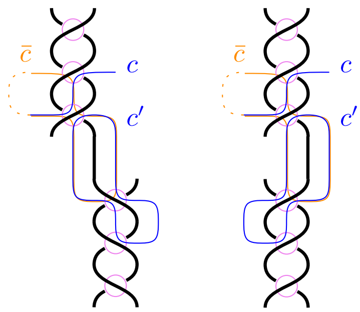


Figure 11: Proof of Lemma 6.9. The four configurations of c'.

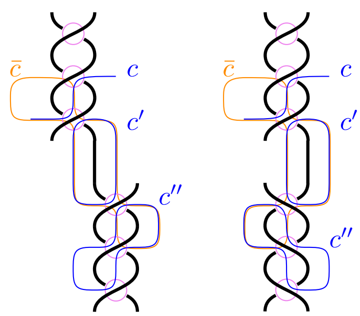
Let $T' \neq T$ be the other twist region which c' meets, let B'_0 denote the extremal bubble in T' through which c' passes, and let B'_1 be its adjacent bubble.

Case 1 If $c' \in \mathcal{C}_{4,0}$ (ie as depicted in the left two subfigures in Figure 11), then consider the curve $\bar{c}$ abutting c and c' (shown in orange in subsequent figures). None of the bubbles of $\bar{c}$ belongs to an arc of $\mathcal{K}$. Therefore, $\bar{c} \in \mathcal{C}_{4,0}$, and it follows that it must close up as shown in the dotted curves here:



As c must follow the dotted bone of $\bar{c}$, we see that c passes through the bubbles B_1 twice. By Lemma 6.6. This contradicts the assumption that c is bad.

Case 2 Let $c' \in \mathcal{C}_{2,0}$ (ie as depicted in the right two subfigures of Figure 11) and assume that c *does not* meet B'_0. Consider, now, the curve c'' whose $\circ$-joint in B'_0 is opposite to that of c'. By the choice of c, the curve c'' is good. It must be good of type (I). And the configuration is as depicted here:



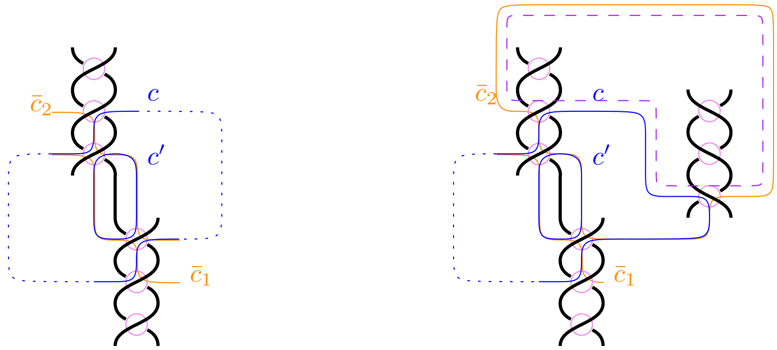


Figure 12: Subcase 3.1.

Considering the curve $\bar{c}$ abutting all of c, c', c'', we see that it must be good by Lemma 6.6 of type (IV) and (V) respectively. As in Case 1, it follows that c meets B_1 twice contradicting the assumption that c is bad.

Case 3 Assume that $c' \in \mathcal{C}_{2,0}$ of type (i) and c *does meet* B_0' (as in the third, counted from the left, subfigure of Figure 11). In this situation are three subcases to consider:

(1) c wiggles through T' at the bubbles B_0' and B_1'.

(2) c turns at B_0'.

(3) $c \cap L$ meets the bubble B_0'.

Subcase 3.1 The curve c wiggles through T' and through the bubbles B_0' and B_1'. After passing through B_1, the curve c must exit T at its right length edge and enter T' on its right length edge before meeting B_0'. Otherwise the curve c would be forced to pass through a bubble twice in contradiction to Lemma 6.6. Thus, the curve c is as in Figure 12, left. If $c \in \mathcal{C}_{4,0}$ then we get a twist-reduction subdiagram, in contradiction to the assumption. Thus, by Lemma 5.18, at least one of the wiggles of c, via B_0, B_1 or via B_0', B_1', must be part of an arc $\kappa \in \mathcal{K}$. Since the curves $\bar{c}_1, \bar{c}_2$ abutting c' are not in $C_{2,0}$ nor in $\mathcal{C}_{1,2}$ it is clear that $\kappa \notin \mathcal{K}_{4,0} \cup \mathcal{K}_{2,1}$. Hence $\kappa \in \mathcal{K}_{3,0}$. Since c has two wiggles, it must be in $\mathcal{K}_{3,0} + \mathcal{K}_{3,0}$ (as in Table 1), and each of the dotted subarcs (in Figure 12, left) must contain a turn.

Let $\kappa \in \mathcal{K}_{3,0}$ be the limb that wiggles through B_0', B_1'. Then, c is the terminal curve of some vertebra τ in a curve in $\mathcal{C}_{2,0}$ (in the sense of Lemma 5.6). By retracing backwards, we see that the sequence of curves terminating in c starts with $c' \in \mathcal{C}_{2,0}$, then produces $\bar{c}_2 \in \mathcal{C}_{4,0}$, and then finally produce the curve c (and the limb κ). In particular, $\bar{c}_2$ is a rib, and the $\circ$-bone connecting its two turns is a vertebra. Figure 12, right, shows an example of such a configuration of curves. As one can see, there are subarcs of $\bar{c}_2$ and c that together bound a twist-reduction subdiagram which is a contradiction.

Subcase 3.2 Assume c turns at the bubble B_0'. As explained at the beginning of the previous Subcase 3.1, the curve c must be as in Figure 13.

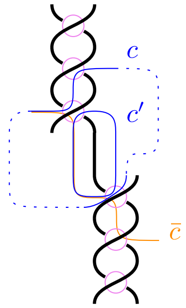
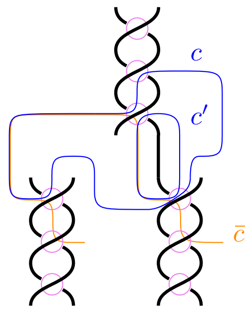


Figure 13: Subcase 3.2. The curve c' is of type (i) and c turns at B'_0.

Let $\bar{c}$ be the curve abutting c' as in Figure 13. The argument proceeds by dividing into cases according to Table 1:

(1) The curve c contains three $\circ$-joints, hence it is not in $C_{2,0}$, $C_{2,2}$ or $C_{0,4}$.

(2) The curve c is *not* in $\mathcal{C}_{4,0}$: Otherwise, c has to be a rib. The bone τ connecting the two turns cannot meet both width edges of the corresponding twist regions by Claim 6.1. Hence it is a vertebra, eg as in this figure:



As $\bar{c}$ has at least five $\circ$-joints, by Table 1 $\bar{c}$ is in $\mathcal{K}_{3,0}+\mathcal{K}_{3,0}$, and its two cores are those of the limbs in $\mathcal{K}_{3,0}$. The core of c ending in B_0 meets the length edge of both twist regions it connects. This is impossible for a core of a limb in $\mathcal{K}_{3,0}$.

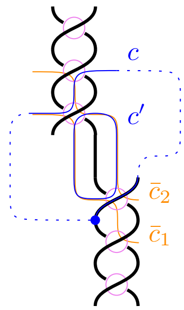
(3) The curve c cannot be in $\mathcal{K}_{4,0}+0,2$, $\mathcal{K}_{2,1}$+$\mathcal{K}_{2,1}$, or $\mathcal{K}_{4,0}+\mathcal{K}_{4,0}$ (see Table 1), since c contains a turn.

(4) The curve c cannot be in $\mathcal{K}_{3,0}+0,2$: Otherwise, the three $\circ$-joints of c are part of the same limb in $\mathcal{K}_{3,0}$. The core of such a limb connects a width edge to a length edge. This is not the case here.

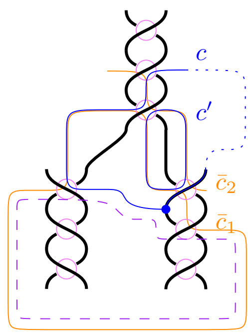
(5) The curve c cannot be in $\mathcal{K}_{2,1}+1,1$ or in $\mathcal{K}_{4,0}+2,0$: Otherwise, the wiggle of c is part of a limb in $\mathcal{K}_{2,1}$ or $\mathcal{K}_{4,0}$, respectively. It follows that the curve $\bar{c}$ abutting c and c' in Figure 13 must be in $\mathcal{C}_{1,2}$ or $\mathcal{C}_{2,0}$, respectively, which is clearly not the case.

(6) Finally, the curve c cannot be in $\mathcal{K}_{3,0}+\mathcal{K}_{3,0}$: Otherwise, it follows that γ_R must contain a single wiggle. However, in this case one can close γ_R with an arc along c' to obtain a curve in P intersecting the diagram in two points contradicting the assumption that the diagram is prime.

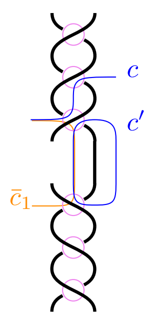
Subcase 3.3 If $c \cap L$ meets the bubble B_0', then it is as depicted here:



If $c \in \mathcal{C}_{2,2}$ then the left dotted line passes through no bubbles or intersection points, and we get a contradiction to the parity. Therefore, by Lemma 5.18, c must contain some arc $\kappa \in \mathcal{K}$. Clearly, κ must contain the subarc of c wiggling through T via B_0, B_1. The curve $\bar{c}_1$ is not in $\mathcal{C}_{1,2} \cup \mathcal{C}_{2,0}$ and therefore $\kappa \notin \mathcal{K}_{2,1} \cup \mathcal{K}_{4,0}$. It follows that κ is an arc in $\mathcal{K}_{3,0}$ and is the terminal layer limb of the process c', then $\bar{c}_1$, then κ, which is discussed in the proof of Lemma 5.6. If this is the case then, as in the end of Subcase 3.1, a subarc of $\bar{c}_1$ and a subarc of c bound a twist-reduction subdiagram as in the example shown here:



Case 4 Assume that $c' \in \mathcal{C}_{2,0}$ of type (ii) and c *does meet* B_0'. (as depicted in the rightmost subfigure of Figure 11). Let $\bar{c}_1$ be the curve abutting c as in:



The wiggle of c passing via B_0, B_1 is not part of any arc κ in $\mathcal{K}$: If it were, then the curve $\bar{c}_1$ will either be in $\mathcal{C}_{2,0} \cup \mathcal{C}_{1,2}$ or will be the nonterminal layer curve of a process terminating in κ. Clearly, $\bar{c}_1$ is not in $\mathcal{C}_{2,0} \cup \mathcal{C}_{1,2}$. It is also not a nonterminal layer of a process defining $\mathcal{K}_{3,0}$ as in Lemma 5.6, since at any

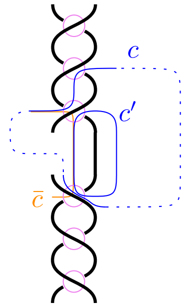


Figure 14: Subcase 4.2. The curve c' is of type (ii) and c turns at B'_0.

step of a process every curve is a rib, ie it has two turns and a wiggle, and the next step of the process abuts the $\circ$-bone connecting its the two turns, however here c does not abut the $\circ$-bone of $\bar{c}_1$ connecting its two turns.

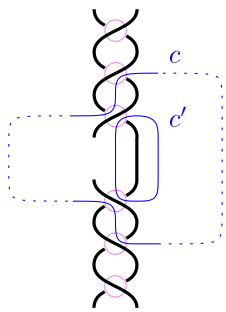
As in the previous subcase there are three further subsubcases to consider:

(1) c wiggles through T' at the bubbles B'_0 and B'_1.

(2) c turns at B'_0.

(3) $c \cap L$ meets the bubble B'_0.

Subcase 4.1 If c wiggles through the bubbles B'_0, B'_1, the exact same argument as above shows that the subarc of c passing through B'_0, B'_1 is not a subarc of any arc in $\mathcal{K}$. It follows that $c \in \mathcal{C}_{4,0}$ and bounds a twist-reduction subdiagram,



which is a contradiction.

Subcase 4.2 If c turns at the bubble B'_0. The curve c must be as in Figure 14. The argument is further divided into cases according to Table 1:

(1) The curve c contains three $\circ$-joints, hence it is not in $C_{2,0}$, $C_{2,2}$ or $C_{0,4}$.

(2) The curve c is not in $\mathcal{C}_{4,0}$: Otherwise, c has to be a rib. The bone τ connecting the two turns cannot meet both width edges of the corresponding twist regions by Claim 6.1. Hence it is a vertebra, as in Figure 15. It follows that the curve $\bar{c}$ abutting c and c' has at least three consecutive turns in contradiction to Lemma 6.2.

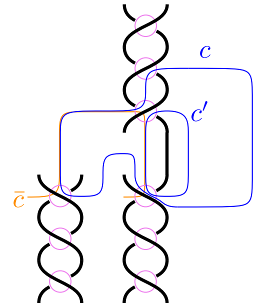


Figure 15: Subcase 4.2(2).

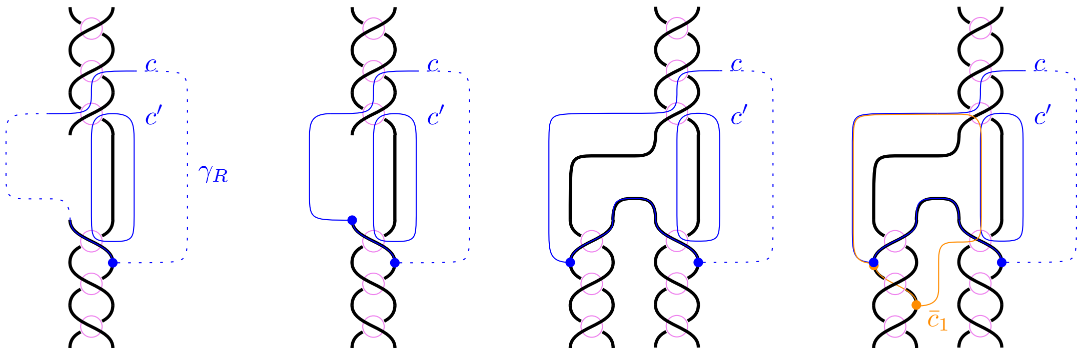


Figure 16: Subcase 4.3.

(3) The curve c cannot be in $\mathcal{K}_{4,0}+0,2$, $\mathcal{K}_{2,1}+\mathcal{K}_{2,1}$ or $\mathcal{K}_{4,0}+\mathcal{K}_{4,0}$ (See Table 1) since c contains a turn.

(4) The curve c cannot be in $\mathcal{K}_{3,0}+0,2$: Otherwise the three $\circ$-joints of c are part of the same limb in $\mathcal{K}_{3,0}$. However this would imply that the three $\circ$-joints of c are consecutive, which is impossible by the checkerboard coloring of the diagram.

(5) The curve c cannot be in $\mathcal{K}_{2,1}+1,1$ or in $\mathcal{K}_{4,0}+2,0$: Otherwise, the wiggle of c is part of a limb in $\mathcal{K}_{2,1}$ or $\mathcal{K}_{4,0}$, respectively. It follows that the curve $\bar{c}$ abutting c and c' in Figure 14 must be in $\mathcal{C}_{1,2}$ or $\mathcal{C}_{2,0}$, respectively, which is clearly not the case.

(6) Finally, the curve c cannot be in $\mathcal{K}_{3,0}+\mathcal{K}_{3,0}$: Otherwise, it follows that the curve $\bar{c}$ abutting c has three consecutive turns. This is impossible by Lemma 6.2.

Subcase 4.3 If $c\cap L$ meets the bubble B_0', then by Table 1 c is either in $\mathcal{C}_{2,2}$, or $\mathcal{K}_{2,1}+1,1$ or $\mathcal{K}_{2,1}+\mathcal{K}_{2,1}$. It cannot be in $\mathcal{K}_{2,1}+1,1$ or $\mathcal{K}_{2,1}+\mathcal{K}_{2,1}$ since otherwise the dotted subarc γ_R has a turn or a wiggle, respectively (see Figure 16, left), which would contradict primeness. Thus, $c\in\mathcal{C}_{2,2}$.

If $c\cap L$ passes over one crossing of L, then c bounds a twist-reduction subdiagram as in Figure 16, middle left. Otherwise, we are in Figure 16, middle right. Now consider how the curve $\bar{c}_1$ abutting

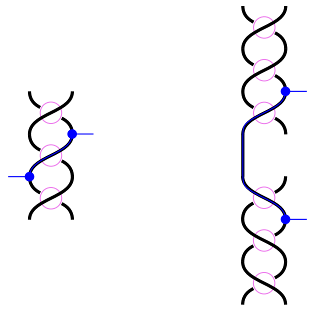
c' on its left can close up. It must be as depicted in Figure 16, right. Thus, $\bar{c}_1$ must be as one of the configurations that were ruled out in Claim 6.1. □

Lemma 6.10 *Assume that there are bad curves which are not in* $\mathcal{C}_{0,4}$. *Let* c *be an innermost bad curve in* P^+ *which is not in* $\mathcal{C}_{0,4}$. *Let* D *be the disk bounded by* c. *Then* $c \notin \mathcal{C}_{4,0}$.

Proof By Lemma 6.9, c only turns. However, this is impossible by Lemma 6.2. □

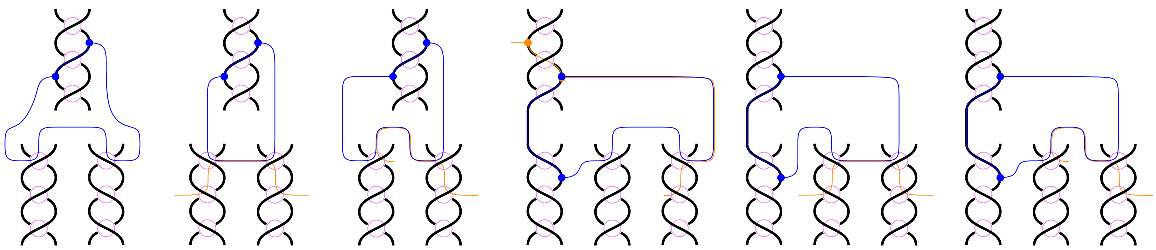
Lemma 6.11 *Assume that there are bad curves which are not in* $\mathcal{C}_{0,4}$. *Let* c *be an innermost bad curve in* P^+ *which is not in* $\mathcal{C}_{0,4}$, *and let* D *be the disk bounded by* c. *Then* $c \notin \mathcal{C}_{2,2}$.

Proof Assume $c \in \mathcal{C}_{2,2}$, Since $c \smallsetminus (c \cap L)$ passes through two bubbles, the endpoints of (a small continuation of) $c \cap L$ have the same color in the checkerboard coloring of $P \smallsetminus D(L)$. Thus it is one of the following:



(a) (b)

By Lemmas 6.8 and 6.9, the complement $c \smallsetminus (c \cap L)$ contains two turns, in twist regions which do not contain bubbles in D. Thus, the possible configurations are:



(a1) (a2) (a3) (b1) (b2) (b3)

In each of the cases, consider the curve $\bar{c}$ abutting c (a subarc of which is shown in orange).

Case (a1) is ruled out by Claim 6.1. In cases (a2), (b1), (b2), it is clear that the curve $\bar{c} \in \mathcal{C}_{4,0} \cup \mathcal{C}_{2,2}$ and bounds a twist-reduction subdiagram, which is a contradiction. Cases (a3), (b3) are impossible by Lemma 6.2. □

Proof of Proposition 6.7 Assume in contradiction that there are bad curves which are not in $\mathcal{C}_{0,4}$. Let c be an innermost bad curve in P^+ which is not in $\mathcal{C}_{0,4}$.

By the last two lemmas, c is not in $\mathcal{C}_{4,0}$ nor in $\mathcal{C}_{2,2}$. Thus, by Lemma 5.18, c must contain an arc of $\mathcal{K}$. Since every arc in $\mathcal{K}$ wiggles through some twist region, we get a contradiction to Lemma 6.9. This contradiction finishes the proof of Proposition 6.7. □

Corollary 6.12 *If the diagram of L is 3-highly twisted, connected, prime, and twist-reduced then $\mathbb{S}^3 \setminus \mathcal{N}(L)$ is atoroidal. In particular, L is prime.*

Proof Let $S \subset \mathbb{S}^3 \setminus \mathcal{N}(L)$ be an incompressible taut torus. Let $\mathcal{C}$ be the curves of intersection of S with $P^{\pm}$. Since S has no boundary, $\mathcal{C}_{0,4} = \varnothing$. By Proposition 6.7 all curves in $\mathcal{C}$ are good. We have seen in Remark 6.5 that if all curves are good then the torus S is boundary parallel.

If L was a composite knot, then the swallow-follow torus would be an essential torus in $\mathbb{S}^3 \setminus L$. □

Proposition 6.13 *If the diagram of L is 3-highly twisted, connected, prime and twist-reduced, then $\mathbb{S}^3 \setminus \mathcal{N}(L)$ is unannular.*

Proof Let S be an essential annulus. Since L is prime by Corollary 6.12, the annulus can be assumed not to have meridional boundary components. Thus, we may assume that S is taut. Let $\mathcal{C}$ be its curves of intersection with P. By Proposition 6.7 all curves in $\mathcal{C}$ are either good or in $\mathcal{C}_{0,4}$. Since S has boundary components, not all curves in $\mathcal{C}$ are good, and there is at least one curve $c \in \mathcal{C}_{0,4}$. We show that this is impossible.

If there exists a curve $c \in \mathcal{C}_{0,4}$ then the curve $\bar{c}$ abutting c has two intersection points which are not connected by an arc of $\bar{c} \cap L$ therefore $\bar{c} \in \mathcal{C}_{0,4}$. Repeating this argument shows that all the curves are in $\mathcal{C}_{0,4}$, ie $\mathcal{C} = \mathcal{C}_{0,4}$.

Case 1 There exists a curve $c \in \mathcal{C}_{0,4}$ which passes twice at the same twist region.

Denote the two connected components of $c \cap L$ by α and β. Let n be the number of crossings of T in-between α and β. We further divide the proof into subcases depending on n.

Subcase 1.0 ($n = 0$) By Lemma 3.9(5), α and β do not meet the same bubble of T. Since we assume $n = 0$, they must meet adjacent bubbles of T. The annulus S must spiral between the strands of $L \cap T$. Thus we obtain a disk of Type 2 (as in Lemma 3.1) and hence, by the definition of $\mathcal{C}^+$ and $\mathcal{C}^-$ as in the beginning of Section 3.2 this curve does not appear in $\mathcal{C}$.

Subcase 1.1 ($n = 1$) The tangle $L \cap T$ has two components λ_1, λ_2. Let l_1, l_2 be the corresponding components of L (possibly $l_1 = l_2$). Because $n = 1$ the arcs α and β meet the same string of $L \cap T$, say λ_1. Hence the two boundary components of S are contained in the same component l_1 of L. If $l_1 = l_2$, then there exists a curve $c' \in \mathcal{C}_{0,4}$ which meets the bridge in-between α and β, and this curve must be as in Subcase 1.0. Thus, we may assume that $l_1 \neq l_2$. Consider the disk Δ as depicted in Figure 17, left. Its interior intersects L in a single point in l_2, and its boundary is the union of an arc on S and an arc on l_1.

Figure 17: Left: the disk Δ. Right: A cross section of the twist box, the annulus S and the tori U, V.

The manifold $\mathcal{N}(S) \cup \mathcal{N}(l_1)$ has two torus boundary components U and V. See Figure 17, right. Let U be the torus that meets Δ. Let U_- be the component of $\mathbb{S}^3 \smallsetminus U$ containing l_2, and let U_+ be the other component. The torus U is incompressible in U_-, as such a compression must be on Δ and Δ does intersect l_2 once. It is also incompressible in U_+, as if a compression disk exists then since it cannot intersect l_1, it gives a compression of the annulus S, in contradiction to the incompressibility of S. By Corollary 6.12, U must be boundary parallel to either $\partial\mathcal{N}(l_2)$ or $\partial\mathcal{N}(l_1)$.

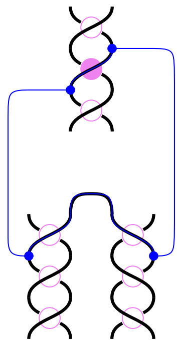
If U is parallel to $\partial\mathcal{N}(l_2)$, then since l_1 is parallel to a curve in U crossing Δ once there exists an annulus $A \subset \mathbb{S}^3 \smallsetminus L$ whose boundary is $l_1 \cup l_2$. This annulus A is incompressible, since otherwise $l_1 \cup l_2$ would be a 2-component unlink that is not linked with L, ie L is split, contradicting Corollary 5.17. The annulus A is trivially boundary-incompressible because the boundary components of A are on two different components of L. If we run the argument for A instead of S, Case 1.1 cannot occur because the boundary components of A are on two different components of L.

If U is parallel to $\partial\mathcal{N}(l_1)$, the intersection $\Delta \cap U$ is a curve on U which meets the meridian of $\mathcal{N}(l_1)$ exactly once: As if it meets it more than once, then the union $\mathcal{N}(\Delta) \cup \mathcal{N}(l_1)$ determines a once-punctured nontrivial lens space contained in $\mathbb{S}^3$, which is impossible. Thus, $\partial\Delta$ which is parallel to $\Delta \cap U$ is also parallel to l_1. Therefore, the arcs $\partial\Delta \smallsetminus l_1 \subset S$ and $l_1 \smallsetminus \partial\Delta \subset L$ bound a disk. Since the arc $\partial\Delta \smallsetminus l_1$ connects different components of S it is an essential arc, and the disk is a boundary compression for S, which is a contradiction.

Subcase 1.2 $(n \geq 2)$ As the boundary of the annulus S must pass through every other bridge in T, there must be another curve of $\mathcal{C}_{0,4}$ in between α and β. By choosing an innermost such curve we are back in one of the previous cases.

Case 2 None of the curves $c \in \mathcal{C}_{0,4}$ passes twice at the same twist region. Let c_1 be a curve in $\mathcal{C}_{0,4}$. Then $c_1 \cap L$ is the disjoint union of two arcs α, β. Since all of the curves are in $\mathcal{C}_{0,4}$ and at least one curve passes over a nonextremal bubble, we may assume, by changing c_1 that one of the components, say α_1, passes over a nonextremal bubble in a twist region. The component β cannot pass over one bubble, as in this case, either c_1 passes twice in the same twist region, or defines a twist-reduction subdiagram. Thus,

β must be the arc connecting two twist regions, passing over their two extremal bubbles, and the situation is as depicted here:



This case was ruled out in Claim 6.1. □

Remark 6.14 Subcase 1.1 ($n = 1$) in the proof of Proposition 6.13 follows also from the following well known general statement:

Nonsplit, annular atoroidal links in $\mathbb{S}^3$ are either torus knots or a link consisting of a torus knot on the "standard torus" $\mathbb{T}^2$ in $\mathbb{S}^3$ and one or both of the core curves of the solid tori components of $\mathbb{S}^3 \smallsetminus \mathcal{N}(\mathbb{T}^2)$.

For completeness we include a proof.

Proof Let $L \subset \mathbb{S}^3$ be a nonsplit and an atoroidal link in $\mathbb{S}^3$ containing an essential annulus A. If L is a knot then boundary A cuts $\partial\mathcal{N}(L)$ into two annuli A_1 and A_2. The surfaces $A \cup A_1$ and $A \cup A_2$ are tori which bound solid tori V_1 and V_2 as L is atoroidal. The two solid tori are glued to each other along A and since the result together with a regular neighborhood of L is $\mathbb{S}^3$, then by Seifert (see [13]), their complement is a regular neighborhood of a torus knot.

If L is a nonhyperbolic, nonsplit link whose exterior is atoroidal then by [14] it is a Seifert link, ie its exterior is a Seifert fiber space. Links in Seifert spaces were classified by Burde and Murasugi in [1]. They are either a connected sum of Hopf links or consist of a union of Seifert fibers in some Seifert fibration of $\mathbb{S}^3$. Atoroidal such links can contain at most three fibers. Hence the link L is a torus knot K on an unknotted solid torus $\mathbb{T}$ and the Hopf link which is the core curves of the complementing solid tori. □

Proof of Theorem A By Thurston [14], it suffices to prove that $\mathbb{S}^3 \smallsetminus \mathcal{N}(L)$ has incompressible boundary, and is irreducible, atoroidal and unannular. By Corollary 5.17 it has incompressible boundary and is irreducible, by Corollary 6.12 it is atoroidal, and by Proposition 6.13 it is unannular. □

7 Essential holed spheres in highly twisted link complements

In this section we use Theorem A to show that certain holed spheres in the complement of a highly twisted links are essential. We begin with three definitions.

Definition 7.1 Let $D(T)$ be a projection of a tangle (B,T) onto a disk $\Delta \subset P$ where $\partial\Delta$ is the simple closed curve $\gamma \subset P$ and so that the end points of the strings of T denoted by ∂T are contained in γ. We call $D(T)$ a *tangle diagram*. Let $\overline{D}(T)$ denote the reflection of $D(T)$ along P (ie the diagram with the same projection, but with reverse crossing data).

Definition 7.2 The diagram $D(T)$ is *relatively prime* (resp. *relatively twist reduced*, *relatively k-highly twisted*) if the link diagram obtained by gluing $D(T)$ and $\overline{D}(T)$ along their boundary is prime (resp. twist reduced, k-highly twisted).

Definition 7.3 A surface S in $(\mathbb{S}^3,\mathcal{L})$ is *pairwise-incompressible* if every disk D in $\mathbb{S}^3$ with $\partial D = D\cap S$, and which intersects $\mathcal{L}$ transversely in a single point, is isotopic to a disk in S by an isotopy preserving $\mathcal{L}$ setwise.

The surface S is *acylindrical* if the complement of S in $\mathbb{S}^3 \smallsetminus \mathcal{N}(\mathcal{L})$ contains no essential annuli whose boundary is on $S \cup \partial\mathcal{N}(\mathcal{L})$.

Theorem B *Let $D(L)$ be a 3-highly twisted diagram and let γ be a simple closed curve in $D(L)$ intersecting $D(L)$ transversely. Assume that both tangle diagrams bounded by γ are connected, relatively prime, relatively twist-reduced, relatively 3-highly twisted and contain at least two twist regions each. Let Σ be the sphere in $\mathbb{S}^3$ which intersects P transversely in γ, and does not intersect $\mathcal{L}$ outside γ. Then the punctured sphere $\Sigma' = \Sigma \smallsetminus \mathcal{N}(\mathcal{L})$ is incompressible, boundary-incompressible, pairwise-incompressible and acylindrical in $\mathbb{S}^3 \smallsetminus \mathcal{N}(\mathcal{L})$.*

Proof Assume in contradiction that $\Sigma' = \Sigma \smallsetminus \mathcal{N}(\mathcal{L})$ is either compressible, boundary-compressible, pairwise-compressible or not acylindrical in $\mathbb{S}^3 \smallsetminus \mathcal{N}(\mathcal{L})$.

Let B_1, B_2 be the two 3-balls that Σ bounds in $\mathbb{S}^3$, and let $E_1 = P \cap B_1,\ E_2 = P \cap B_2$ be the corresponding two disks in P bounded by γ. The induced tangle diagrams on E_1 and E_2 are assumed to be relatively prime, relatively twist-reduced and relatively 3-highly twisted. After doubling each of E_1 and E_2, as in Definition 7.2, we get two link diagrams $D(L_1), D(L_2)$ which are connected, prime, twist-reduced and 3-highly twisted. By Theorem A, the associated links $\mathcal{L}_1, \mathcal{L}_2$ are hyperbolic.

(1) If Σ' is compressible then the doubling of an innermost compressing disk $\Delta \subset B_i$ will give rise to an essential 2-sphere in $\mathbb{S}^3 \smallsetminus \mathcal{N}(\mathcal{L}_i)$.

(2) If Σ' is boundary compressible with boundary compression disk $\Delta \subset B_i$, then the doubling of Δ along $\Sigma' \cap \Delta$ results in a disk which is bounded by a component of $\mathcal{L}_i$. Hence either $\mathcal{L}_i$ is the unknot or has a split unknot component.

(3) If Σ' is pairwise boundary compressible, then the doubling of the essential disk $\Delta \subset B_i$ intersecting $\mathcal{L}$ once is a 2-sphere intersecting the boundary of $\mathcal{N}(\mathcal{L}_i)$ in two meridians. Thus $\mathcal{L}_i$ is not prime.

(4) If Σ' contains an essential annulus $A \subset B_i$ (ie Σ' is cylindrical), then $\mathcal{L}_i$ is toroidal if both boundaries of A are on Σ', or annular if one of these boundaries is on Σ' and the other on $\partial\mathcal{N}(\mathcal{L})$.

In all these cases we get that one of the links $\mathcal{L}_1, \mathcal{L}_2$ is not hyperbolic and thus a contradiction. □

References

[1] **G Burde**, **K Murasugi**, *Links and Seifert fiber spaces*, Duke Math. J. 37 (1970) 89–93 MR Zbl

[2] **M Culler**, **C M Gordon**, **J Luecke**, **P B Shalen**, *Dehn surgery on knots*, Bull. Amer. Math. Soc. 13 (1985) 43–45 MR Zbl

[3] **D Futer**, **E Kalfagianni**, **J S Purcell**, *Hyperbolic semi-adequate links*, Comm. Anal. Geom. 23 (2015) 993–1030 MR Zbl

[4] **D Futer**, **J S Purcell**, *Links with no exceptional surgeries*, Comment. Math. Helv. 82 (2007) 629–664 MR Zbl

[5] **A Giambrone**, *Combinatorics of link diagrams and volume*, J. Knot Theory Ramifications 24 (2015) art. id. 1550001 MR Zbl

[6] **C M Gordon**, **J Luecke**, *Knots are determined by their complements*, Bull. Amer. Math. Soc. 20 (1989) 83–87 MR Zbl

[7] **A Hatcher**, **W Thurston**, *Incompressible surfaces in* 2*–bridge knot complements*, Invent. Math. 79 (1985) 225–246 MR Zbl

[8] **M Lackenby**, *Word hyperbolic Dehn surgery*, Invent. Math. 140 (2000) 243–282 MR Zbl

[9] **T Li**, *Rank and genus of* 3*–manifolds*, J. Amer. Math. Soc. 26 (2013) 777–829 MR Zbl

[10] **M Lustig**, **Y Moriah**, *Are large distance Heegaard splittings generic?*, J. Reine Angew. Math. 670 (2012) 93–119 MR Zbl

[11] **W Menasco**, *Closed incompressible surfaces in alternating knot and link complements*, Topology 23 (1984) 37–44 MR Zbl

[12] **Y N Minsky**, **Y Moriah**, *Discrete primitive-stable representations with large rank surplus*, Geom. Topol. 17 (2013) 2223–2261 MR Zbl

[13] **H Seifert**, *Topologie Dreidimensionaler Gefaserter Räume*, Acta Math. 60 (1933) 147–238 MR Zbl

[14] **W P Thurston**, *The geometry and topology of three-manifolds*, lecture notes, Princeton University (1979) Available at `https://url.msp.org/gt3m`

Department of Mathematics, Technion
Haifa, Israel

`lazarovich@technion.ac.il`, `ymoriah@technion.ac.il`, `talipi@technion.ac.il`

Received: 6 September 2022 Revised: 23 July 2023

Algebraic & Geometric Topology **25**:1 (2025) 245–253
DOI: 10.2140/agt.2025.25.245
Published: 5 March 2025

Rational homology ribbon cobordism is a partial order

STEFAN FRIEDL
FILIP MISEV
RAPHAEL ZENTNER

We show that ribbon rational homology cobordism is a partial order within the class of irreducible 3-manifolds. This makes essential use of the methods recently employed by Ian Agol to show that ribbon knot concordance is a partial order.

57K10, 57K31, 57K40

1 Introduction

It was recently proved by Agol [1] that ribbon concordance of knots is a partial order. We prove an analogous statement for the preorder on irreducible, closed, connected, oriented 3-manifolds that is given by rational homology cobordisms.

Let Y_0 and Y_1 be closed, connected, oriented 3-manifolds. We say that a compact, connected, oriented, smooth 4-manifold W is a *cobordism from Y_0 to Y_1* if $\partial W = Y_1 \sqcup (-Y_0)$. A cobordism W is called a *$\mathbb{Q}$-homology cobordism from Y_0 to Y_1* if the inclusions of Y_0 and Y_1 into W both induce an isomorphism in homology with $\mathbb{Q}$-coefficients. Finally we say that W is a *ribbon cobordism from Y_0 to Y_1* if W is built from $Y_0 \times [0,1]$ by attaching only 1-handles and 2-handles.

We write $Y_1 \geq Y_0$ if there exists a ribbon $\mathbb{Q}$-homology cobordism from Y_0 to Y_1. Clearly this defines a preorder, ie the relation $\geq$ is reflexive and transitive. It is less clear whether this is antisymmetric and so defines a partial order, ie if $Y_0 \geq Y_1$ and $Y_1 \geq Y_0$, are then Y_0 and Y_1 homeomorphic?

The following conjecture was formulated by Daemi, Lidman, Vela-Vick, and Wong [3, Conjecture 1.1]:

Conjecture 1.1 *The preorder on the set of homeomorphism classes of closed, connected, oriented 3-manifolds given by ribbon $\mathbb{Q}$-homology cobordism is a partial order, ie if one has $Y_0 \geq Y_1$ and $Y_1 \geq Y_0$ then Y_0 and Y_1 are homeomorphic.*

We prove this conjecture for irreducible 3-manifolds.

Theorem 1.2 *The preorder $\geq$ is a partial order on the set of homeomorphism classes of irreducible, closed, connected, oriented 3-manifolds. In particular, if $Y_0 \geq Y_1$ and $Y_1 \geq Y_0$, then Y_0 and Y_1 are homeomorphic.*

In fact for aspherical 3-manifolds we can prove a refinement.

Theorem 1.3 *The preorder $\geq$ is a partial order on the set of orientation-preserving homeomorphism classes of aspherical, closed, connected, oriented 3-manifolds. In particular, if $Y_0 \geq Y_1$ and $Y_1 \geq Y_0$, then there exists an orientation-preserving homeomorphism from Y_0 to Y_1.*

It is not clear to us whether the conclusion of Theorem 1.3 also holds for irreducible 3-manifolds that are not aspherical. This leads us to the following question. To formulate it, we call an oriented 3-manifold *chiral* if it does not admit a self-homeomorphism which is orientation-reversing.

Question 1.4 *Does there exist a chiral spherical oriented 3-manifold Y with $Y \leq -Y$?*

In the proofs of Theorems 1.2 and 1.3 we make essential use of the methods employed by Agol to prove that ribbon concordance is a partial order. More precisely, Agol's methods apply to the situation we are considering and provide us with the following theorem.

Theorem 1.5 *Suppose Y is a closed, connected, oriented 3-manifold. Suppose W is a ribbon $\mathbb{Q}$-homology cobordism from $Y_- \cong Y$ to $Y_+ \cong Y$, ie $Y_+ \geq Y_-$ (and so from Y to itself). Then the inclusion $\iota_+ : Y_+ \to W$ induces an isomorphism on fundamental groups.*

Using Theorem 1.5 one can easily prove the following corollary which is the key ingredient in the proofs of Theorems 1.2 and 1.3.

Corollary 1.6 *Suppose W_0 is a ribbon $\mathbb{Q}$-homology cobordism from Y_0 to Y_1, so that $Y_1 \geq Y_0$. Suppose that W_1 is a ribbon $\mathbb{Q}$-homology cobordism from Y_1 to Y_0, so that $Y_0 \geq Y_1$. Then the injection $\iota_0 : Y_1 \to W_0$ induces an isomorphism of fundamental groups, and likewise for the injection $\iota_1 : Y_0 \to W_1$.*

Remark We use the convention of [3], in which Conjecture 1.1 has been formulated, where $Y_1 \geq Y_0$ means that there is a ribbon cobordism *from* Y_0 *to* Y_1, as defined above. We would like to point out that this may result in some confusion when comparing with Gordon's initial convention [5] and the one employed in [1], where a ribbon concordance goes from K_1 to K_0 if the exterior $E(C)$ of the concordance C is obtained by adding only 1-handles and 2-handles to $E(K_0) \times [0,1]$. In other words, in their convention, a ribbon cobordism goes *from* the more complex object *to* the simpler one.

Outline 1 In Section 2 we will state and prove some auxiliary results that we will need for the proof of our main results, Theorems 1.2 and 1.3. In Section 3 we provide the proof of Theorem 1.3. Afterwards in Section 4 we will use Theorem 1.3 as an ingredient in our proof of Theorem 1.2. Finally Section 5 contains a sketch of proof of Theorem 1.5 which is almost verbatim identical to Agol's proof in the context of knot complements, together with the proof of Corollary 1.6.

Acknowledgments

Friedl and Misev were supported by the SFB 1085 "higher invariants" at the University of Regensburg, funded by the DFG. Zentner thanks the DFG for support through the Heisenberg program. We also wish to thank Ederson Dutra and Lukas Lewark for helpful conversations, and Lukas Lewark for comments on a preliminary version of this article. Finally, we thank the anonymous referee for their comments.

Shortly after our article first appeared on arXiv, Marius Huber [8] published a paper containing essentially the same main result.

2 Preparations

In this short section we collect a few basic facts that we will use in the proof of Theorem 1.2. The following appears as [3, Proposition 2.1].

Proposition 2.1 (Daemi, Lidman, Vela-Vick, Wong) *Let Y_0 and Y_1 be closed, connected, oriented 3-manifolds and let W be a cobordism from Y_0 to Y_1. We denote by $\iota_0\colon Y_0 \to W$ and $\iota_1\colon Y_1 \to W$ the obvious inclusion maps.*

(1) *If W is a ribbon cobordism, then the map $(\iota_1)_*\colon \pi_1(Y_1) \to \pi_1(W)$ is an epimorphism.*

(2) *If W is a ribbon $\mathbb{Q}$-homology cobordism, then the map $(\iota_0)_*\colon \pi_1(Y_0) \to \pi_1(W)$ is a monomorphism.*

By definition, a ribbon cobordism W from Y_0 to Y_1 is obtained from $Y_0 \times [0,1]$ by attaching 1-handles and 2-handles. Flipping W upside down, we see that W can equivalently be viewed as $Y_1 \times [0,1]$ with some 2-handles and 3-handles attached. This immediately implies the first statement. For the second statement, Daemi, Lidman, Vela-Vick, and Wong use Gersenhaber and Rothaus' Theorem 2 in [4], much in the same way as initially done by Gordon in [5, Proof of Lemma 3.1], using the residual finiteness of 3-manifold groups.

The next proposition gives us some homological information about $\mathbb{Q}$-homology cobordisms.

Proposition 2.2 *Let Y_0 and Y_1 be closed, connected, oriented 3-manifolds. We consider their fundamental classes $[Y_0] \in H_3(Y_0;\mathbb{Z})$ and $[Y_1] \in H_3(Y_1;\mathbb{Z})$. Let W be a cobordism from Y_0 to Y_1. If W is a $\mathbb{Q}$-homology cobordism, then the maps $(\iota_0)_*\colon H_3(Y_0;\mathbb{Z}) \to H_3(W;\mathbb{Z})$ and $(\iota_1)_*\colon H_3(Y_1;\mathbb{Z}) \to H_3(W;\mathbb{Z})$ are both isomorphisms. Furthermore $(\iota_0)_*([Y_0]) = (\iota_1)_*([Y_1]) \in H_3(W;\mathbb{Z})$.*

Proof Let $i \in \{0,1\}$. We have the following commutative diagram:

$$\begin{tikzcd}
\cdots \arrow[r] & H_4(W,Y_i;\mathbb{Z}) \arrow[r] & H_3(Y_i;\mathbb{Z}) \arrow[r, "(\iota_i)_*"] \arrow[d] & H_3(W;\mathbb{Z}) \arrow[r] \arrow[d] & H_3(W,Y_i;\mathbb{Z}) \arrow[r] & \cdots \\
& & H_3(Y_i;\mathbb{Q}) \arrow[r, "(\iota_i)_*", "\cong"'] & H_3(W;\mathbb{Q}) & &
\end{tikzcd}$$

Since W is a $\mathbb{Q}$-homology cobordism we know that the bottom horizontal map is an isomorphism. It follows from this observation and the universal coefficient theorem that the kernel and the cokernel of the map $(\iota_i)_*\colon H_3(Y_i;\mathbb{Z}) \to H_3(W;\mathbb{Z})$ are finite. But by Poincaré duality and the universal coefficient theorem we see that

$$H_3(W, Y_i;\mathbb{Z}) \cong H^1(W, Y_{1-i};\mathbb{Z}) \cong \mathrm{Hom}(H_1(W, Y_{1-i};\mathbb{Z}),\mathbb{Z}) \oplus \mathrm{Ext}(H_0(W, Y_{1-i};\mathbb{Z}),\mathbb{Z})$$

is torsion-free. Similarly, we see that $H_4(W, Y_i;\mathbb{Z})$ is torsion-free, since

$$H_4(W, Y_i;\mathbb{Z}) \cong H^0(W, Y_{1-i};\mathbb{Z}) \cong \mathrm{Hom}(H_0(W, Y_{1-i};\mathbb{Z}),\mathbb{Z}).$$

In summary, we see that the two maps $(\iota_0)_*\colon H_3(Y_0;\mathbb{Z}) \to H_3(W;\mathbb{Z})$ and $(\iota_1)_*\colon H_3(Y_1;\mathbb{Z}) \to H_3(W;\mathbb{Z})$ are isomorphisms.

It remains to show that $(\iota_0)_*([Y_0]) = (\iota_1)_*([Y_1]) \in H_3(W;\mathbb{Z})$. We consider the long exact sequence of the pair $(W, Y_0 \cup Y_1)$:

$$\cdots \to H_4(W, Y_0 \cup Y_1;\mathbb{Z}) \xrightarrow{\partial} H_3(Y_0 \cup Y_1;\mathbb{Z}) \to H_3(W;\mathbb{Z}) \to \cdots .$$

It is well known that $\partial([W]) = [\partial W]$. Since W is a cobordism, we have $\partial W = Y_1 \sqcup (-Y_0)$; in particular $[\partial W] = [Y_1] - [Y_0]$. Since the sequence is exact we see that the image of $[Y_1] - [Y_0]$ is zero in $H_3(W;\mathbb{Z})$, ie we have $(\iota_0)_*([Y_0]) - (\iota_1)_*([Y_1]) = 0 \in H_3(W;\mathbb{Z})$. □

Lemma 2.3 *Suppose that we have a continuous map $g\colon X \to Z$, between path-connected topological spaces X and Z which admit the structure of a CW-complex. Then for any k the following diagram commutes:*

$$\begin{tikzcd} H_k(X;\mathbb{Z}) \arrow[r, "g_*"] \arrow[d, "(j_X)_*"'] & H_k(Z;\mathbb{Z}) \arrow[d, "(j_Z)_*"] \\ H_k(\pi_1(X);\mathbb{Z}) \arrow[r, "g_*"] & H_k(\pi_1(Z);\mathbb{Z}) \end{tikzcd}$$

Here the group homology $H_i(G;\mathbb{Z})$ of a group G is defined to be the singular homology of its associated $K(G,1)$-space: $H_i(G;\mathbb{Z}) := H_i(K(G,1);\mathbb{Z})$. Furthermore $j_X\colon X \to K(\pi_1(X),1)$ and $j_Z\colon Z \to K(\pi_1(Z),1)$ are the natural maps of the respective spaces to the Eilenberg–Mac Lane space associated to their fundamental groups.

Proof We denote by $K(g)\colon K(\pi_1(X),1) \to K(\pi_1(Z),1)$ the map induced by the group homomorphism $g_*\colon \pi_1(X) \to \pi_1(Z)$. By construction this map induces the map

$$g_*\colon \pi_1(X) \xrightarrow{\cong} \pi_1(K(\pi_1(X),1)) \to \pi_1(K(\pi_1(Z)),1) \xrightarrow{\cong} \pi_1(Z).$$

Therefore, the two maps $K(g) \circ j_X$ and $j_Z \circ g$ induce the same maps on the fundamental group. Since their image space $K(\pi_1(Z),1)$ is aspherical, these two maps are homotopic by Whitehead's theorem, and hence induce the same maps in homology. □

Theorem 2.4 *Let Y_0 and Y_1 be irreducible, closed, orientable 3-manifolds and let $\alpha\colon \pi_1(Y_0) \to \pi_1(Y_1)$ be an isomorphism.*

(1) *If Y_0 and Y_1 are not lens spaces, then Y_0 and Y_1 are homeomorphic.*

(2) *If Y_0 and Y_1 are not spherical, then there exists a homeomorphism $g\colon Y_0 \to Y_1$ which induces α, ie which satisfies $g_* = \alpha\colon \pi_1(Y_0) \to \pi_1(Y_1)$.*

Proof The theorem is stated as [2, Theorem 2.1.2]. It is a consequence of the geometrization theorem, the Mostow rigidity theorem, work of Waldhausen [11, Corollary 6.5] and Scott [10, Theorem 3.1], and classical work on spherical 3-manifolds (see [9, page 113]). $\square$

We conclude our section of preparations with the following theorem, recently proved by Huber [7].

Theorem 2.5 (Huber) *Let $L(p_1, q_1)$ and $L(p_2, q_2)$ be lens spaces. If $L(p_1, q_1) \le L(p_2, q_2)$, then one of the following holds:*

(1) *The lens spaces are homeomorphic.*

(2) *There exists an $n \ge 2$ with $L(p_1, q_1) \cong L(n, 1)$ and $p_2/q_2 \in \mathcal{F}_n$, where*
$$\mathcal{F}_n := \left\{ \frac{nm^2}{nmk+1} \;\middle|\; m > k > 0,\ \gcd(m,k) = 1 \right\}.$$

(3) $L(p_1, q_1) \cong S^3$.

3 Proof of Theorem 1.3

We first provide the proof of Theorem 1.3, since the statement of Theorem 1.3 gives us also most of Theorem 1.2.

Proof of Theorem 1.3 Let Y_0 and Y_1 be aspherical, closed, connected, oriented 3-manifolds. We assume that $Y_0 \ge Y_1$ and that $Y_1 \ge Y_0$. We need to show that there exists an orientation-preserving homeomorphism from Y_0 to Y_1.

Let W be a ribbon $\mathbb{Q}$-homology cobordism from Y_0 to Y_1. We denote by $\iota_0\colon Y_0 \to W$ and $\iota_1\colon Y_1 \to W$ the obvious inclusion maps. Since we also assume that $Y_1 \ge Y_0$, we obtain from Corollary 1.6 that $(\iota_1)_*\colon \pi_1(Y_1) \to \pi_1(W)$ is an isomorphism. Furthermore by Proposition 2.1(2) we know that $(\iota_0)_*\colon \pi_1(Y_0) \to \pi_1(W)$ is a monomorphism. We write $\alpha := ((\iota_1)_*)^{-1} \circ (\iota_0)_*\colon \pi_1(Y_0) \to \pi_1(Y_1)$.

Since Y_1 is aspherical, there exists a map $f\colon Y_0 \to Y_1$ with $f_* = \alpha\colon \pi_1(Y_0) \to \pi_1(Y_1)$. We will show that this map f has degree equal to ± 1. For this we consider the following diagram, which is commutative by Lemma 2.3:

$$\begin{tikzcd} H_3(Y_0;\mathbb{Z}) \arrow[r, "(\iota_0)_*", "\cong"'] \arrow[d, "(j_0)_*"', "\cong"] & H_3(W;\mathbb{Z}) \arrow[d, "(j_W)_*"] & H_3(Y_1;\mathbb{Z}) \arrow[l, "(\iota_1)_*"', "\cong"] \arrow[d, "\cong"', "(j_1)_*"] \\ H_3(\pi_1(Y_0);\mathbb{Z}) \arrow[r, "(\iota_0)_*"] & H_3(\pi_1(W);\mathbb{Z}) & H_3(\pi_1(Y_1);\mathbb{Z}) \arrow[l, "(\iota_1)_*"', "\cong"] \end{tikzcd} \tag{1}$$

Here, $j_0 \colon Y_0 \hookrightarrow K(\pi_1(Y_0), 1)$ is the natural map of Y_0 to the Eilenberg–MacLane space of its fundamental group, and likewise j_1 and j_W are defined.

It follows from our hypothesis that W is a $\mathbb{Q}$-homology cobordism and Proposition 2.2 that the horizontal maps in the first line of this diagram are isomorphisms. By hypothesis Y_0 and Y_1 are already aspherical, ie they are Eilenberg–MacLane spaces. Therefore, the maps $(j_0)_*$ and $(j_1)_*$ are isomorphisms.

Now by commutativity of the right square in the diagram (1), the vertical map $(j_W)_*$ induced by the inclusion map j_W must be an isomorphism. Therefore, by commutativity of the left square in this diagram, we conclude that the map $(\iota_0)_* \colon H_3(\pi_1(Y_0); \mathbb{Z}) \to H_3(\pi_1(W); \mathbb{Z})$ is an isomorphism.

As in the proof of Lemma 2.3 above, we denote by $K(\varphi)\colon K(G, 1) \to K(H, 1)$ the map induced on Eilenberg–MacLane spaces by a group homomorphism $\varphi \colon G \to H$. In our situation, we have the two maps $f \colon Y_0 \to Y_1$ and $K((\iota_1)_*^{-1}) \circ K((\iota_0)_*) \colon Y_0 \to Y_1$, where we identify Y_0 and Y_1 with Eilenberg–MacLane spaces of their respective fundamental groups. Both induce the same map at the level of fundamental groups. Since Y_1 is aspherical, these maps are homotopic by Whitehead's theorem. Therefore, they induce the same map on homology. By the above observation that the bottom maps in diagram (1) are isomorphisms, we conclude that $f_* \colon H_3(Y_0) \to H_3(Y_1)$ is an isomorphism, and therefore f has degree ± 1.

By a standard argument a map $f \colon Y_0 \to Y_1$ of degree ± 1 induces an epimorphism of fundamental groups. Since we already know that $f_* = \alpha$ is a monomorphism, we see that $f_* \colon \pi_1(Y_0) \to \pi_1(Y_1)$ is an isomorphism. Thus it follows from Theorem 2.4(2) that there exists a homeomorphism $g \colon Y_0 \to Y_1$ that induces $f_* = \alpha$. It remains to show that g is orientation-preserving. To do so we consider the following commutative diagram:

(2)
$$\begin{tikzcd}
H_3(Y_0;\mathbb{Z}) \arrow[r, "(\iota_0)_*", "\cong"'] \arrow[d, "(j_0)_*"', "\cong"] & H_3(W;\mathbb{Z}) \arrow[d, "(j_W)_*"] & H_3(Y_1;\mathbb{Z}) \arrow[l, "(\iota_1)_*"', "\cong"] \arrow[d, "\cong"', "(j_1)_*"] \\
H_3(\pi_1(Y_0);\mathbb{Z}) \arrow[r, "(\iota_0)_*"] \arrow[d, "\mathrm{id}"'] & H_3(\pi_1(W);\mathbb{Z}) & H_3(\pi_1(Y_1);\mathbb{Z}) \arrow[l, "(\iota_1)_*"', "\cong"] \arrow[d, "\mathrm{id}"'] \\
H_3(\pi_1(Y_0);\mathbb{Z}) \arrow[rr, "\alpha=f_*=g_*"] & & H_3(\pi_1(Y_1);\mathbb{Z}) \\
H_3(Y_0;\mathbb{Z}) \arrow[u, "\cong", "(j_0)_*"'] \arrow[rr, "g_*", "\cong"'] & & H_3(Y_1;\mathbb{Z}) \arrow[u, "(j_1)_*", "\cong"']
\end{tikzcd}$$

We already know that the top part of the diagram (2) commutes. The center part of diagram (2) commutes since $\alpha = ((\iota_1)_*)^{-1} \circ (\iota_0)_*$. The bottom part of diagram (2) commutes again by Lemma 2.3. Finally by design we have $\alpha = f_* = g_*$.

By Proposition 2.2 we know that $(\iota_0)_*([Y_0]) = (\iota_1)_*([Y_1]) \in H_3(W; \mathbb{Z})$. But by the above this implies that $g_*([Y_0]) = [Y_1]$. □

4 Proof of Theorem 1.2

Proof of Theorem 1.2 Let Y_0 and Y_1 be irreducible closed connected oriented 3-manifolds. We assume that $Y_0 \geq Y_1$ and that $Y_1 \geq Y_0$. We need to show that Y_0 and Y_1 are homeomorphic. If $\pi_1(Y_0)$ and $\pi_1(Y_1)$ are infinite, it follows by a standard argument, using the sphere theorem, that Y_1 is aspherical; see [2, page 48]. Thus we see that the statement follows from Theorem 1.3.

Therefore it suffices to consider the case that $\pi_1(Y_0)$ or $\pi_1(Y_1)$ is finite. We assume that $Y_0 \geq Y_1$ and that $Y_1 \geq Y_0$. We need to show that Y_0 and Y_1 are homeomorphic. By symmetry we can assume that $\pi_1(Y_1)$ is finite.

Let W_0 be a ribbon $\mathbb{Q}$-homology cobordism from Y_0 to Y_1 and let W_1 be a ribbon $\mathbb{Q}$-homology cobordism from Y_1 to Y_0. By Corollary 1.6 we know that the inclusion-induced maps $\pi_1(Y_1) \to \pi_1(W_0)$ and $\pi_1(Y_0) \to \pi_1(W_1)$ are isomorphisms, and by Proposition 2.1 we know that the inclusion-induced maps $\pi_1(Y_0) \to \pi_1(W_0)$ and $\pi_1(Y_1) \to \pi_1(W_1)$ are monomorphisms.

It follows that $|\pi_1(Y_1)| \leq |\pi_1(W_1)| = |\pi_1(Y_0)| \leq |\pi_1(W_0)| = |\pi_1(Y_1)|$. Since $\pi_1(Y_1)$ is finite we see that we have equalities throughout and we see that all the inclusion-induced maps are isomorphisms. In particular we see that $\pi_1(Y_0) \cong \pi_1(Y_1)$.

First we consider the case that $\pi_1(Y_0)$, and thus also $\pi_1(Y_1)$, is not cyclic. In this setting it follows from Theorem 2.4 that Y_0 is homeomorphic to Y_1.

Finally we consider the case that $\pi_1(Y_0)$, and thus also $\pi_1(Y_1)$, is cyclic. It follows from [2, page 25] that neither Y_0 nor Y_1 are lens spaces. But it follows almost immediately from Theorem 2.5 that if for two lens spaces Y_0 and Y_1 we have $Y_1 \geq Y_0$ and if they have isomorphic fundamental groups, then Y_0 and Y_1 are homeomorphic. □

5 Sketch of proof of Theorem 1.5

In this section we provide a sketch of proof of Theorem 1.5 and Corollary 1.6, both of which are essentially due to [1], although his formulation is in the context of knot complements.

Sketch of proof of Theorem 1.5 This follows almost verbatim in the same way as in all but the last paragraph of [1, Proof of Theorem 1.2]. The only comment to make is that his proof uses residual finiteness of fundamental groups of knot complements. This is due to [6], using Thurston's proof of his geometrization conjecture for Haken manifolds. In our situation, we need the fact that all fundamental groups of 3-manifolds are residually finite, and this uses the full geometrization conjecture, together with Hempel's result. We will outline this proof for the sake of completeness.

Suppose that W is a rational ribbon homology cobordism from Y_- to Y_+, where both Y_+ and Y_- are homeomorphic to Y. For a finitely presented group π, Agol considers the representation variety $R_N(\pi) = \mathrm{Hom}(\pi, \mathrm{SO}(N))$ for some $N \geq 1$, and in the case of a path-connected topological space X, he defines $R_N(X) := R_N(\pi_1(X))$. This representation variety is a real algebraic set. Since the inclusion $\iota_+ \colon Y_+ \to W$ defines a surjection at the level of fundamental groups, we obtain an injection $\iota_+^* \colon R_N(W) \to R_N(Y_+)$ by precomposition of representations with $(\iota_+)_* \colon \pi_1(Y_+) \to \pi_1(W)$. In fact, since there is a presentation of $\pi_1(W)$ obtained by one from $\pi_1(Y_+)$ by only possibly adding relations, but no generators, one can realize $R_N(W)$ as an algebraic subset of $R_N(Y_+)$, $R_N(W) \subseteq R_N(Y_+)$.

On the other hand, the inclusion $\iota_- \colon Y_- \to W$ induces an injection of fundamental groups, and Daemi, Lidman, Vela-Vick, and Wong [3, Proposition 2.1] showed that the induced map $(\iota_-)^* \colon R_N(W) \to R_N(Y_-)$ is surjective. Both of these results build on work of Gerstenhaber and Rothaus: the first statement uses [4, Theorem 2], using the residual finiteness of $\pi_1(Y_-)$, and the second statement builds on [4, Theorem 1], where it is essential that the Lie group that Agol considers, the group $\mathrm{SO}(N)$, is compact.

At this stage, $R_N(Y_-)$ and $R_N(Y_+)$ have a priori been considered using different presentations, but a sequence of Tietze moves between the presentations induces a polynomial isomorphism $\phi \colon R_N(Y_-) \to R_N(Y_+)$ between these. Together with the statement in the last paragraph, one obtains a surjective polynomial map $\phi \circ (\iota_-)^* \colon R_N(W) \to R_N(Y_+)$. At this stage Agol uses the following algebraic-geometric lemma [1, Lemma A.2]: if X and Z are real algebraic sets, with $X \subseteq Z$, and if there is a surjective polynomial map $\varphi \colon X \to Z$, then $X = Z$. Applied to our problem, this implies that $R_N(Y_+) = R_N(W)$, induced by the inclusion $(\iota_+)^* \colon R_N(W) \to R_N(Y_+)$. Finally, by using residual finiteness of $\pi_1(Y_+)$ again, and by the fact that any finite group embeds into some $\mathrm{SO}(N)$ for sufficiently large N, we conclude as in Agol's situation that $(\iota_+)_* \colon \pi_1(Y_+) \to \pi_1(W)$ is an isomorphism, using $R_N(Y_+) = R_N(W)$. □

Proof of Corollary 1.6 Suppose that W_0 is a ribbon $\mathbb{Q}$-homology cobordism from Y_0 to Y_1, and that W_1 is a ribbon $\mathbb{Q}$-homology cobordism from Y_1 to Y_0. We denote by $\iota_{10} \colon Y_1 \to W_0$ the natural inclusion. We form a $\mathbb{Q}$-homology ribbon cobordism from $Y_- := Y_1$ to $Y_+ := Y_1$ by gluing W_0 and W_1 along Y_0, and we denote by $\iota_+ \colon Y_+ \to W$ the natural inclusion. Finally, we denote by $j \colon W_0 \to W$ the natural inclusion. Then clearly $\iota_+ = j \circ \iota_{10}$. This induces a commutative diagram between fundamental groups:

(3)
$$\begin{tikzcd} \pi_1(Y_1) \arrow[d, "(\iota_{10})_*"'] \arrow[dr, "(\iota_+)_*"] & \\ \pi_1(W_0) \arrow[r, "j_*"] & \pi_1(W) \end{tikzcd}$$

By Theorem 1.5, the map $(\iota_+)_*$ is an isomorphism. Since W_0 is a ribbon $\mathbb{Q}$-homology cobordism from Y_0 to Y_1 it follows from Proposition 2.1(1) that the map $(\iota_{10})_*$ is a surjection. By commutativity of the diagram we conclude that $(\iota_{10})_*$ is a monomorphism. Thus in summary it is in fact an isomorphism. □

References

[1] **I Agol**, *Ribbon concordance of knots is a partial ordering*, Commun. Am. Math. Soc. 2 (2022) 374–379 MR Zbl

[2] **M Aschenbrenner**, **S Friedl**, **H Wilton**, *3-manifold groups*, Eur. Math. Soc., Zürich (2015) MR Zbl

[3] **A Daemi**, **T Lidman**, **D S Vela-Vick**, **C-M M Wong**, *Ribbon homology cobordisms*, Adv. Math. 408 (2022) art. id. 108580 MR Zbl

[4] **M Gerstenhaber**, **O S Rothaus**, *The solution of sets of equations in groups*, Proc. Nat. Acad. Sci. U.S.A. 48 (1962) 1531–1533 MR Zbl

[5] **C M Gordon**, *Ribbon concordance of knots in the 3-sphere*, Math. Ann. 257 (1981) 157–170 MR Zbl

[6] **J Hempel**, *Residual finiteness for 3-manifolds*, from "Combinatorial group theory and topology", Ann. of Math. Stud. 111, Princeton Univ. Press (1987) 379–396 MR Zbl

[7] **M Huber**, *Ribbon cobordisms between lens spaces*, Pacific J. Math. 315 (2021) 111–128 MR Zbl

[8] **M Huber**, *Ribbon cobordisms as a partial order*, Math. Res. Lett. 30 (2023) 1511–1519 MR Zbl

[9] **P Orlik**, *Seifert manifolds*, Lecture Notes in Math. 291, Springer (1972) MR Zbl

[10] **P Scott**, *There are no fake Seifert fibre spaces with infinite π_1*, Ann. of Math. 117 (1983) 35–70 MR Zbl

[11] **F Waldhausen**, *On irreducible 3-manifolds which are sufficiently large*, Ann. of Math. 87 (1968) 56–88 MR Zbl

Fakultät für Mathematik, Universität Regensburg
Regensburg, Germany

sfriedl@gmail.com, filip.misev@mathematik.uni-regensburg.de,
raphael.zentner@mathematik.uni-regensburg.de

Received: 1 December 2022 Revised: 26 August 2023

Algebraic & Geometric Topology **25**:1 (2025) 255–266
DOI: 10.2140/agt.2025.25.255
Published: 5 March 2025

A cubulation with no factor system

SAM SHEPHERD

The primary method for showing that a given cubulated group is hierarchically hyperbolic is by constructing a factor system on the cube complex. We show that such a construction is not always possible, namely we construct a cubulated group for which the cube complex does not have a factor system. We also construct a cubulated group for which the induced action on the contact graph is not acylindrical.

20F65; 20F67

1 Introduction

A *cubulated group* $G \curvearrowright X$ is a group G together with a proper cocompact action of G on a CAT(0) cube complex X (and if G is fixed then each such action is called a *cubulation of* G). Numerous groups can be cubulated, including small cancellation groups, finite volume hyperbolic 3-manifold groups and many Coxeter groups — see Wise [15] for further background and examples. In turn, many cubulated groups are examples of hierarchically hyperbolic groups (HHGs), a class of groups that includes hyperbolic groups, relatively hyperbolic groups and mapping class groups among others — see Behrstock, Hagen and Sisto [3; 4] for relevant definitions and background. The primary method for showing that a given cubulated group $G \curvearrowright X$ is an HHG is by constructing a certain family of subcomplexes of X, called a factor system, which we define below following [3]. Many cubulated groups are known to have factor systems, including virtually special cubulated groups [3, Proposition B] — see also Hagen and Susse [9].

Definition 1.1 Let X be a CAT(0) cube complex. Each hyperplane H in X has an associated carrier $H \times [-1, 1] \subset X$, and we call the convex subcomplexes $H \times \{\pm 1\}$ *combinatorial hyperplanes*. For a convex subcomplex $K \subset X$ we let $\mathfrak{g}_K \colon X \to K$ denote the closest point projection to K. A collection $\mathfrak{F}$ of subcomplexes of X is called a *factor system* if it satisfies the following:

(1) $X \in \mathfrak{F}$.
(2) Each $F \in \mathfrak{F}$ is a nonempty convex subcomplex of X.
(3) There exists $\Delta \geq 0$ such that for all $x \in X^{(0)}$, at most Δ elements of $\mathfrak{F}$ contain x.
(4) Every nontrivial convex subcomplex parallel to a combinatorial hyperplane of X is in $\mathfrak{F}$.
(5) There exists $\xi \geq 0$ such that for any pair of subcomplexes $F, F' \in \mathfrak{F}$, either $\mathfrak{g}_F(F') \in \mathfrak{F}$ or $\operatorname{diam}(\mathfrak{g}_F(F')) < \xi$.

Our first theorem is as follows, which answers a question of Behrstock, Hagen and Sisto [3, Question 8.13].

Theorem 1.2 *There is a cubulated group $G \curvearrowright X$ such that X does not have a factor system.*

Hagen and Susse [9, Theorem A] provided three separate sufficient conditions for a cubulated group $G \curvearrowright X$ to admit a factor system:

(1) the action is rotational,

(2) it satisfies the weak finite height condition for hyperplanes, and

(3) it satisfies the essential index condition together with the Noetherian intersection of conjugates condition on hyperplane stabilizers.

Theorem 1.2 gives the first known example of a cubulated group that fails all of these conditions (see Remark 2.4 for more on this). The example behind Theorem 1.2 also contains pairs of hyperplanes that are L-well-separated but not $(L-1)$-well-separated for arbitrarily large L (Remark 2.3), which provides a negative answer to a question of Genevois [7, Question 6.69 (first part)].

Associated to a CAT(0) cube complex X is the *contact graph* $\mathscr{C}X$: the vertices are the hyperplanes of X, and edges correspond to pairs of hyperplanes whose carriers intersect (equivalently, pairs of hyperplanes that are not separated by a third hyperplane). The contact graph is always a quasitree — see Hagen [8] — so in particular it is hyperbolic. Moreover, the contact graph is a key ingredient of the HHG structure that one usually builds for cubulated groups. More precisely, if a cubulated group $G \curvearrowright X$ has a G-invariant factor system $\mathfrak{F}$, then one can build an HHG structure for G by taking the contact graph $\mathscr{C}F$ of each $F \in \mathfrak{F}$ and coning off certain subgraphs of $\mathscr{C}F$ that correspond to smaller elements of $\mathfrak{F}$ — see [3] for details. The existence of a factor system for $G \curvearrowright X$ also implies that the induced action of G on $\mathscr{C}X$ is acylindrical [3, Theorem D]. (Recall that the action of a group G on a metric space (M,d) is *acylindrical* if for all $\epsilon > 0$ there exist $R, N > 0$ such that $d(x,y) \geq R$ implies that there are at most N elements $g \in G$ satisfying $d(x,gx), d(y,gy) < \epsilon$.) The following theorem is therefore a strengthening of Theorem 1.2.

Theorem 1.3 *There is a cubulated group $G \curvearrowright X$ for which the induced action on the contact graph $\mathscr{C}X$ is not acylindrical.*

This theorem is even more surprising in light of Genevois [6, Theorem 1.1], which implies that every cubulated group $G \curvearrowright X$ has a *nonuniformly acylindrical* action on $\mathscr{C}X$ (*nonuniformly* meaning that "at most N elements" is replaced by "finitely many elements" in the definition of acylindrical).

Briefly, the construction for Theorem 1.2 is to take a free cocompact action of a group Γ on a product of trees $T_1 \times T_2$ that contains an antitorus, and then Γ-equivariantly attach infinite strips to $T_1 \times T_2$ along antitori axes. The details are in Section 2. The construction for Theorem 1.3 builds on this by defining a certain HNN extension $\Lambda = \Gamma *_{\mathbb{Z}}$, and an action of Λ on a cube complex that splits as a tree of spaces, with vertex spaces being copies of $T_1 \times T_2$ and edge spaces corresponding to the infinite strips described above. The arguments for this are in Section 3. Although Γ admits a cubulation without a factor system, it is still an HHG because $\Gamma \curvearrowright T_1 \times T_2$ is another cubulation that does have a factor system. On the other hand, we do not know whether our second group Λ admits a cubulation with a factor system, and we do not know whether Λ is an HHG. In particular, the question of whether all cubulated groups are HHGs is still open [4, Question A]. One possible strategy is to find a cubulated group with no largest acylindrical action (see Definition 3.4), since all HHGs have a largest acylindrical action — see Abbott, Behrstock and Durham [2]. This does not work for the group Λ however, as we prove in Proposition 3.6 that Λ does have a largest acylindrical action.

Acknowledgements I am grateful for Mark Hagen's suggestion of considering antitori, which made my construction for Theorem 1.2 more general. Thanks go to Anthony Genevois for his comments, in particular regarding Remark 3.3. And I thank the referee for their comments and corrections. I am also thankful for the support of the Institut Henri Poincaré (UAR 839 CNRS-Sorbonne Université), and LabEx CARMIN (ANR-10-LABX-59-01).

2 Example with no factor system

Let T_1 and T_2 be locally finite trees, and let Γ be a group acting freely and cocompactly on $T_1 \times T_2$. Suppose that elements $g_1, g_2 \in \Gamma$ form an *antitorus*, meaning firstly that they translate nontrivially along intersecting axes $\ell_1 \times \{p_2\}, \{p_1\} \times \ell_2 \subset T_1 \times T_2$ respectively (so $p_1 \in \ell_1$ and $p_2 \in \ell_2$), and secondly that no nonzero powers of g_1 and g_2 commute. In addition, suppose that g_1 and g_2 are not proper powers in Γ. The condition that no powers of g_1 and g_2 commute is equivalent to saying that the flat $\Pi = \ell_1 \times \ell_2$ is not periodic. We also note that the existence of an antitorus implies that Γ is *irreducible* [10, Lemma 18], meaning it does not have a finite-index subgroup that splits as a product $\Gamma_1 \times \Gamma_2$ with Γ_i acting trivially on T_{3-i}. Examples of antitori were constructed by Wise [14], Janzen and Wise [10] and Rattaggi [12]. The smallest example is in [10], where $(T_1 \times T_2)/\Gamma$ consists of one vertex, four edges and four 2-cells. See [5] for more about antitori and irreducible lattices in products of trees.

Choose orientations for the edges in the finite quotient $(T_1 \times T_2)/\Gamma$, and label them with distinct letters from an alphabet A. Lift this labeling to $T_1 \times T_2$. Each finite edge path in $T_1 \times T_2$ or its quotient is thus labeled by some word w on $A^{\pm}$, and we denote the length of w by $|w|$. The axes $\ell_1 \times \{p_2\}$ and $\{p_1\} \times \ell_2$ descend to loops in $(T_1 \times T_2)/\Gamma$ based at $\Gamma \cdot (p_1, p_2)$; say these loops are labeled by words w_1 and w_2 respectively. Lifts of the w_1-loop to $T_1 \times T_2$ will be referred to as w_1*-geodesics* (equivalently, these are Γ-translates of $\ell_1 \times \{p_2\}$).

Lemma 2.1 *For any $n \geq 1$ there exists a w_1-geodesic whose intersection with Π is a finite path of the form $\gamma \times \{y\} \subset \ell_1 \times \ell_2$, with $p_1 \in \gamma$ and γ of length at least n.*

Proof For each $i \geq 1$, consider the rectangle in Π with two sides labeled by w_1^n and w_2^i that meet at the vertex (p_1, p_2), as shown in Figure 1. Note that the bottom side is a subpath of the axis $\ell_1 \times \{p_2\}$, while

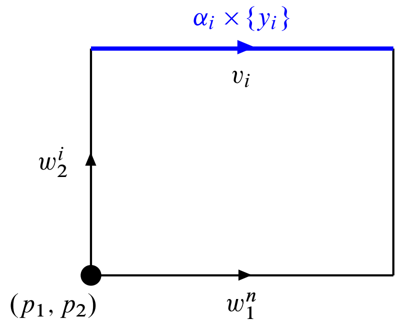


Figure 1: Rectangle in Π with two sides labeled by w_1^n and w_2^i that meet at the vertex (p_1, p_2).

the left-hand side is a subpath of the axis $\{p_1\} \times \ell_2$. Let $\alpha_i \times \{y_i\}$ denote the top side, and suppose that it is labeled by the word v_i. There are only finitely many words of length $|w_1^n|$, so $v_i = v_{i+j}$ for some $i, j \geq 1$. Since $\alpha_i \times \{y_i\}$ and $\alpha_{i+j} \times \{y_{i+j}\}$ are both labeled by v_i, the element g_2^j maps $\alpha_i \times \{y_i\}$ to $\alpha_{i+j} \times \{y_{i+j}\}$. Moreover, g_2^j preserves the axis $\{p_1\} \times \ell_2$, so it maps the rectangle shown in Figure 1 to another rectangle in Π. Restricting to the bottom sides of the rectangles, we see that g_2^j maps the subpath of $\ell_1 \times \{p_2\}$ labeled w_1^n to $\alpha_j \times \{y_j\}$, so $v_j = w_1^n$.

The path $\alpha \times \{y\} := \alpha_j \times \{y_j\}$ extends to a unique w_1-geodesic. Let $\gamma \times \{y\} \subset \ell_1 \times \ell_2$ denote the intersection of this w_1-geodesic with Π. Since $p_1 \in \alpha \subset \gamma$, and since α has length $|w_1^n| \geq n$, it remains to show that γ is finite. Say that p_1 splits γ into subpaths γ_1 and γ_2, with α being an initial segment of γ_2. We will show that γ_2 is finite — finiteness of γ_1 follows by a similar argument.

For $k \geq n$, consider the rectangle in Π with two sides labeled by w_1^k and w_2^j that meet at the vertex (p_1, p_2), as shown in Figure 2. Let $\beta_k \times \{y\}$ denote the top side. Note that $\alpha \times \{y\}$ is an initial segment of $\beta_k \times \{y\}$. Say the right-hand side is labeled by the word v'_k. The same argument we used earlier in the proof shows that $v'_k = w_2^j$ for some k. For this k, we then argue that γ_2 has length less than $|w_1^k|$. Indeed, otherwise $\gamma_2 \times \{y\}$ would contain $\beta_k \times \{y\}$ as an initial segment, so $\beta_k \times \{y\}$ would be labeled by w_1^k, but then the labels on the rectangle would imply that g_1^k and g_2^j commute, contradicting the fact that g_1 and g_2 form an antitorus. Thus γ_2 is finite, as required. □

To construct a cubulation of Γ with no factor system we first take the quotient $(T_1 \times T_2)/\Gamma$, then we attach an annulus by gluing one boundary component along the edge loop labeled by w_1, and then we let X be the universal cover. If the annulus is subdivided into $|w_1|$ squares then X is a CAT(0) cube complex. Attaching the annulus to $(T_1 \times T_2)/\Gamma$ doesn't change the fundamental group, so Γ acts on X by deck transformations. The picture upstairs is that X is obtained from $T_1 \times T_2$ by attaching an infinite strip to each w_1-geodesic (and only one strip since g_1 is not a proper power). We already remarked that $(T_1 \times T_2)/\Gamma$ has only four 2-cells for the example in [10]; moreover the word w_1 has length two in this case, so the cube complex X/Γ would be a $\mathcal{VH}$-complex consisting of just six 2-cells.

Theorem 2.2 *X has no factor system.*

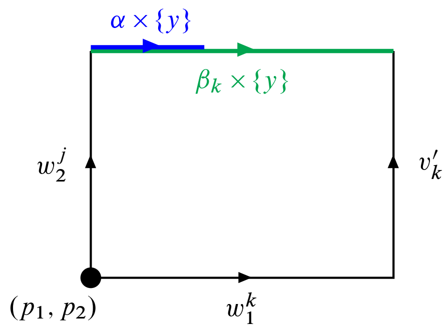


Figure 2: Rectangle in Π with two sides labeled by w_1^k and w_2^j that meet at the vertex (p_1, p_2).

Proof Suppose for contradiction that X has a factor system $\mathfrak{F}$. In X, there is an infinite strip glued to each w_1-geodesic, and there is a hyperplane that runs along the middle of each infinite strip (shown as dotted red lines in Figure 3). Hence each w_1-geodesic is a combinatorial hyperplane in X, and is an element of $\mathfrak{F}$ by Definition 1.1(4). In particular $F := \ell_1 \times \{p_2\} \in \mathfrak{F}$.

Choose an integer $n \geq 1$ and apply Lemma 2.1. This provides us with a w_1-geodesic F' whose intersection with Π is a finite path of the form $\gamma \times \{y\} \subset \ell_1 \times \ell_2 = \Pi$, with $p_1 \in \gamma$ and γ of length at least n. The projection $\mathfrak{g}_F \colon X \to F$ maps $\gamma \times \{y\}$ to $\gamma \times \{p_2\}$, and it maps the rest of F' to the endpoints of $\gamma \times \{p_2\}$. By Definition 1.1(5), the image $\mathfrak{g}_F(F') = \gamma \times \{p_2\}$ is in $\mathfrak{F}$ for sufficiently large n. But we have $(p_1, p_2) \in \mathfrak{g}_F(F')$ for all n, so we contradict Definition 1.1(3). □

Remark 2.3 If H and H' are the hyperplanes that run along the strips glued to F and F' from the above proof (see Figure 3), then the hyperplanes that are transverse to both H and H' are precisely the hyperplanes that cross the path $\gamma \times \{y\}$. Moreover, this collection of hyperplanes has no facing triples, so if γ has length L then H and H' are L-well-separated but not $(L-1)$-well-separated. We can choose H and H' so that L is arbitrarily large, so this provides a negative answer to a question of Genevois [7, Question 6.69 (first part)].

Remark 2.4 Hagen and Susse [9, Theorem A] provided three separate sufficient conditions for a cubulated group to admit a factor system. Since X has no factor system, we know that $\Gamma \curvearrowright X$ does not satisfy any of these three conditions. We describe what these conditions are below, and outline some direct arguments for why they fail for $\Gamma \curvearrowright X$:

(1) A cubulated group $G \curvearrowright Z$ is *rotational* if for each hyperplane B there is a finite-index subgroup $K_B < \mathrm{Stab}_G(B)$ such that, for any hyperplane A disjoint from B and all $k \in K_B$, the carriers of A and kA are either equal or disjoint.

For our cubulated group $\Gamma \curvearrowright X$, we can consider $B = H$ to be the hyperplane shown in Figure 3, and we can show that for any $1 \neq k \in \mathrm{Stab}_\Gamma(H)$ there is a hyperplane A disjoint from H such that the carriers of A and kA are distinct but not disjoint. Indeed, for $1 \neq k \in \mathrm{Stab}_\Gamma(H) = \langle g_1 \rangle$ we know that

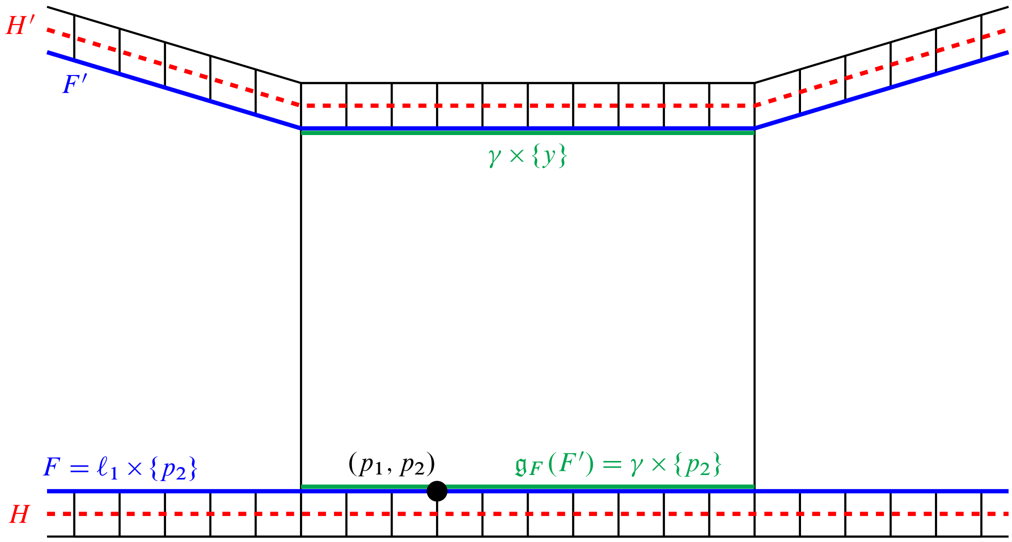


Figure 3: The w_1-geodesics F and F' with their attached strips.

$k(\{p_1\} \times \ell_2) \cap \Pi$ is a finite path (else k would commute with some power of g_2 by a similar argument to the proof of Lemma 2.1, contradicting the fact that g_1 and g_2 form an antitorus), so there must be a vertex $x \in \ell_2$ incident to an edge $e \subset \ell_2$ such that $k(p_1, x) \in \ell_1 \times \{x\}$ but $k(\{p_1\} \times e) \not\subseteq \Pi$. With A being the hyperplane dual to $\{p_1\} \times e$, the intersection of the carriers of A and kA is $T_1 \times \{x\}$, which is a proper subset of the carrier of A (it is one of the combinatorial hyperplanes of A).

(2) A cubulated group $G \curvearrowright Z$ satisfies the *weak finite height condition for hyperplanes* if the following holds for each hyperplane A and its stabilizer $K = \mathrm{Stab}_G(A)$: if $\{g_i\} \subset G$ is an infinite set such that $K \cap \bigcap_{i \in J} K^{g_i}$ is infinite for all finite $J \subset I$, then there exist distinct g_i and g_j such that $K \cap K^{g_i} = K \cap K^{g_j}$.

This condition fails for our cubulated group $\Gamma \curvearrowright X$ (in fact it also fails for $\Gamma \curvearrowright T_1 \times T_2$) by considering an edge $e \subset \ell_2$ and taking A to the hyperplane dual to $\{p_1\} \times e$. Then the stabilizer $K = \mathrm{Stab}_\Gamma(A)$ is just the stabilizer of e with respect to the action $\Gamma \curvearrowright T_2$. For any $i \geq 1$ we know that $g_1^i(\{p_1\} \times \ell_2) \cap \Pi$ is finite (as in case (1)), or equivalently $g_1^i \ell_2 \cap \ell_2 \subset T_2$ is finite. The element g_2 translates along ℓ_2, so for any power g_2^j the conjugate $K^{g_2^j}$ is the Γ-stabilizer of the edge $g_2^j e \subset \ell_2$. Thus $g_1^i \notin K^{g_2^j}$ for all sufficiently large $j \geq 1$. Since T_2 is locally finite, K and $K^{g_2^j}$ are commensurable in Γ for any j, and $\langle g_1 \rangle \cap K^{g_2^j}$ is infinite since g_1 fixes the vertex $p_2 \in T_2$. Hence, we may construct increasing sequences of positive integers (i_k) and (j_k) such that $g_1^{i_k}$ is not in $K^{g_2^{j_k}}$ but $g_1^{i_{k+1}}$ is. Therefore,

$$K \cap K^{g_2^{j_1}} \supsetneq K \cap K^{g_2^{j_2}} \supsetneq K \cap K^{g_2^{j_3}} \supsetneq \cdots$$

is a strictly descending chain of commensurable (hence infinite) subgroups of Γ, which means the weak finite height condition for hyperplanes does not hold for $\Gamma \curvearrowright X$.

(3) The third condition has two parts. A cubulated group $G \curvearrowright Z$ satisfies the *essential index condition* if there is a constant ζ such that for any $F \in \mathfrak{F}$ (where $\mathfrak{F}$ is the smallest collection of convex subcomplexes of Z that contains Z, contains all combinatorial hyperplanes, and is closed under closest point projection) the G-stabilizer of F has index at most ζ in the G-stabilizer of the essential core of F. A cubulated group $G \curvearrowright Z$ satisfies the *Noetherian intersection of conjugates* (*NIC*) *condition on hyperplane stabilizers* if the following holds for each hyperplane stabilizer K: given $\{g_i\} \subset G$ such that $K_n = K \cap \bigcap_{i=0}^{n} K^{g_i}$ is infinite for all n, there exists l such that K_n and K_l are commensurable for $n \geq l$.

Our cubulated group $\Gamma \curvearrowright X$ satisfies the NIC condition for hyperplane stabilizers (because all hyperplane stabilizers are either cyclic or commensurated) but fails the essential index condition as follows. Suppose the geodesic $\ell_1 \times \{p_2\}$ (from Figure 3) crosses the hyperplanes $H_1, H_2, H_3, \ldots$, respectively, when starting at (p_1, p_2) and moving in the direction of translation of g_1. Let F_i be the combinatorial hyperplane of H_i that is on the same side of H_i as (p_1, p_2). Let $\mathfrak{F}$ be as described above for our cubulated group $\Gamma \curvearrowright X$, and consider the subcomplexes $\mathfrak{g}_{F_1}(F_i) \in \mathfrak{F}$. Given an integer $n \geq 1$, as in the proof of Theorem 2.2 we can apply Lemma 2.1 to obtain a w_1-geodesic F' whose intersection with Π is a finite path of the form $\gamma \times \{y\} \subset \ell_1 \times \ell_2 = \Pi$, with $p_1 \in \gamma$ and γ of length at least n. Moreover, it follows from the proof of Lemma 2.1 that we can take $(p_1, y) = g_2^k(p_1, p_2)$ for some $k \neq 0$, and we may assume that the subpath of $\gamma \times \{y\}$ that starts at (p_1, y) and moves in the direction of translation of $g_2^k g_1 g_2^{-k}$ has length at least n. Let e and e' be the edges in F_1 incident at vertices (p_1, p_2) and (p_1, y), respectively, that cross the hyperplanes H and H' respectively (again from Figure 3). Note that $g_2^k e = e'$. One can show that $e \subset \mathfrak{g}_{F_1}(F_i)$ for all i. On the other hand, $e' \subset \mathfrak{g}_{F_1}(F_i)$ for $1 \leq i \leq n$ but for only finitely many i in total. One can then argue that g_2^k is in the Γ-stabilizer of $\mathfrak{g}_{F_1}(F_i)$ for $1 \leq i \leq n$ but for only finitely many i in total. This can be done for any $n \geq 1$, so the Γ-stabilizers of the $\mathfrak{g}_{F_1}(F_i)$ is a descending sequence of subgroups that never terminates. Meanwhile, all the $\mathfrak{g}_{F_1}(F_i)$ have essential core $\{p_1\} \times T_2$, so the essential index condition fails.

3 Example with nonacylindrical action on the contact graph

We now construct a free cocompact action of a group Λ on a CAT(0) cube complex Y, such that the induced action on the contact graph $\mathscr{C}Y$ is not acylindrical. We start with the action of Γ on $T_1 \times T_2$ from Section 2, with elements $g_1, g_2 \in \Gamma$ forming an antitorus. We retain all the notation from Section 2, so in particular g_1 translates along an axis $\ell_1 \times \{p_2\} \subset T_1 \times T_2$ which descends to a loop in $(T_1 \times T_2)/\Gamma$ labeled by w_1. Next we attach an annulus to $(T_1 \times T_2)/\Gamma$ by gluing both its boundary components (with matching orientations) along the loop labeled by w_1. We make this a nonpositively curved cube complex by subdividing the annulus into $|w_1|$ squares, and we let Y be the universal cover. Unlike for the construction of X in Section 2, we have glued *both* boundary components of the annulus to $(T_1 \times T_2)/\Gamma$, so the gluing changes the fundamental group from Γ to the HNN extension

$$\Lambda := \Gamma *_{\langle g_1 \rangle} = \langle \Gamma, t \mid t g_1 t^{-1} = g_1 \rangle, \tag{3-1}$$

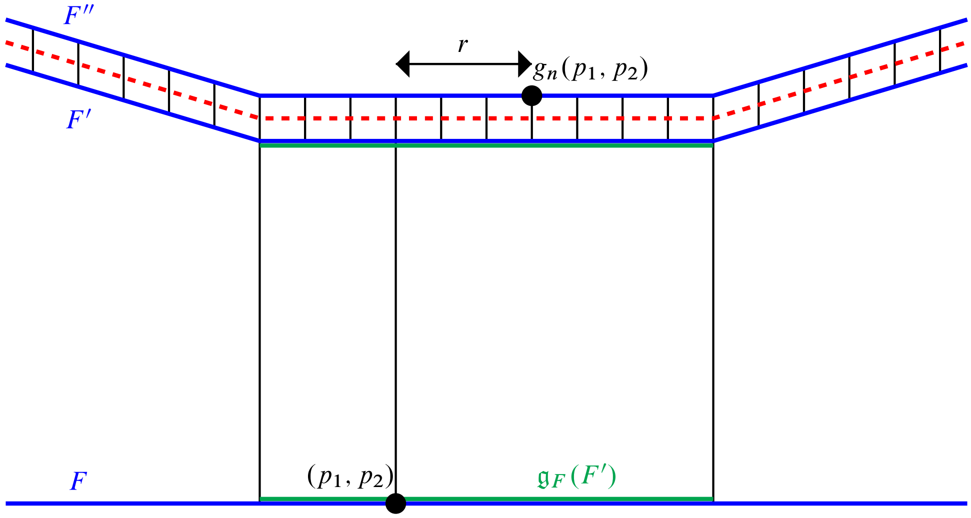


Figure 4: The positioning of $g_n(p_1, p_2)$.

and Λ acts on Y by deck transformations. Observe that Y has the structure of a tree of spaces, where the vertex spaces are copies of $T_1 \times T_2$, and the edge spaces are infinite strips. The edge labeling of $(T_1 \times T_2)/\Gamma$ induces an edge labeling of Y/Λ (apart from the edges that cross the annulus) and we can lift this to an edge labeling of Y. As in Section 2, lifts of the w_1-loop in Y/Λ to Y will be referred to as w_1-*geodesics*. Each w_1-geodesic in Y is attached to two edge spaces since the w_1-loop in Y/Λ is attached to both boundary components of the annulus.

Theorem 3.1 *The action of Λ on the contact graph $\mathscr{C}Y$ is not acylindrical.*

Proof We will show that for any $R, N > 0$ there exist $H, H' \in \mathscr{C}Y$ such that $d_{\mathscr{C}Y}(H, H') \geq R$ and there are more than N elements $g \in \Lambda$ satisfying $d_{\mathscr{C}Y}(H, gH), d_{\mathscr{C}Y}(H', gH') \leq 2$.

Consider a vertex space in Y, and identify it with $T_1 \times T_2$. Given an integer $n \geq 1$, we can choose w_1-geodesics F and F' as in the proof of Theorem 2.2 such that the projection $\mathfrak{g}_F(F')$ is a finite path of length at least n that contains the vertex (p_1, p_2). Now take one of the edge spaces in Y glued to F', and let F'' be the geodesic on the other side of the edge space. F'' is in a different vertex space, but it is again a w_1-geodesic, so it contains vertices in the orbit $\Lambda \cdot (p_1, p_2)$. Moreover, the spacing between these vertices is at most $|w_1|$, so we can choose $g_n \in \Lambda$ such that $g_n(p_1, p_2)$ lies on F'' but is shifted to the right relative to (p_1, p_2) by an integer $0 < r \leq |w_1|$, as shown in Figure 4.

Note that $g_n F = F''$. Furthermore, applying powers of g_n to F and F' produces a staircase-like picture as shown in Figure 5, where each step has depth r. Each pair $g_n^i F, g_n^i F'$ lies in a different vertex space of Y. If H is the hyperplane that runs along the edge space between $g_n^{-1} F'$ and F, then the hyperplanes $g_n^i H$ run along a sequence of edge spaces that connect the aforementioned sequence of vertex spaces.

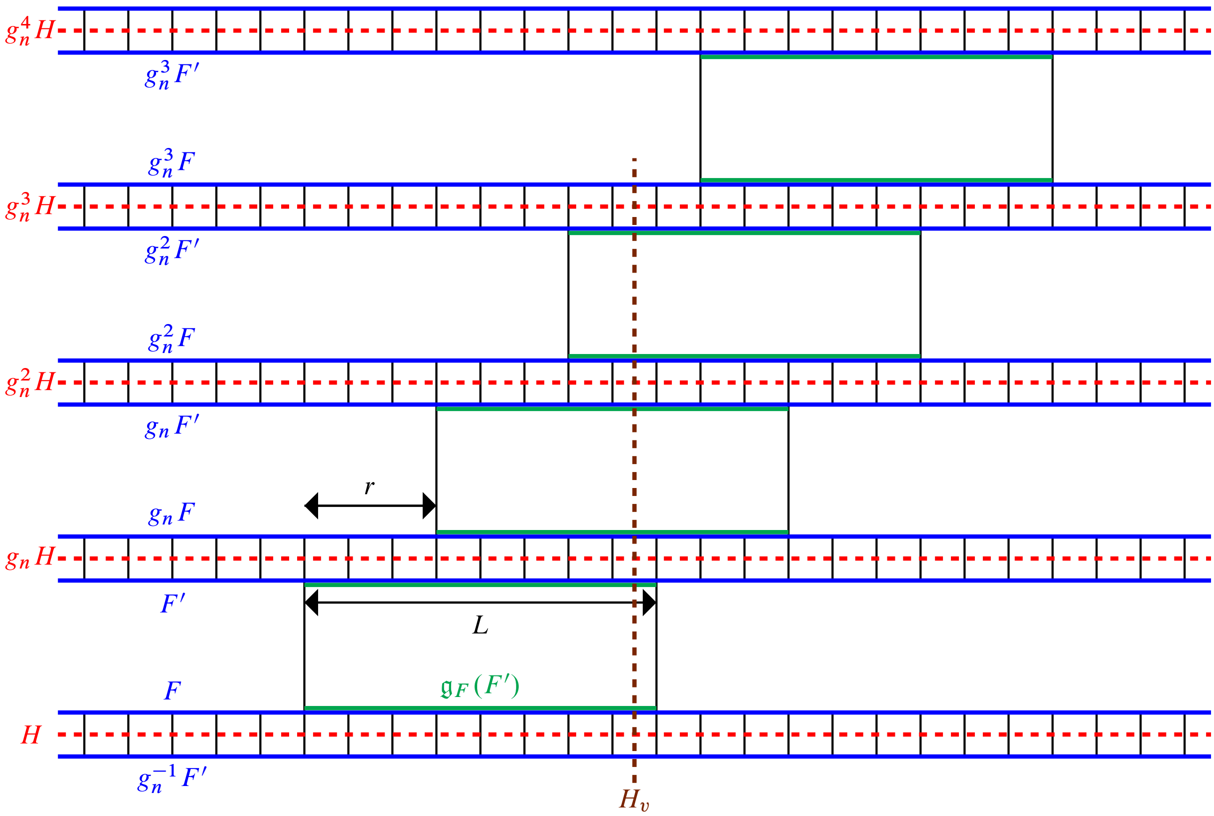


Figure 5: The arrangement of the w_1-geodesics $g_n^i F$ and $g_n^i F'$.

Suppose that the path $\mathfrak{g}_F(F')$ has length L (remembering that $L \geq n$). If a hyperplane intersects more than one of the hyperplanes $g_n^i H$ then it must cross some $\langle g_n \rangle$-translate of the projection $\mathfrak{g}_F(F')$. The hyperplane H_v that is dual to the last edge of $\mathfrak{g}_F(F')$ intersects exactly $M := \lceil L/r \rceil + 1$ of the hyperplanes $g_n^i H$, and no other hyperplane intersects more of them. In particular, for $1 \leq i < M$, H_v intersects H and $g_n^i H$, so the distance between H and $g_n^i H$ in the contact graph $\mathscr{C}Y$ is

$$d_{\mathscr{C}Y}(H, g_n^i H) = 2. \tag{3-2}$$

On the other hand, $d_{\mathscr{C}Y}(H, g_n^p H) \to \infty$ as $p \to \infty$. Indeed, suppose the geodesic in $\mathscr{C}Y$ from H to $g_n^p H$ consists of hyperplanes $H = H_0, H_1, \ldots, H_d = g_n^p H$. We know that H and $g_n^p H$ are separated by the hyperplanes $g_n^i H$ for each $0 < i < p$, so each of these $g_n^i H$ either equals one of the H_j or intersects one of the H_j. But we know from earlier that each H_j intersects at most M of the $g_n^i H$; hence

$$d = d_{\mathscr{C}Y}(H, g_n^p H) \geq \frac{p}{M}.$$

Putting $H' = g_n^p H$, we have $d_{\mathscr{C}Y}(H, H') \geq R$ provided $p \geq MR$. But (3-2) implies that H and H' are both moved at most distance 2 in $\mathscr{C}Y$ by the elements $\{1, g_n, g_n^2, \ldots, g_n^{M-1}\} \subset \Lambda$; and $M = \lceil L/r \rceil + 1 > n/r$, so if we choose $n \geq rN$ then we get more than N elements $g \in \Lambda$ satisfying $d_{\mathscr{C}Y}(H, gH), d_{\mathscr{C}Y}(H', gH') \leq 2$, as required. So we conclude that the action of Λ on $\mathscr{C}Y$ is not acylindrical. □

Remark 3.2 The staircase in Figure 5 is not a staircase as defined in [9] (visually speaking, the former looks like it has empty space below the staircase whereas the latter does not). However, they are both obstructions to the existence of a factor system. Indeed, the existence of staircases as in Figure 5 for arbitrarily large ratios L/r is the key to proving Theorem 3.1, which in turn implies that Y has no factor system. Meanwhile, the existence of just a single convex staircase in the sense of [9] rules out the possibility of a factor system. It remains an open question whether any cubulation of a group contains a convex staircase in the sense of [9].

Remark 3.3 Genevois [7] defined a metric δ_K for a CAT(0) cube complex X as the maximal number of pairwise K-well-separated hyperplanes separating two given vertices. The space (X, δ_K) is hyperbolic for all K, and it is quasi-isometric to the contact graph for $K = 0$. Moreover, if $G \curvearrowright X$ is a cubulated group, then the induced action on (X, δ_K) is nonuniformly acylindrical for all K. However, for our cubulated group $\Lambda \curvearrowright Y$, the action of Λ on (Y, δ_K) is not acylindrical for any K. The argument is similar to that in the proof of Theorem 3.1: you can define the element $g_n \in \Lambda$ in the same way, and show that g_n acts loxodromically on (Y, δ_K), and you can exhibit points on the axis of g_n that are arbitrarily far apart but are moved at most distance 2 by many powers of g_n — with the number of such powers tending to infinity as $n \to \infty$.

As discussed in the introduction, we do not know whether Λ is an HHG. A possible strategy to prove that Λ is not an HHG would be to show that Λ has no largest acylindrical action (definition below), because all HHGs possess a largest acylindrical action [2].

Definition 3.4 Let G be a group that acts on metric spaces R and S. We say that $G \curvearrowright R$ is *dominated* by $G \curvearrowright S$, written $G \curvearrowright R \preceq G \curvearrowright S$, if there exist $r \in R$, $s \in S$ and a constant C such that

$$d_R(r, gr) \leq C d_S(s, gs) + C$$

for all $g \in G$. The actions $G \curvearrowright R$ and $G \curvearrowright S$ are *equivalent* if $G \curvearrowright R \preceq G \curvearrowright S$ and $G \curvearrowright S \preceq G \curvearrowright R$. We denote the equivalence class by $[G \curvearrowright R]$. The relation $\preceq$ defines a partial order on the set of equivalence classes of actions of G on metric spaces. The *largest acylindrical action* of G (if it exists) is the largest element of the set of equivalence classes of cobounded acylindrical actions of G on hyperbolic metric spaces.

Alas, we show below in Proposition 3.6 that Λ does have a largest acylindrical action, which is defined as follows. Let T be the Bass–Serre tree for the splitting $\Lambda = \Gamma *_{\langle g_1 \rangle}$. We say that edges $e_1, e_2 \in ET$ are *equivalent* if the stabilizers Λ_{e_1} and Λ_{e_2} are commensurable. Each equivalence class defines a subtree of T called a *cylinder*. The *tree of cylinders* T_c is the bipartite tree with vertex set $V_0T_c \sqcup V_1T_c$, where

V_0T_c are the vertices of T and V_1T_c is the set of cylinders. The edges of T_c are of the form (v, C) where v is a vertex in T that lies in the cylinder $C \subset T$. The action of Λ on T induces an action on T_c.

To help us prove Proposition 3.6 we will use the following lemma, which is equivalent to [1, Corollary 4.14].

Lemma 3.5 *Let G act cocompactly on a connected graph Δ and let $G \curvearrowright R$ be another cobounded action on a metric space. If the vertex stabilizers of Δ have bounded orbits in R, then $G \curvearrowright R \preceq G \curvearrowright \Delta$.*

Proposition 3.6 *The largest acylindrical action of Λ is its action on the tree of cylinders $\Lambda \curvearrowright T_c$.*

Proof First we show that the action of Λ on T_c is acylindrical. It suffices to show that only the identity element of Λ fixes a path in T_c of the form v_1, C_1, u, C_2, v_2, where $u, v_i \in V_0T_c$ and $C_i \in V_1T_c$. Indeed if $g \in \Lambda$ fixes such a path, then it must also fix any pair of edges $e_1, e_2 \in ET$ that lie on the geodesics $[u, v_1]$ and $[u, v_2]$ respectively. It follows that e_1 and e_2 lie in the cylinders C_1 and C_2, and so the stabilizers Λ_{e_1} and Λ_{e_2} are not commensurable. But these stabilizers are infinite cyclic (they are conjugates of $\langle g_1 \rangle$), hence they have trivial intersection, so $g \in \Lambda_{e_1} \cap \Lambda_{e_2}$ is trivial.

The action $\Lambda \curvearrowright T_c$ is clearly cobounded and T_c is obviously hyperbolic, so it remains to show that $\Lambda \curvearrowright T_c$ dominates all other cobounded acylindrical actions of Λ on hyperbolic spaces. Let $\Lambda \curvearrowright R$ be such an action. By Lemma 3.5, it suffices to show that the vertex stabilizers of T_c have bounded orbits in R.

First consider $v \in V_0T_c$. If the stabilizer Λ_v does not have bounded orbits in R then there must exist a loxodromic element $g \in \Lambda_v$ by [11, Theorem 1.1], and g must be contained in a virtually cyclic hyperbolically embedded subgroup of Λ_v by [11, Theorem 1.4]. It then follows from [13, Theorem 1] that g is Morse in Λ_v. But this is impossible since $\Lambda_v \cong \Gamma$ is quasi-isometric to a product of trees.

Now consider $C \in V_1T_c$. Without loss of generality we may assume that C contains the edge that corresponds to the subgroup $\langle g_1 \rangle < \Gamma < \Lambda$. We know that Λ_C contains the element g_1 as well as the stable letter t from the presentation (3-1), so Λ_C is not virtually cyclic. If Λ_C does not have bounded orbits in R then there must exist a loxodromic element $g \in \Lambda_C$ by [11, Theorem 1.1]. Fix a point $r \in R$. The element g_1 lies in the V_0T_c-vertex stabilizer $\Gamma < \Lambda$, so g_1 is elliptic in R by the previous paragraph, and the orbit $\langle g_1 \rangle \cdot r$ lies in some ϵ-ball about r. By definition of C, for any integer k the subgroups $\langle g_1 \rangle$ and $g^k \langle g_1 \rangle g^{-k}$ are commensurable, so have infinite intersection; and any element h of this intersection satisfies $d_R(r, hr), d_R(g^k r, hg^k r) < \epsilon$. But $d(r, g^k r) \to \infty$ as $k \to \infty$ since g is loxodromic, which contradicts the acylindricity of the action $\Lambda \curvearrowright R$. □

References

[1] **C Abbott**, **S H Balasubramanya**, **D Osin**, *Hyperbolic structures on groups*, Algebr. Geom. Topol. 19 (2019) 1747–1835 MR Zbl

[2] **C Abbott**, **J Behrstock**, **M G Durham**, *Largest acylindrical actions and stability in hierarchically hyperbolic groups*, Trans. Amer. Math. Soc. Ser. B 8 (2021) 66–104 MR Zbl

[3] **J Behrstock**, **M F Hagen**, **A Sisto**, *Hierarchically hyperbolic spaces, I: Curve complexes for cubical groups*, Geom. Topol. 21 (2017) 1731–1804 MR Zbl

[4] **J Behrstock**, **M Hagen**, **A Sisto**, *Hierarchically hyperbolic spaces, II: Combination theorems and the distance formula*, Pacific J. Math. 299 (2019) 257–338 MR Zbl

[5] **P-E Caprace**, *Finite and infinite quotients of discrete and indiscrete groups*, from "Groups St Andrews 2017 in Birmingham", Lond. Math. Soc. Lect. Note Ser. 455, Cambridge Univ. Press (2019) 16–69 MR Zbl

[6] **A Genevois**, *Acylindrical hyperbolicity from actions on* $\mathrm{CAT}(0)$ *cube complexes: a few criteria*, New York J. Math. 25 (2019) 1214–1239 MR Zbl

[7] **A Genevois**, *Hyperbolicities in* $\mathrm{CAT}(0)$ *cube complexes*, Enseign. Math. 65 (2019) 33–100 MR Zbl

[8] **M F Hagen**, *Weak hyperbolicity of cube complexes and quasi-arboreal groups*, J. Topol. 7 (2014) 385–418 MR Zbl

[9] **M F Hagen**, **T Susse**, *On hierarchical hyperbolicity of cubical groups*, Israel J. Math. 236 (2020) 45–89 MR Zbl

[10] **D Janzen**, **D T Wise**, *A smallest irreducible lattice in the product of trees*, Algebr. Geom. Topol. 9 (2009) 2191–2201 MR Zbl

[11] **D Osin**, *Acylindrically hyperbolic groups*, Trans. Amer. Math. Soc. 368 (2016) 851–888 MR Zbl

[12] **D Rattaggi**, *Anti-tori in square complex groups*, Geom. Dedicata 114 (2005) 189–207 MR Zbl

[13] **A Sisto**, *Quasi-convexity of hyperbolically embedded subgroups*, Math. Z. 283 (2016) 649–658 MR Zbl

[14] **D T Wise**, *Complete square complexes*, Comment. Math. Helv. 82 (2007) 683–724 MR Zbl

[15] **D T Wise**, *From riches to raags:* 3*-manifolds, right-angled Artin groups, and cubical geometry*, CBMS Regional Conf. Ser. Math. 117, Amer. Math. Soc., Providence, RI (2012) MR Zbl

Institut für Mathematische Logik und Grundlagenforschung
Fachbereich Mathematik und Informatik, Universität Münster
Münster, Germany

sam.shepherd@uni-muenster.de

Received: 27 February 2023 Revised: 13 June 2023

Algebraic & Geometric Topology **25**:1 (2025) 267–285
DOI: 10.2140/agt.2025.25.267
Published: 5 March 2025

Relative h-principle and contact geometry

JACOB TAYLOR

We show that if $F(M)$ is a space of holonomic solutions with space of formal solutions $F^f(M)$ that satisfies a certain relative h-principle, then the nonrelative map $F(M) \to F^f(M)$ admits a section up to homotopy. We apply this to the relative h-principle for overtwisted contact structures proved by Borman, Eliashberg and Murphy to find infinite cyclic subgroups in the homotopy groups of contactomorphism groups.

53D35, 57R17; 58D99

1 Introduction

Gromov [12] showed that if M is an open manifold then the inclusion $\mathrm{Cont}(M) \to \mathrm{AlmCont}(M)$ is a weak equivalence, where $\mathrm{Cont}(M)$ is the space of contact structures on M and $\mathrm{AlmCont}(M)$ is the space of almost contact structures on M. The case of compact manifolds is not so simple. For example, there exist contact structures on closed 3-manifolds that are formally homotopic but not homotopic; see Bennequin [1]. In [2], Matthew Borman, Yakov Eliashberg and Emmy Murphy advanced the field of contact geometry by first extending the definition of an overtwisted contact manifold from 3-dimensional manifolds to manifolds of dimension $2n+1 \geq 3$, and then proving an h-principle result for overtwisted contact manifolds. Essentially, an overtwisted contact manifold is a contact manifold M that contains an embedded overtwisted disk, ie an embedded $2n$-disk Δ with a certain model germ of a contact structure on a neighborhood of Δ (see [2, Definition 3.6]). If $\mathrm{Cont}^{\mathrm{OT}}(M,\Delta)$ and $\mathrm{AlmCont}(M,\Delta)$ denote, respectively, the spaces of contact and formal contact structures that are overtwisted with fixed disk Δ, then the main result of [2] is that

$$\mathrm{Cont}^{\mathrm{OT}}(M,\Delta) \to \mathrm{AlmCont}(M,\Delta)$$

is a weak equivalence. However, it is known that in general the map $\mathrm{Cont}^{\mathrm{OT}}(M) \to \mathrm{AlmCont}(M)$ from overtwisted contact structures to almost contact structures is not a weak equivalence; see for example Vogel [18]. Given this, one may wonder how much can be learned about the maps $\mathrm{Cont}^{\mathrm{OT}}(M) \to \mathrm{AlmCont}(M)$ and $\mathrm{Cont}(M) \to \mathrm{AlmCont}(M)$ using the h-principle when one fixes a disk. In fact, there is a much more general question here about relative h-principles, motivated by this example.

Question Let Δ be some subset of M and γ be the germ of some holonomic solution on Δ. Let $F(M \text{ rel}(\Delta,\gamma))$ denote the set of all holonomic solutions that have germ γ on Δ, and $F^f(M \text{ rel}(\Delta,\gamma))$

denote the set of formal solutions that have germ γ on Δ. If the map

$$F(M\ \mathrm{rel}(\Delta,\gamma)) \to F^f(M\ \mathrm{rel}(\Delta,\gamma))$$

is a weak equivalence for all pairs $(\Delta,\gamma) \in \mathcal{W}$ for some collection $\mathcal{W}$, what can be said about the map

$$F(M) \to F^f(M)?$$

One of the main results of this paper is a partial answer to this question. Suppose M is a manifold and $A \subset M$ is a (possibly empty) closed subset of M such that $M \setminus A$ is a manifold without boundary. By assumption $F^f(M)$ is the space of sections of some bundle $\pi\colon E \to M$, and furthermore we suppose the fibers of E are path connected. Let $\mathcal{W}$ be a *sufficiently separated collection* for the bundle E relative to A (see Definition 2.2), such that each germ in the collection is holonomic. Finally, let ξ_0 be a holonomic solution near A. Using this, we construct a semisimplicial space $F_\bullet(M\ \mathrm{rel}(A,\xi_0);\mathcal{W})$ with an augmentation to $F(M)$, and prove the following theorem:

Theorem A *If the map*

$$F\big(M\ \mathrm{rel}(\Delta_1,\gamma_1),\dots,(\Delta_k,\gamma_k),(A,\xi_0)\big) \to F^f(M\ \mathrm{rel}(\Delta_1,\gamma_1),\dots,(\Delta_k,\gamma_k),(A,\xi_0))$$

is a weak equivalence for all finite sets of disjoint elements $\{(\Delta_i,\gamma_i)\}_{i=1}^k \subset \mathcal{W}$, $k \geq 1$, *then the following diagram commutes*:

$$\begin{tikzcd} \|F_\bullet(M\ \mathrm{rel}(A,\xi_0);\mathcal{W})\| \arrow[r] \arrow[dr, "\simeq"'] & F(M\ \mathrm{rel}(A,\xi_0)) \arrow[d] \\ & F^f(M\ \mathrm{rel}(A,\xi_0)) \end{tikzcd}$$

and the map $\|F_\bullet(M\ \mathrm{rel}(A,\xi_0);\mathcal{W})\| \to F^f(M\ \mathrm{rel}(A,\xi_0))$ *is a weak equivalence.*

Remark 1.1 As the above theorem is stated, the so-called "nonrelative map" is still relative to the set A. This is because there are two notions of relative h-principle here, one that gives an h-principle relative to fairly arbitrary sets and functions near those sets, and one that gives an h-principle relative to very specific sets and functions near those sets. The former usually includes the empty set and is thus strictly stronger than the nonrelative statement, while the latter should be thought of as asking for the existence of some sort of nice local model. So, the above theorem should be thought of as saying that if one has an h-principle relative to some nice local models, then even when you do not fix any models the map admits a section up to homotopy.

This has some immediate consequences in contact geometry, as the above theorem allows us to find a subgroup of $\pi_k\,\mathrm{Cont}^{\mathrm{OT}}(M)$ isomorphic to $\pi_k\,\mathrm{AlmCont}(M)$, induced by a map of spaces, for all k. This is an improvement on the current tool used to analyze the difference between the homotopy groups of $\mathrm{Cont}^{\mathrm{OT}}(M)$ and $\mathrm{AlmCont}(M)$, the *overtwisted group* (see Casals, del Pino and Presas [4, Proposition 1] or Fernández and Gironella [8, Definition 10]), which only allows one to realize $\pi_k\,\mathrm{AlmCont}(M)$ as a subgroup of $\pi_k\,\mathrm{Cont}^{\mathrm{OT}}(M)$ when $1 \leq k \leq 2n$. Furthermore, we show that these subgroups agree when

$1 \leq k \leq 2n$. Finally, we use these results to help study certain homotopy groups of the contactomorphism group of an overtwisted contact manifold. Let $\mathcal{C}_0(M, \xi_{\mathrm{OT}} \operatorname{rel} \partial M) \subset \mathrm{Diff}_0(M \operatorname{rel} \partial M)$ denote the space of contactomorphisms of the contact manifold (M, ξ_{OT}) which are smoothly isotopic to the identity and agree with the identity near the boundary of M.

Theorem B *If (M, ξ_{OT}) is a compact, coorientable, overtwisted contact manifold of dimension $2n+1$, then $\pi_k \mathcal{C}_0(M, \xi_{\mathrm{OT}} \operatorname{rel} \partial M)$ contains an infinite cyclic subgroup whenever*

$$\pi_k \mathrm{Diff}_0(\mathbb{D}^{2n+1} \operatorname{rel} \partial \mathbb{D}^{2n+1}) \otimes \mathbb{Q} \neq 0, \quad \text{for } k \leq \phi^{\mathbb{Q}}(\mathbb{D}^{2n}) - 1,\ k \neq 0.$$

Here $\phi^{\mathbb{Q}}(\mathbb{D}^{2n})$ is the rational concordance stable range for $\mathbb{D}^{2n}$. The stability theorem of Igusa [13, page 6] implies $\phi^{\mathbb{Q}}(M^d) \geq \min\big(\frac{1}{2}(d-7), \frac{1}{3}(d-4)\big)$ for any compact d-dimensional manifold M. Better lower bounds exist in different cases, for example Corollary C of Goodwillie, Krannich and Kupers [11]. For $d \geq 10$ it is known that $\phi^{\mathbb{Q}}(\mathbb{D}^d) = d-4$ by Corollary B of Kranniich and Randal-Williams [15]. Also, it is known that $\pi_k \mathrm{Diff}_0(\mathbb{D}^{2n+1} \operatorname{rel} \partial \mathbb{D}^{2n+1}) \otimes \mathbb{Q} \neq 0$ for many k in this range, for example Krannich [14, Corollary B] computes $\pi_k \mathrm{Diff}_0(\mathbb{D}^{2n+1} \operatorname{rel} \partial \mathbb{D}^{2n+1}) \otimes \mathbb{Q}$ for $k < 2n-4$ in terms of the algebraic K-theory of the integers.

Overview of the paper

In Section 2, we show that for a fiber bundle $E \to M$ with space of continuous sections $\Gamma(E)$, if $\mathcal{W}$ is a *sufficiently separated collection* relative to A of pairs (Δ, γ) for E (see Definition 2.2) then the space $\Gamma(E \operatorname{rel}(A, \xi_0))$ is weakly equivalent to the geometric realization of a certain semisimplicial space built from the spaces of relative sections. Then in Section 3, we use this result to prove Theorem A. In Section 4, we show that the collection of all overtwisted disks on a manifold is sufficiently separated, and so the map $\mathrm{Cont}^{\mathrm{OT}}(M \operatorname{rel} A) \to \mathrm{AlmCont}(M \operatorname{rel} A)$ admits a section up to homotopy. Then we use this to show that in the range it is defined, the usual overtwisted group agrees with the image of the map induced by this section. In Section 5 we use these results to prove Theorem B, which finds infinite cyclic subgroups in the homotopy groups of the contactomorphism groups of compact overtwisted contact manifolds, in degrees different than those found in [8]. Finally, in Section 6 we note other applications of Theorem A, specifically coming from Engel geometry.

Acknowledgements

I would like to thank my supervisor Alexander Kupers for suggesting this topic and for his patience, support and guidance throughout this project. I would also like to thank the referee for their comments, which have improved both the clarity and the scope of the paper. This work is partly supported by an Ontario Graduate Scholarship.

2 Semisimplicial resolutions of section spaces

For technical reasons, suppose we are in the category of compactly generated spaces. Let M be a d-dimensional manifold, possibly with boundary. Let X be a path connected space and $\pi\colon E \to M$ be a fiber bundle with fiber X. Let $\Gamma(E)$ denote the space of sections of E.

Definition 2.1 Let $A \subset M$ be some subset of M. A *germ of a section on* A is a pair (γ, U), where U is an open neighborhood of A and γ is a section on U, with the equivalence relation that two germs are the same if they agree on some neighborhood of A.

For convenience we usually omit the neighborhood U and just let γ denote the germ, with the understanding that γ is defined on some arbitrarily small neighborhood of A. If $A_1, \dots, A_k \subset M$ and γ_i is a germ of a section on A_i for $1 \le i \le k$, we will let $\Gamma(E \operatorname{rel}(A_1, \gamma_1), \dots, (A_k, \gamma_k))$ denote the space of sections of E which agree with γ_i near A_i for all $1 \le i \le k$. By abuse of notation, we will let $\Gamma(E \operatorname{rel} A_1, \dots, A_k)$ denote the same space, and more generally we will use this convention for many function spaces appearing in this paper.

Definition 2.2 A *sufficiently separated collection relative to* A for the bundle $\pi\colon E \to M$ is a collection $\mathcal{W}$ of pairs (Δ, γ) of (a) a contractible compact subset $\Delta \subset M \setminus A$ and (b) a germ of a section near Δ. These are required to satisfy:

(1) For each $(\Delta, \gamma) \in \mathcal{W}$, there exists some neighborhood $D \cong \mathbb{D}^d$ such that Δ is contained in the interior of D, $D \subset M \setminus A$, and the inclusion map $\iota\colon \Delta \to D$ is a closed cofibration.

(2) Given any finite collection $(\Delta_i, \gamma_i)_{i=1}^k$ in $\mathcal{W}$, there exists some $(\Delta', \gamma') \in \mathcal{W}$ so that Δ' is contained in the interior of a closed ball $D \cong \mathbb{D}^d$ that is disjoint from each Δ_i, $D \subset M \setminus A$, and the inclusion $\iota\colon \Delta' \to D$ is a closed cofibration.

We say that two elements (Δ_1, γ_1), (Δ_2, γ_2) of $\mathcal{W}$ are *disjoint* if $\Delta_1 \cap \Delta_2 = \varnothing$, and a collection of elements is *disjoint* if each pair of elements in the collection is disjoint. Next, let $A \subset M$ and ξ_0 be some germ of a section on A. Suppose $\mathcal{W}$ is a sufficiently separated collection relative to A for the bundle E. Then let $\Gamma_\bullet(E \operatorname{rel} A; \mathcal{W})$ be the semisimplicial space defined by

$$\Gamma_p(E \operatorname{rel} A; \mathcal{W}) := \coprod_{\substack{\text{tuples } ((\Delta_0,\gamma_0),\dots,(\Delta_p,\gamma_p)) \\ \text{of } p+1 \text{ disjoint elements of } \mathcal{W}}} \Gamma(E \operatorname{rel} \Delta_0, \dots, \Delta_p, A).$$

The i^{th} face map is given by forgetting that the sections agree with γ_i near Δ_i. Our goal is to compare the space $\|\Gamma_\bullet(E \operatorname{rel} A; \mathcal{W})\|$ to $\Gamma(E \operatorname{rel} A)$. In order to do this, we need a few lemmas. First, let $(\Delta, \gamma) \in \mathcal{W}$ and $\Gamma(E, \Delta \operatorname{rel} A)$ denote the space of sections s of E which agree with ξ_0 near A and satisfy $s|_\Delta = \gamma|_\Delta$. Then we have the following lemma, which will allow us to forget about germs and just work with spaces of sections that have fixed values on subsets.

Lemma 2.3 *The map $\Gamma(E \operatorname{rel} \Delta, A) \to \Gamma(E, \Delta \operatorname{rel} A)$ given by inclusion is a weak equivalence.*

Proof By definition, $\Gamma(E \operatorname{rel} \Delta, A)$ is topologized as the colimit

$$\Gamma(E \operatorname{rel} \Delta, A) := \varinjlim \Gamma(E, B_i \operatorname{rel} A),$$

where the B_i are some neighborhood basis of Δ. Let $\alpha\colon \mathbb{D}^n \to \Gamma(E, \Delta \operatorname{rel} A)$ be a continuous map such that we have the following diagram:

$$\begin{tikzcd} S^{n-1} \arrow[r, "\partial\alpha"] \arrow[d] & \Gamma(E \operatorname{rel} \Delta, A) \arrow[d] \\ \mathbb{D}^n \arrow[r, "\alpha"] & \Gamma(E, \Delta \operatorname{rel} A) \end{tikzcd}$$

We will show there exists a lift $\beta\colon \mathbb{D}^n \to \Gamma(E \operatorname{rel} \Delta, A)$ of α up to homotopy relative to the boundary. By Lemma 3.6 of [17], any map from a compact space to a colimit of closed inclusions factors over one of the inclusions, so $S^{n-1} \to \Gamma(E \operatorname{rel} \Delta, A)$ factors as

$$S^{n-1} \to \Gamma(E, B_i \operatorname{rel} A) \to \Gamma(E \operatorname{rel} \Delta, A)$$

for some neighborhood B_i. We can choose B_i to be contractible and such that the inclusion $B_i \to M$ is a closed cofibration. We can also ensure B_i is disjoint from A. Then it suffices to show that $\Gamma(E, B_i \operatorname{rel} A) \to \Gamma(E, \Delta \operatorname{rel} A)$ is a weak equivalence, which holds due to the following commutative diagram; the rows are fiber sequences and $\Gamma(E|_{B_i}) \to \Gamma(E|_\Delta)$ is a weak equivalence since both Δ and B_i are contractible:

$$\begin{tikzcd} \Gamma(E, B_i \operatorname{rel} A) \arrow[r] \arrow[d, "\iota"'] & \Gamma(E \operatorname{rel} A) \arrow[r] \arrow[d, "\mathrm{id}"'] & \Gamma(E|_{B_i}) \arrow[d, "\mathrm{res}"'] \\ \Gamma(E, \Delta \operatorname{rel} A) \arrow[r] & \Gamma(E \operatorname{rel} A) \arrow[r] & \Gamma(E|_\Delta) \end{tikzcd}$$

□

Remark 2.4 A similar argument can be used if one replaces (Δ, γ) with finitely many disjoint elements of $\mathcal{W}$.

This allows us to use the following abuse of notation. Let $\Gamma_\bullet(E \operatorname{rel} A; \mathcal{W})$ denote the semisimplicial space given by

$$\Gamma_p(E \operatorname{rel} A; \mathcal{W}) := \coprod_{\substack{\text{tuples } ((\Delta_0,\gamma_0),\dots,(\Delta_p,\gamma_p)) \\ \text{of } p+1 \text{ disjoint elements of } \mathcal{W}}} \Gamma(E, \Delta_0, \dots, \Delta_p \operatorname{rel} A),$$

where again the face maps just forget the fixed sets. We can do this because Lemma 2.3 implies the resulting semisimplicial space is levelwise weakly equivalent to the one constructed before, and so the geometric realizations of both are weakly equivalent. We can also consider the semisimplicial space $W_\bullet$ given by

$$W_p := \coprod_{\substack{\text{tuples } ((\Delta_0,\gamma_0),\dots,(\Delta_p,\gamma_p)) \\ \text{of } p+1 \text{ disjoint elements of } \mathcal{W}}} \{\star\},$$

where the face maps are given by forgetting elements of $\mathcal{W}$. Here $\{\star\}$ is just a singleton set, so W_p contains a point for each tuple of $p+1$ disjoint elements of $\mathcal{W}$, with the discrete topology. These semisimplicial spaces are related via the following lemma.

Lemma 2.5 *The space $\|\Gamma_\bullet(E \operatorname{rel} A; \mathcal{W})\|$ is homeomorphic to the subspace of $\Gamma(E \operatorname{rel} A) \times \|W_\bullet\|$ given by*

$$\{(f, \vec{w}, \vec{t}) \mid f(m) = \gamma_i(m) \text{ whenever } m \in \Delta_i \text{ and } t_i \neq 0\}/\sim,$$

where (Δ_i, γ_i) is the i^{th} component of $\vec{w}$, and $\sim$ is just the usual geometric realization equivalence on the second factor.

Proof It is clear these are the same as sets, so we just need to show that this map is a homeomorphism onto some subspace. First, since the quotient of a subspace is naturally a subspace of the quotient in compactly generated spaces, we have

$$\|\Gamma_\bullet(E \operatorname{rel} A; \mathcal{W})\| \subset \|\Gamma(E \operatorname{rel} A) \times W_\bullet\|.$$

On the other hand, by [7, page 2106] we have that

$$\|\Gamma(E \operatorname{rel} A) \times W_\bullet\| = \|\Gamma(E \operatorname{rel} A) \otimes W_\bullet\| \cong \|\Gamma(E \operatorname{rel} A)\| \times \|W_\bullet\| \cong \Gamma(E \operatorname{rel} A) \times \|W_\bullet\|,$$

where we are treating $\Gamma(E \operatorname{rel} A)$ as a semisimplicial space with only 0-simplices in order to use the exterior product defined in [7, page 2103]. So, $\|\Gamma_\bullet(E \operatorname{rel} A; \mathcal{W})\|$ is homeomorphic to the subspace of $\Gamma(E \operatorname{rel} A) \times \|W_\bullet\|$ described above. $\square$

We can now introduce the main result of this section:

Theorem 2.6 *Let $A \subset M$ be a closed subset of a manifold M and ξ_0 be a germ of a section on A, such that $M \setminus A$ is a manifold without boundary. Furthermore, let $\pi : E \to M$ be a fiber bundle with connected fiber X, and $\mathcal{W}$ be a sufficiently separated collection relative to A for the bundle E. Let $\|\Gamma_\bullet(E \operatorname{rel} A; \mathcal{W})\|$ be as before. Then the map $\|\Gamma_\bullet(E \operatorname{rel} A; \mathcal{W})\| \to \Gamma(E \operatorname{rel} A)$ given by forgetting the fixed subsets is a weak equivalence.*

Proof We will prove this by showing that the relative homotopy groups of the map are zero. So, let $\alpha : \mathbb{D}^n \to \Gamma(E \operatorname{rel} A)$ be a continuous map such that we have the following diagram:

$$\begin{array}{ccc} S^{n-1} & \xrightarrow{\partial\alpha} & \|\Gamma_\bullet(E \operatorname{rel} A; \mathcal{W})\| \\ \downarrow & & \downarrow \\ \mathbb{D}^n & \xrightarrow{\alpha} & \Gamma(E \operatorname{rel} A) \end{array}$$

where by abuse of notation $\partial\alpha$ is the map $S^{n-1} \to \|\Gamma_\bullet(E \operatorname{rel} A; \mathcal{W})\|$, $p \mapsto (\alpha(p), \vec{w}_p, \vec{t}_p)$ for some finite ordered set of elements $\vec{w}_p$ of $\mathcal{W}$ and weights $\vec{t}_p$ (where of course we just exclude any elements when their weight goes to zero). Then we need to show that there exists a continuous map $\alpha' \simeq \alpha$ relative to the boundary, and a lift $\beta : \mathbb{D}^n \to \|\Gamma_\bullet(E \operatorname{rel} A; \mathcal{W})\|$ of α' such that the resulting diagram still commutes.

First, consider the section

$$\alpha_1(p)(m) := \begin{cases} \alpha(2p)(m), & 0 \le |p| \le \frac{1}{2}, \\ \alpha(p/|p|), & \frac{1}{2} \le |p| \le 1. \end{cases}$$

So α_1 is just α compressed to a smaller ball, with the boundary extended to an annulus. It is clear that $\alpha_1 \simeq \alpha$ relative to the boundary, so we can work with α_1, which gives us a buffer away from the boundary. Next, consider the set

$$\widetilde{W} := \{m \in M \mid \text{for some } p \in S^{n-1} \text{ and some positive integer } i, m \in \Delta_{p,i}, \\ \text{where } (\Delta_{p,i}, \gamma_{p,i}) \text{ is the } i^{\text{th}} \text{ component of } \vec{w}_p\}.$$

Note that α_1 lifts to $\|\Gamma_\bullet(E \text{ rel } A; \mathcal{W})\|$ on the annulus, and furthermore if we only change α_1 away from $\widetilde{W}$ then it will still lift in the annulus. Now, using the natural projection map from $\|\Gamma_\bullet(E \text{ rel } A; \mathcal{W})\| \to \|W_\bullet\|$ given in Lemma 2.5, we can consider the map

$$S^{n-1} \to \|\Gamma_\bullet(E \text{ rel } A; \mathcal{W})\| \to \|W_\bullet\|,$$

where $p \mapsto (\vec{w}_p, \vec{t}_p)$. Since S^{n-1} is compact we know that the image of this map is compact, and so it hits finitely many cells of $\|W_\bullet\|$. Also, each cell consists of finitely many elements of $\mathcal{W}$, so the set of all elements of $\mathcal{W}$ that are a component of $\vec{w}_p$ for some p is finite. But, if $\{(\Delta_1, \gamma_1), \dots, (\Delta_k, \gamma_k)\}$ is that set, then clearly $\widetilde{W} = \Delta_1 \cup \cdots \cup \Delta_k$. Let $(\Delta, \gamma) \in \mathcal{W}$ be such that there is some neighborhood $D \cong \mathbb{D}^d$ of Δ contained in $M \setminus A$ that is disjoint from $\widetilde{W}$. Since a map $\mathbb{D}^n \to \Gamma(E \text{ rel } A)$ is the same data as a section of the bundle $\mathbb{D}^n \times E \to \mathbb{D}^n \times M$, $(p, e) \mapsto (p, \pi(e))$ we will from now on consider α, α_1 as maps from $\mathbb{D}^n \times M \to \mathbb{D}^n \times E$ so that composition with the projection map is the identity. Now, let $g_0 = \alpha_1|_{\mathbb{D}^n \times D}$, $\mathbb{D}^n_{3/4} := \{p \in \mathbb{D}^n \mid |p| \le \frac{3}{4}\}$. Since $\mathbb{D}^n \times D$ is a contractible submanifold of $\mathbb{D}^n \times M$, we know that the restriction of the bundle to $\mathbb{D}^n \times D$ is trivial. So, g_0 and any other sections on (a subset of) $\mathbb{D}^n \times D$ just become maps from (a subset of) $\mathbb{D}^n \times D$ to X. Consider the homotopy

$$(\mathbb{D}^n_{3/4} \times \Delta) \cup \partial(\mathbb{D}^n \times D) \times [0,1] \to X$$

given by

$$H(p, m, s) := \begin{cases} \alpha_1(p, m), & p \in \partial\mathbb{D}^n, \\ \alpha_1(p, m), & m \in \partial D, \\ h(p, m, s), & (p, m) \in \mathbb{D}^n_{3/4} \times \Delta, \end{cases}$$

where $h(p, m, s)$ is a homotopy between $\alpha_1|_{\mathbb{D}^n_{3/4} \times \Delta}$ and the map $(p, m) \mapsto \gamma(m)$. Such an h exists since $\mathbb{D}^n_{3/4} \times \Delta$ is contractible and X is path connected, which implies any two maps are homotopic. Since the inclusion of Δ into D is a cofibration, and the inclusion of $\mathbb{D}^n_{3/4}$ into $\mathbb{D}^n$ is a cofibration, the inclusion of $\mathbb{D}^n_{3/4} \times \Delta$ into $\mathbb{D}^n \times D$ also is. This implies $(\mathbb{D}^n \times D, (\mathbb{D}^n_{3/4} \times \Delta) \cup \partial(\mathbb{D}^n \times D))$ satisfies the homotopy extension property, so there exists a homotopy

$$\mathbb{D}^n \times D \times [0,1] \to X$$

between g_0 and a function $g_1 : \mathbb{D}^n \times D \to X$, relative to the boundary, so that $g_1 = g_0$ on $\partial\mathbb{D}^n$, ∂D, and $(p,m) \mapsto \gamma(m)$ on $\mathbb{D}^n_{3/4} \times \Delta$. Now, we can extend this to a section on all of M by letting

$$\alpha_2(p,m) := \begin{cases} \alpha_1(p,m), & m \notin D, \\ g_1(p,m), & m \in D. \end{cases}$$

Clearly $\alpha_2 \simeq \alpha_1$ relative to the boundary, $\alpha_2(p,m) = \gamma(m)$ when $|p| \le \frac{3}{4}$, $m \in \Delta$, and for $m \in \widetilde{W}$ we have $\alpha_2(p,m) = \alpha_1(p,m) = \gamma_{p/|p|,i}(m)$ when $|p| \ge \frac{1}{2}$. So, all we need to do now is show that α_2 lifts. Let us again view these as maps from $\mathbb{D}^n \to \Gamma(E \text{ rel } A)$, and consider the map $\beta : \mathbb{D}^n \to \|\Gamma_\bullet(E \text{ rel } A; \mathcal{W})\|$ given by

$$\beta(p) := \begin{cases} (\alpha_2(p), (\Delta,\gamma), 1), & 0 \le |p| \le \frac{1}{2}, \\ \big(\alpha_2(p), (\vec{w}_{p/|p|}, (\Delta,\gamma)), ((4|p|-2)\vec{t}_{p/|p|}, 3-4|p|)\big), & \frac{1}{2} < |p| \le \frac{3}{4}, \\ (\alpha_2(p), \vec{w}_{p/|p|}, \vec{t}_{p/|p|}), & \frac{3}{4} < |p| \le 1. \end{cases}$$

Clearly β is a lift of α_2, as required. □

Remark 2.7 In particular, the above theorem applies when M has no boundary and $A = \varnothing$, and when M has boundary and $A = \partial M$.

3 *h*-principle

Let M be a d-dimensional manifold and A be a (possibly empty) closed subset of M such that $M \setminus A$ is a manifold without boundary. Suppose $F(M)$ is a space of holonomic solutions with space of formal solutions $F^f(M)$, and ξ_0 is a germ of a holonomic solution on A. We would like to use Theorem 2.6 to show that the map $F(M \text{ rel } A) \to F^f(M \text{ rel } A)$ from the space of holonomic solutions relative to A to the space of formal solutions relative to A admits a section up to homotopy under the right conditions. We assume there exists some bundle $E \to M$ such that $F^f(M \text{ rel } A)$ is the space $\Gamma(E \text{ rel } A)$ of sections of E relative to A. Suppose furthermore that the fibers of E are path connected, and let $\mathcal{W}$ be a sufficiently separated collection for E relative to A. In addition, for each $(\Delta,\gamma) \in \mathcal{W}$, suppose γ is the germ of a holonomic solution. Then we can define semisimplicial spaces $F_\bullet(M \text{ rel } A; \mathcal{W})$ and $F^f_\bullet(M \text{ rel } A; \mathcal{W})$ by letting

$$F_p(M \text{ rel } A; \mathcal{W}) := \coprod_{\substack{\text{tuples } ((\Delta_0,\gamma_0),\dots,(\Delta_p,\gamma_p)) \\ \text{of } p+1 \text{ disjoint elements of } \mathcal{W}}} F(M \text{ rel } A, \Delta_0, \dots, \Delta_p),$$

$$F^f_p(M \text{ rel } A; \mathcal{W}) := \coprod_{\substack{\text{tuples } ((\Delta_0,\gamma_0),\dots,(\Delta_p,\gamma_p)) \\ \text{of } p+1 \text{ disjoint elements of } \mathcal{W}}} F^f(M \text{ rel } A, \Delta_0, \dots, \Delta_p),$$

where as before the face maps come from forgetting that the functions agree with the given germs near the given sets. There is a map of semisimplicial spaces

$$F_\bullet(M \text{ rel } A; \mathcal{W}) \to F^f_\bullet(M \text{ rel } A; \mathcal{W})$$

induced by the maps

$$F(M \text{ rel } A, \Delta_0, \dots, \Delta_p) \to F^f(M \text{ rel } A, \Delta_0, \dots, \Delta_p).$$

Also, there are natural maps $\| F_\bullet(M \text{ rel } A; \mathcal{W}) \| \to F(M \text{ rel } A)$ and $\| F^f_\bullet(M \text{ rel } A; \mathcal{W}) \| \to F^f(M \text{ rel } A)$ which forget the elements of $\mathcal{W}$.

Theorem 3.1 *If the natural inclusion map*

$$F(M \text{ rel } A, \Delta_0, \dots, \Delta_p) \to F^f(M \text{ rel } A, \Delta_0, \dots, \Delta_p)$$

is a weak equivalence for all finite sets of disjoint elements $\{(\Delta_i, \gamma_i)\}_{i=1}^k \subset \mathcal{W}$, $k \geq 1$, *then the composition map* $\| F_\bullet(M \text{ rel } A; \mathcal{W}) \| \to F(M \text{ rel } A) \to F^f(M \text{ rel } A)$ *is a weak homotopy equivalence.*

Proof First, we are given that

$$F(M \text{ rel } A, \Delta_0, \dots, \Delta_p) \to F^f(M \text{ rel } A, \Delta_0, \dots, \Delta_p)$$

is a weak equivalence for all tuples of elements of $\mathcal{W}$, so the map

$$\| F_\bullet(M \text{ rel } A; \mathcal{W}) \| \to \| F^f_\bullet(M \text{ rel } A; \mathcal{W}) \|$$

is a levelwise weak equivalence and hence a weak equivalence. Also, by Theorem 2.6,

$$\| F^f_\bullet(M \text{ rel } A; \mathcal{W}) \| \to F^f(M \text{ rel } A)$$

is a weak equivalence, so we get the following commutative diagram:

$$\begin{tikzcd}
\| F_\bullet(M \text{ rel } A; \mathcal{W}) \| \arrow[r] \arrow[d, "\simeq"] & F(M \text{ rel } A) \arrow[d] \\
\| F^f_\bullet(M \text{ rel } A; \mathcal{W}) \| \arrow[r, "\simeq"] & F^f(M \text{ rel } A)
\end{tikzcd}$$

which gives the required weak homotopy equivalence. □

Remark 3.2 A more common formulation of such a relative h-principle result is that there is an h-principle relative to any fixed closed set A and one fixed subset Δ, where Δ is from some special collection. Such a result will usually imply the result for many fixed subsets $\Delta_1, \dots, \Delta_k$ in such a collection, since we can take A to be the union of the first $k-1$ such sets, and use Δ_k as our fixed subset Δ.

4 Improved *h*-principle for contact geometry

Let us briefly recall some basic definitions from contact geometry. A cooriented contact structure on a connected, orientable, $(2n+1)$-dimensional manifold M is a "maximally nonintegrable" hyperplane distribution $\xi = \ker(\alpha)$ for a 1-form α. Maximally nonintegrable means $\alpha \wedge d\alpha^n \neq 0$. Unless otherwise stated we assume all contact structures are cooriented. This naturally induces a reduction of the structure group of M to $U(n) \times 1$, and so an *almost contact structure* is just a reduction of the structure group of

M to $U(n)\times 1$. Equivalently, an almost contact structure is a triple (ξ, J, R), where ξ is a hyperplane distribution, J is a complex structure on ξ, and R is a trivial sub-line-bundle of TM such that $\xi \oplus R = TM$. We let $\mathrm{Cont}(M)$ denote the space of contact structures on M and $\mathrm{AlmCont}(M)$ denote the space of almost contact structures on M.

Next, we recall the notion of an overtwisted contact structure. An *overtwisted disk* in a manifold M is a pair (Δ,γ), where $\Delta \subset M$ is an embedded $2n$-dimensional disk and γ is a certain model germ of a contact structure on Δ. Then a contact manifold (M,ξ) is said to be overtwisted if there exists an embedding of an overtwisted disk (Δ,γ) such that the contact germ γ agrees with ξ on some neighborhood of Δ (ie the embedding is a contact embedding). In this case we say that Δ is overtwisted for ξ. The details of this definition in any dimension can be found in [2, Definition 3.6]. Let $\mathrm{Cont}^{\mathrm{OT}}(M)$ denote the space of overtwisted contact structures on M. It has been shown by [2, Theorem 1.2] that if M is a closed $2n+1$-dimensional manifold, A is a closed subset of M such that $M \setminus A$ is connected, (Δ,γ) is an overtwisted disk in $M\setminus A$, and ξ_0 is an almost contact structure on M that is a genuine contact structure on a neighborhood of A, then the map

$$\mathrm{Cont}^{\mathrm{OT}}(M \text{ rel } A, \Delta) \to \mathrm{AlmCont}(M \text{ rel } A, \Delta)$$

is a weak equivalence. Here $\mathrm{Cont}^{\mathrm{OT}}(M \text{ rel } A, \Delta)$ is the space of contact structures on M that agree with ξ_0 in a neighborhood of A and are overtwisted with disk Δ, and $\mathrm{AlmCont}(M \text{ rel } A, \Delta)$ is the corresponding space of almost contact structures. Next, let $\mathcal{W}$ be the collection of all overtwisted disks in $M\setminus A$, and let $\mathrm{Cont}^{\mathrm{OT}}_\bullet(M \text{ rel } A; \mathcal{W})$ and $\mathrm{AlmCont}_\bullet(M \text{ rel } A; \mathcal{W})$ be the semisimplicial spaces defined by

$$\mathrm{Cont}^{\mathrm{OT}}_p(M \text{ rel } A; \mathcal{W}) := \coprod_{\substack{\text{tuples } ((\Delta_0,\gamma_0),\dots,(\Delta_p,\gamma_p)) \\ \text{of } p+1 \text{ disjoint elements of } \mathcal{W}}} \mathrm{Cont}^{\mathrm{OT}}(M \text{ rel } A, \Delta_0, \dots, \Delta_p),$$

$$\mathrm{AlmCont}_p(M \text{ rel } A; \mathcal{W}) := \coprod_{\substack{\text{tuples } ((\Delta_0,\gamma_0),\dots,(\Delta_p,\gamma_p)) \\ \text{of } p+1 \text{ disjoint elements of } \mathcal{W}}} \mathrm{AlmCont}(M \text{ rel } A, \Delta_0, \dots, \Delta_p),$$

where as usual the face maps just forget the overtwisted disks. From this, we would like to prove the following theorem:

Theorem 4.1 *Let M be a $(2n+1)$-dimensional manifold and $A \subset M$ be a closed subset of M such that $M\setminus A$ is connected and without boundary. Then the composition*

$$\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M \text{ rel } A; \mathcal{W})\| \to \mathrm{Cont}^{\mathrm{OT}}(M \text{ rel } A) \to \mathrm{AlmCont}(M \text{ rel } A)$$

is a weak homotopy equivalence.

Proof It is known (see [8, page 191]) that $\mathrm{AlmCont}(M)$ is naturally the section space of a certain fiber bundle $\pi\colon E \to M$ with connected fiber $\mathrm{SO}(2n+1)/U(n)$, which comes from viewing $\mathrm{AlmCont}(M)$ as the space of reductions of the structure group of M to $U(n)\times 1$, so $\mathrm{AlmCont}(M \text{ rel } A)$ is naturally $\Gamma(E \text{ rel } A)$. Also, if we have some collection of disjoint overtwisted disks $\Delta_1,\dots,\Delta_k$, we can let $A' = A \cup \Delta_1 \cup \dots \cup \Delta_{k-1}$ and Δ_k be the overtwisted disk, so the h-principle given in [2] implies an

h-principle of the form required in Theorem 3.1. So, we just need to show that the collection of all overtwisted disks on $M \setminus A$ is sufficiently separated relative to A. Clearly such an embedded disk is closed, contractible and compact. Also, by finding a sufficiently small regular neighborhood, it is clear that any embedded $2n$-dimensional disk is a neighborhood deformation retract of a $(2n+1)$-dimensional ball in $M \setminus A$ containing it, so the inclusion is a cofibration and condition (1) of Definition 2.2 is satisfied. Finally, it is clear that a finite collection of embedded overtwisted disks in $M \setminus A$ can't cover $M \setminus A$, so given some finite set of overtwisted disks in $M \setminus A$ there will always be some point $m \in M \setminus A$ that is not in any of them, and since an overtwisted disk is just some embedded disk with a local germ, we can always introduce a new overtwisted disk at the point m that doesn't intersect the rest of the overtwisted disks. Again, we can choose a sufficiently small regular neighborhood of the disk that doesn't intersect the given overtwisted disks or the set A, so that the inclusion is a cofibration and condition (2) of Definition 2.2 is satisfied. □

Remark 4.2 This argument also shows that the map $\mathrm{Cont}(M \text{ rel } A) \to \mathrm{AlmCont}(M \text{ rel } A)$ admits a section up to homotopy, but this section factors through the overtwisted contact structures.

Remark 4.3 In particular, this holds if M is closed and connected and A is empty.

From Theorem 4.1 we get as an immediate consequence that $\pi_k \,\mathrm{AlmCont}(M \text{ rel } A)$ is isomorphic to a subgroup of $\pi_k \,\mathrm{Cont}^{\mathrm{OT}}(M \text{ rel } A)$ for all k. For the remainder of this section, suppose M is a closed manifold and $A = \varnothing$. Then our result is an improvement on the current overtwisted group $\mathrm{OT}_k(M)$ (see [4, Proposition A.2]), which gives an isomorphism between $\pi_k \,\mathrm{AlmCont}(M)$ and a subgroup of $\pi_k \,\mathrm{Cont}^{\mathrm{OT}}(M)$ when $1 \le k \le 2n$. In fact, when $1 \le k \le 2n$ the image of $\pi_k \|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M; \mathcal{W})\|$ in $\pi_k \,\mathrm{Cont}^{\mathrm{OT}}(M)$ is $\mathrm{OT}_k(M)$:

Theorem 4.4 *The overtwisted group* $\mathrm{OT}_k(M)$ *is the image of* $\pi_k \|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M; \mathcal{W})\|$ *induced by the natural forgetful map* $\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M; \mathcal{W})\| \to \mathrm{Cont}^{\mathrm{OT}}(M)$ *when* $1 \le k \le 2n$.

Before we can prove this, we need to define some intermediary spaces that will help us understand the relationship between $\pi_k \|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M; \mathcal{W})\|$ and $\mathrm{OT}_k(M)$. Recall that an element of $\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M; \mathcal{W})\|$ is a contact structure, along with a list of disks and weights, such that each disk is overtwisted for the contact structure as long as its weight is nonzero. Since we defined this for an arbitrary section space and an arbitrary sufficiently separated collection, we did not make use of any topology on the space of disks. So, in $\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M; \mathcal{W})\|$ the fixed disks are not allowed to move through M, and the only way to change the disks is to introduce a new overtwisted disk somewhere, or delete a disk by letting its weight go to zero. However, the space of overtwisted disks does have a topology coming from the space of embeddings of $\mathbb{D}^{2n}$ into M. We can use this to define new semisimplicial spaces that are more clearly related to $\mathrm{OT}_k(M)$.

Remark 4.5 Strictly speaking, overtwisted disks are only piecewise smooth, so instead of embeddings of the standard disk into M we want to take a specific piecewise structure coming from the model overtwisted disk (see [2, Definition 3.6]) and consider the space of embeddings of this into M that preserve the piecewise smooth structure. However, none of our arguments depend on this distinction, so by abuse of notation we just denote this space as $\mathrm{Emb}(\mathbb{D}^{2n}, M)$.

Definition 4.6 Let $\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})^C$ be the semisimplicial space defined by

$$\mathrm{Cont}^{\mathrm{OT}}_p(M;\mathcal{W})^C \subset \mathrm{Cont}^{\mathrm{OT}}(M) \times \mathrm{Emb}(\mathbb{D}^{2n}, M)^{p+1},$$

the set of all $(\xi, \Delta_0, \dots, \Delta_p)$ such that each Δ_i is an overtwisted disk for ξ, and Δ_i, Δ_j are disjoint when $i \neq j$. The face maps d_i are given by forgetting the i^{th} disk.

Similarly, let $\mathrm{AlmCont}_\bullet(M;\mathcal{W})^C$ be defined in the same way except with $\mathrm{AlmCont}(M)$ instead of $\mathrm{Cont}^{\mathrm{OT}}(M)$. The geometric realizations of these differ from $\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})\|$ and $\|\mathrm{AlmCont}_\bullet(M;\mathcal{W})\|$ since continuous maps into $\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})^C\|$ and $\|\mathrm{AlmCont}_\bullet(M;\mathcal{W})^C\|$ can also deform disks along families inside of M. For example, if $\alpha\colon S^k \to \mathrm{Cont}^{\mathrm{OT}}(M)$ is a family of contact structures, and $\Delta\colon S^k \to \mathrm{Emb}(\mathbb{D}^{2n}, M)$ is a certificate of overtwistedness for α, then $(\alpha, \Delta, 1)$ is naturally a continuous map $S^k \to \|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})^C\|$. Furthermore, since constant embeddings are still continuous embeddings, there are natural maps

$$\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})\| \to \|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})^C\| \quad \text{and} \quad \|\mathrm{AlmCont}_\bullet(M;\mathcal{W})\| \to \|\mathrm{AlmCont}_\bullet(M;\mathcal{W})^C\|$$

given by viewing the fixed disks as constant embeddings. Then we have the following commutative diagram:

$$\begin{tikzcd}
\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})\| \arrow[r] \arrow[d] & \|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})^C\| \arrow[r] \arrow[d] & \mathrm{Cont}^{\mathrm{OT}}(M) \arrow[d] \\
\|\mathrm{AlmCont}_\bullet(M;\mathcal{W})\| \arrow[r] & \|\mathrm{AlmCont}_\bullet(M;\mathcal{W})^C\| \arrow[r] & \mathrm{AlmCont}(M)
\end{tikzcd}$$

We already know the map $\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})\| \to \|\mathrm{AlmCont}_\bullet(M;\mathcal{W})\|$ is a weak equivalence by the h-principle given in [2]. Also, we have the following lemma, which is a direct consequence of [2].

Lemma 4.7 *The map*

$$\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})^C\| \to \|\mathrm{AlmCont}_\bullet(M;\mathcal{W})^C\|$$

induced by $\mathrm{Cont}^{\mathrm{OT}}(M) \to \mathrm{AlmCont}(M)$ *is a weak equivalence.*

Proof First, $\mathrm{Cont}^{\mathrm{OT}}_p(M;\mathcal{W})^C \to \mathrm{AlmCont}_p(M;\mathcal{W})^C$ is a weak equivalence as a direct consequence of Theorem 1.6 of [2]. Indeed, suppose we have the following diagram:

$$\begin{tikzcd}
S^{k-1} \arrow[r, "\partial\alpha"] \arrow[d] & \mathrm{Cont}^{\mathrm{OT}}_p(M;\mathcal{W})^C \arrow[d] \\
\mathbb{D}^k \arrow[r, "\alpha"] & \mathrm{AlmCont}_p(M;\mathcal{W})^C
\end{tikzcd}$$

where $\alpha(t) = (\xi(t), \Delta_0(t), \dots, \Delta_p(t))$. First, we can homotope α relative to the boundary by extending the boundary to an annulus, so that α is genuine on a neighborhood of the boundary. Then we can consider $V = M \times \mathbb{D}^k$, so that ξ can be viewed as a leafwise almost contact structure on V. If we let $A = S^{k-1} \times M \subset V$, $\xi_0 = \xi|_A$, and $h_i = \Delta_i$ for $0 \le i \le p$, we have that

$$\xi \in \mathrm{AlmCont}(V; A, \xi_0, h_0, \dots, h_p),$$

where $\mathrm{AlmCont}(V; A, \xi_0, h_0, \dots, h_p)$ is the space of leafwise almost contact structures that agree with ξ_0 near A, with overtwisted basis $\{h_i\}_{i=0}^p$ (see [2, Theorem 1.6 and definitions immediately preceding]). Then ξ is a representative of an element in $\pi_0 \mathrm{AlmCont}(V; A, \xi_0, h_0, \dots, h_p)$. But, if $\mathrm{Cont}(V; A, \xi_0, h_0, \dots, h_p)$ is the space of leafwise contact structures that agree with ξ_0 near A, with overtwisted basis $\{h_i\}_{i=0}^p$, then by [2, Theorem 1.6] we have that

$$\pi_0 \mathrm{Cont}(V; A, \xi_0, h_0, \dots, h_p) \to \pi_0 \mathrm{AlmCont}(V; A, \xi_0, h_0, \dots, h_p)$$

is an isomorphism, so there is a path from ξ to $\tilde{\xi}$ in $\mathrm{AlmCont}(V; A, \xi_0, h_0, \dots, h_p)$ for some $\tilde{\xi} \in \mathrm{Cont}(V; A, \xi_0, h_0, \dots, h_p)$. But, a path in this space is a homotopy of ξ relative to the boundary and relative to the families of overtwisted disks $\Delta_0, \dots, \Delta_p$. Clearly such a homotopy is a homotopy of $\alpha\colon \mathbb{D}^k \to \mathrm{AlmCont}_p(M; \mathcal{W})^C$ relative to the boundary to a map $\beta = (\tilde{\xi}, \Delta_0, \dots, \Delta_p)$, which has image in $\mathrm{Cont}_p^{\mathrm{OT}}(M; \mathcal{W})^C$. So indeed $\mathrm{Cont}_p^{\mathrm{OT}}(M; \mathcal{W})^C \to \mathrm{AlmCont}_p(M; \mathcal{W})^C$ is a weak equivalence and so $\|\mathrm{Cont}_\bullet^{\mathrm{OT}}(M; \mathcal{W})^C\| \to \|\mathrm{AlmCont}_\bullet(M; \mathcal{W})^C\|$ is also a weak equivalence. □

Furthermore, we have the following, which relates $\|\mathrm{AlmCont}_\bullet(M; \mathcal{W})^C\|$ to $\mathrm{AlmCont}(M)$:

Lemma 4.8 *The map $\|\mathrm{AlmCont}_\bullet(M; \mathcal{W})^C\| \to \mathrm{AlmCont}(M)$ is a weak equivalence.*

Proof The proof is similar to the proof of Theorem 2.6, so we will omit some technical details that were included there. Let $\alpha\colon \mathbb{D}^k \to \mathrm{AlmCont}(M)$ be a continuous map such that we have a commutative diagram

$$\begin{array}{ccc} S^{k-1} & \xrightarrow{\partial\alpha} & \|\mathrm{AlmCont}_\bullet(M; \mathcal{W})^C\| \\ \downarrow & & \downarrow \\ \mathbb{D}^k & \xrightarrow{\alpha} & \mathrm{AlmCont}(M) \end{array}$$

where by abuse of notation $\partial\alpha$ is the map $S^{k-1} \to \|\mathrm{AlmCont}_\bullet(M; \mathcal{W})^C\|$, $p \mapsto (\alpha(p), \vec{w}_p, \vec{t}_p)$, where $\vec{w}_p = (w_1(p), \dots w_\ell(p))$ is a finite ordered set of overtwisted disks for $\alpha(p)$ and $\vec{t}_p = (t_1(p), \dots, t_\ell(p))$ are their corresponding weights. We can assume all of these weights appearing are nonzero.

There is only one part of the proof of Theorem 2.6 that does not go through immediately, which is finding a disk that is disjoint from all the disks $w_i(p)$ for all p, i. The problem is that since the embedded disks are no longer locally constant but rather can vary from point to point, we have S^{k-1} families of disks instead of finitely many, so it is possible that they cover all of M. However, we can get around this as follows. First, we can make a buffer away from the boundary by replacing α with a radial compression to the disk of radius $\frac{1}{2}$, which is homotopic to α relative to the boundary. Again by abuse of notation we will let α denote this new map. Let A be the annulus of radius $\frac{1}{2}$, so now the map α lifts to $\|\mathrm{AlmCont}_\bullet(M;\mathcal{W})^C\|$ in A. Let $S^{k-1}_{1/2}$ be the sphere of radius $\frac{1}{2}$. By arguing as in Section 6.2 of [9], we can assume that for each $p \in S^{k-1}_{1/2}$ there is a contractible open neighborhood V_p of $p \in \mathbb{D}^k$ such that there is some $m \in M$ so that m is not contained in any $w_i(q)$ for any $q \in V_p \cap A$ and $1 \le i \le \ell(q)$. Furthermore, we can let $U_p \subset V_p$ be a slightly smaller contractible open neighorhood of p so that the closure of U_p, $\mathrm{cl}(U_p)$, is contractible and contained in V_p. These can be chosen so that the inclusion of $\mathrm{cl}(U_p)$ into V_p is a closed cofibration, for example by choosing them as small balls, so we assume that this is the case. Since spheres are compact, we can find a finite subcover by these smaller neighborhoods, $U_1, \dots, U_j$, which also gives us a cover by the larger neighborhoods $V_1, \dots, V_j$. By construction the disks $w_i(q)$ for $q \in V_a \cap A$ don't cover M for any given $1 \le a \le j$, so in particular we can find an embedded disk $\Delta_1 \subset M$ and a regular neighborhood D_1 of Δ_1 in M, such that D_1 is disjoint from all such disks $w_i(q)$, $q \in V_1 \cap A$. Finally, we can pick some $V_0 \subset \mathrm{int}(\mathbb{D}^k_{1/2})$ and some slightly smaller U_0 so that $U_1, \dots, U_j, U_0, A$ cover $\mathbb{D}^k$.

With all of this set up, we can now do the following. Since $\mathrm{cl}(U_1) \times \Delta_1$ is contractible, if we restrict α to this we can homotope it to agree with the overtwisted germ that comes with Δ_1. Also, we can use the homotopy extension property to extend this homotopy to one on $V_1 \times D_1$, such that on ∂D_1, ∂V_1 the homotopy is just α. Then we can extend the homotopy by α to all of $\mathbb{D}^k$, M so that we have a new map $\alpha_1 \colon \mathbb{D}^k \to \mathrm{AlmCont}(M)$ that is homotopic to α, agrees with α away from $V_1 \times D_1$, and satisfies that Δ_1 is overtwisted for $\alpha_1(p)$ for all $p \in U_1$. Also, $w_i(q)$ is still overtwisted for $\alpha_1(q)$ for all $q \in A$, since in V_1, Δ_1 is away from all of the $w_i(q)$, and outside of V_1, $\alpha_1 = \alpha$. We can repeat this on U_2, now being careful to choose Δ_2, D_2 so that D_2 is disjoint from Δ_1 as well as $w_i(q)$ for $q \in V_2 \cap A$. Then we can find α_2 homotopic to α_1 so that $\alpha_2 = \alpha_1$ outside of $V_2 \times D_2$ and Δ_2 is overtwisted for $\alpha_2(p)$ for all $p \in \mathrm{cl}(U_2)$. So, by the same reasoning Δ_1 is still overtwisted for $\alpha_2(p)$ for $p \in U_1$ and $w_i(q)$ is still overtwisted for $\alpha_2(q)$ for all $q \in A$. Repeat this for $U_3, \dots, U_j$, and get α_j, which still has the same overtwisted disks in A as well as Δ_a is overtwisted for α_j on U_a. Finally, since V_0 is disjoint from the annulus A by definition, on U_0 we just need to find an overtwisted disk Δ_0 with regular neighborhood D_0 disjoint from $\Delta_1, \dots, \Delta_j$, which we can always do. Then do the same homotopy trick as before to get α_0 that agrees with α_j outside of $V_0 \times D_0$ and has overtwisted disk Δ_0 on $\mathrm{cl}(U_0)$. Finally, let $s_0, \dots, s_j, s_A$ be a partition of unity subordinate to the cover $U_0, U_1, \dots, U_j, A$. Then we have a map $\beta \colon \mathbb{D}^k \to \|\mathrm{AlmCont}_\bullet(M;\mathcal{W})^C\|$ given by

$$p \mapsto \bigl(\alpha_0(p), \Delta_0, \Delta_1, \dots, \Delta_j, \vec{w}_{p/|p|}, s_0(p), s_1(p), \dots, s_j(p), s_A(p)\vec{t}_{p/|p|}\bigr).$$

By construction the specified disk is an overtwisted disk at p precisely when its weight is nonzero, all of the weights add up to one, and this is clearly a lift of α_0 which is homotopic to α relative to the boundary. □

Using this lemma, we can now prove Theorem 4.4; $\mathrm{OT}_k(M)$ is the image of $\pi_k\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})\|$ induced by the natural forgetful map $\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})\| \to \mathrm{Cont}^{\mathrm{OT}}(M)$ when $1 \le k \le 2n$.

Proof Recall the commutative diagram relating the different semisimplicial spaces:

$$\begin{array}{ccccc} \|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})\| & \longrightarrow & \|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})^C\| & \longrightarrow & \mathrm{Cont}^{\mathrm{OT}}(M) \\ \big\downarrow & & \big\downarrow & & \big\downarrow \\ \|\mathrm{AlmCont}_\bullet(M;\mathcal{W})\| & \longrightarrow & \|\mathrm{AlmCont}_\bullet(M;\mathcal{W})^C\| & \longrightarrow & \mathrm{AlmCont}(M) \end{array}$$

First, we will show that $\mathrm{OT}_k(M)$ is a subgroup of the image of $\pi_k\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})\|$. By abuse of notation we will use specific representatives of elements of homotopy groups when we mean their homotopy classes, so we are really working up to homotopy relative to the basepoint. Let $\alpha\colon S^k \to \mathrm{Cont}^{\mathrm{OT}}(M)$ be an element of $\mathrm{OT}_k(M)$ and $\Delta\colon S^k \to \mathrm{Emb}(\mathbb{D}^{2n}, M)$ be a certificate of overtwistedness for α. Then $(\alpha, \Delta, 1) \in \pi_k\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})^C\|$ maps to α. However, we know that the maps

$$\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})\| \to \|\mathrm{AlmCont}_\bullet(M;\mathcal{W})\| \quad \text{and} \quad \|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})^C\| \to \|\mathrm{AlmCont}_\bullet(M;\mathcal{W})^C\|$$

in the above diagram are weak equivalences. Furthermore, by the previous lemma

$$\|\mathrm{AlmCont}_\bullet(M;\mathcal{W})^C\| \to \mathrm{AlmCont}(M)$$

is a weak equivalence, and since we know that $\|\mathrm{AlmCont}_\bullet(M;\mathcal{W})\| \to \mathrm{AlmCont}(M)$ is also a weak equivalence by Theorem 2.6 we have that $\|\mathrm{AlmCont}_\bullet(M;\mathcal{W})\| \to \|\mathrm{AlmCont}_\bullet(M;\mathcal{W})^C\|$ is a weak equivalence. Combining these equivalences with the previous commutative diagram, we have that $\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})\| \to \|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})^C\|$ is a weak equivalence, and so there exists some $\beta \in \pi_k\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})\|$ such that $\beta \mapsto (\alpha, \Delta, 1)$ and hence maps to α. So indeed, $\mathrm{OT}_k(M)$ is a subgroup of the image of $\pi_k\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})\|$. However, we know that the isomorphisms

$$\mathrm{OT}_k(M) \to \pi_k\,\mathrm{AlmCont}(M) \quad \text{and} \quad \pi_k\|\mathrm{Cont}^{\mathrm{OT}}_\bullet(M;\mathcal{W})\| \to \pi_k\,\mathrm{AlmCont}(M)$$

are both induced by the natural inclusion $\mathrm{Cont}^{\mathrm{OT}}(M) \to \mathrm{AlmCont}(M)$, so we have a group isomorphism that remains an isomorphism when restricted to a subgroup. This is only possible if the subgroup is the whole group, so indeed $\mathrm{OT}_k(M)$ is this image. □

5 Infinite cyclic subgroups in the homotopy groups of the contactomorphism group

We can now use Theorem 4.1 to generalize the results from [8]. Let (M, ξ_{OT}) be a compact, connected, cooriented, overtwisted contact manifold of dimension $2n+1$ with (possibly empty) boundary, and let $\mathcal{C}_0(M, \xi_{\mathrm{OT}} \operatorname{rel} \partial M) \subset \mathrm{Diff}_0(M \operatorname{rel} \partial M)$ be the group of contactomorphisms of (M, ξ_{OT}) relative to the boundary, ie all diffeomorphisms of M that are isotopic to the identity, agree with the identity near the boundary, and preserve the contact structure ξ_{OT}. Then from [10, Lemma 1.1] we have a fiber sequence

$$\mathcal{C}_0(M, \xi_{\mathrm{OT}} \operatorname{rel} \partial M) \to \mathrm{Diff}_0(M \operatorname{rel} \partial M) \to \mathrm{Cont}(M \operatorname{rel} \partial M),$$

where $\mathrm{Diff}_0(M \mathrm{\ rel\ } \partial M) \to \mathrm{Cont}(M \mathrm{\ rel\ } \partial M)$ is given by $f \mapsto f^\star \xi_{\mathrm{OT}}$, which induces a long exact sequence of homotopy groups.

Remark 5.1 In the literature the fibration is given by pushforward not pullback, ie the map $f \mapsto f_\star \xi_{\mathrm{OT}}$. While this may be more natural geometrically, since we are using diffeomorphisms pullback is just pushforward by the inverse, so our map is still a fibration. Also, we will see later that it is convenient to factor this map through something more general, where pushforward is no longer well defined but pullback is.

We would like to find infinite cyclic subgroups inside $\pi_k(\mathcal{C}_0(M, \xi_{\mathrm{OT}} \mathrm{\ rel\ } \partial M), \mathrm{id})$, ie we want to find nonzero elements of the rational homotopy groups of $\mathcal{C}_0(M, \xi_{\mathrm{OT}} \mathrm{\ rel\ } \partial M)$. To do this, we will prove a few lemmas. Let $\mathrm{Bun}_\partial(TM)$ denote the space of all pairs $(f, \delta f)$, where $f : M \to M$ is a smooth map which agrees with the identity map near the boundary, and $\delta f : TM \to f^\star TM$ is some vector bundle map over M that agrees with the identity near the boundary and is a fiberwise isomorphism, ie the space of bundle isomorphisms of TM.

Lemma 5.2 *The map*

$$\mathrm{Diff}_0(M \mathrm{\ rel\ } \partial M) \to \mathrm{Cont}(M \mathrm{\ rel\ } \partial M) \to \mathrm{AlmCont}(M \mathrm{\ rel\ } \partial M)$$

factors through the space of bundle isomorphisms of TM as

$$\begin{tikzcd}
\mathrm{Diff}_0(M \mathrm{\ rel\ } \partial M) \arrow[r, "f \mapsto f^\star \xi_{\mathrm{OT}}"] \arrow[d] & \mathrm{Cont}(M \mathrm{\ rel\ } \partial M) \arrow[d] \\
\mathrm{Bun}_\partial(TM) \arrow[r, "(f,\delta f) \mapsto (f,\delta f)^\star \xi_{\mathrm{OT}}"] & \mathrm{AlmCont}(M \mathrm{\ rel\ } \partial M)
\end{tikzcd}$$

where the left vertical map is the derivative $f \mapsto (f, df)$, and $(f, \delta f)^* \xi_{\mathrm{OT}}$ is the almost contact structure obtained by realizing $f^* \xi_{\mathrm{OT}} \subset f^* TM$ as a subbundle of TM via the isomorphism $\delta f : TM \to f^* TM$.

Proof It is clear the diagram commutes as long as the bottom map is well defined, so to verify this, we just need to ensure that if $(f, \delta f) \in \mathrm{Bun}_\partial(TM)$ then $(f, \delta f)^\star \xi_{\mathrm{OT}}$ is still an almost contact structure on M, which agrees with ξ_{OT} near the boundary. First, $f^\star \xi_{\mathrm{OT}}$ and $f^\star R$ are always a hyperplane bundle and line bundle on M respectively, for any smooth function $f : M \to M$. Also, the Whitney sum decomposition and the almost complex structure are naturally preserved by pullback. Since both f and δf agree with the identity near the boundary, the pullback agrees with ξ_{OT} near the boundary, so the only possible obstruction is that $f^\star \xi_{\mathrm{OT}}$ and $f^\star R$ are not necessarily isomorphic to subbundles of TM for an arbitrary smooth function. However, we are given a fiberwise isomorphism $\delta f : TM \to f^\star TM$ which allows us to realize these pullbacks as subbundles of the tangent bundle, as required. □

Next, we need some results about how the diffeomorphism group of a disk glued into M maps to $\mathrm{Cont}(M \mathrm{\ rel\ } \partial M)$. First, we can use a recent result in [6, Theorem 1.4] which says that the inclusion map

$$\mathrm{Diff}_0(\mathbb{D}^{2n+1} \mathrm{\ rel\ } \partial \mathbb{D}^{2n+1}) \to \mathrm{Diff}_0(M \mathrm{\ rel\ } \partial M)$$

is injective on rational homotopy in degrees k in the rational concordance range $k \leq \phi^{\mathbb{Q}}(\mathbb{D}^{2n}) - 1$, $k \neq 0$. So, if we can find something nontrivial in $\pi_k \mathrm{Diff}_0(\mathbb{D}^{2n+1} \operatorname{rel} \partial \mathbb{D}^{2n+1}) \otimes \mathbb{Q}$ that maps to zero in $\pi_k \operatorname{Cont}(M \operatorname{rel} \partial M) \otimes \mathbb{Q}$ in this range, then by the injectivity result we will have a nontrivial element of $\pi_k \mathrm{Diff}_0(M \operatorname{rel} \partial M) \otimes \mathbb{Q}$ that maps to zero in $\pi_k \operatorname{Cont}(M \operatorname{rel} \partial M) \otimes \mathbb{Q}$, which will give us a nontrivial element of $\pi_k \mathcal{C}_0(M, \xi_{\mathrm{OT}} \operatorname{rel} \partial M) \otimes \mathbb{Q}$ by exactness. Let $\mathrm{Diff}_0(\mathbb{D}^{2n+1} \operatorname{rel} \partial \mathbb{D}^{2n+1}) \to \operatorname{Cont}(M \operatorname{rel} \partial M)$ be given by composing through $\mathrm{Diff}_0(M \operatorname{rel} \partial M)$.

Lemma 5.3 *Let* $\alpha \in \pi_k \mathrm{Diff}_0(\mathbb{D}^{2n+1} \operatorname{rel} \partial \mathbb{D}^{2n+1})$ *be such that* $\alpha \mapsto 0$ *under the map*

$$\pi_k \mathrm{Diff}_0(\mathbb{D}^{2n+1} \operatorname{rel} \partial \mathbb{D}^{2n+1}) \to \pi_k \operatorname{Cont}(M \operatorname{rel} \partial M) \to \pi_k \operatorname{AlmCont}(M \operatorname{rel} \partial M).$$

Then we also have that $\alpha \mapsto 0$ *under the map*

$$\pi_k \mathrm{Diff}_0(\mathbb{D}^{2n+1} \operatorname{rel} \partial \mathbb{D}^{2n+1}) \to \pi_k \operatorname{Cont}(M \operatorname{rel} \partial M).$$

Proof Let Δ be an overtwisted disk for ξ_{OT}, and let $D \simeq \mathbb{D}^{2n+1}$ be an embedded disk in the interior of M disjoint from a neighborhood of Δ. Then since we get a diffeomorphism of M by extending a diffeomorphism of the disk by the identity, we have that Δ is overtwisted for $f^\star \xi_{\mathrm{OT}}$ for all $f \in \mathrm{Diff}_0(D \operatorname{rel} \partial D)$. Then we have that the map $\mathrm{Diff}_0(\mathbb{D}^{2n+1} \operatorname{rel} \partial \mathbb{D}^{2n+1}) \to \operatorname{Cont}(M \operatorname{rel} \partial M)$ factors through $\|\operatorname{Cont}^{\mathrm{OT}}_\bullet(M \operatorname{rel} \partial M; \mathcal{W})\|$, where $f \mapsto (f^\star \xi_{\mathrm{OT}}, \Delta, 1)$, and of course the map

$$\|\operatorname{Cont}^{\mathrm{OT}}_\bullet(M \operatorname{rel} \partial M; \mathcal{W})\| \to \operatorname{Cont}(M \operatorname{rel} \partial M)$$

just comes from forgetting the overtwisted disks. So, we have the following commutative diagram:

$$\begin{tikzcd}
\mathrm{Diff}_0(\mathbb{D}^{2n+1} \operatorname{rel} \partial \mathbb{D}^{2n+1}) \arrow[r] \arrow[d] & \operatorname{Cont}(M \operatorname{rel} \partial M) \arrow[d] \\
\|\operatorname{Cont}^{\mathrm{OT}}_\bullet(M \operatorname{rel} \partial M; \mathcal{W})\| \arrow[r] \arrow[ur] & \operatorname{AlmCont}(M \operatorname{rel} \partial M)
\end{tikzcd}$$

But, as we saw in a previous section the map $\|\operatorname{Cont}^{\mathrm{OT}}_\bullet(M \operatorname{rel} \partial M; \mathcal{W})\| \to \operatorname{AlmCont}(M \operatorname{rel} \partial M)$ is a weak equivalence, so indeed anything mapping to zero in $\pi_k \operatorname{AlmCont}(M \operatorname{rel} \partial M)$ must map to zero in $\pi_k \|\operatorname{Cont}^{\mathrm{OT}}_\bullet(M \operatorname{rel} \partial M; \mathcal{W})\|$ and thus in $\pi_k \operatorname{Cont}(M \operatorname{rel} \partial M)$ as well. □

Now, we have enough to prove the following lemma:

Lemma 5.4 *The map* $\mathrm{Diff}_0(\mathbb{D}^{2n+1} \operatorname{rel} \partial \mathbb{D}^{2n+1}) \to \operatorname{Cont}(M \operatorname{rel} \partial M)$ *is trivial on rational homotopy groups.*

Proof From Lemma 5.2, we can factor the map through bundle isomorphisms, as follows:

$$\begin{tikzcd}
\mathrm{Diff}_0(\mathbb{D}^{2n+1} \operatorname{rel} \partial \mathbb{D}^{2n+1}) \arrow[r] \arrow[d] & \mathrm{Diff}_0(M \operatorname{rel} \partial M) \arrow[r] \arrow[d] & \operatorname{Cont}(M \operatorname{rel} \partial M) \arrow[d] \\
\mathrm{Bun}_\partial(\mathbb{D}^{2n+1}) \arrow[r] & \mathrm{Bun}_\partial(TM) \arrow[r] & \operatorname{AlmCont}(M \operatorname{rel} \partial M)
\end{tikzcd}$$

where the left vertical map is the derivative

$$d\colon \mathrm{Diff}_0(\mathbb{D}^{2n+1} \operatorname{rel} \partial\mathbb{D}^{2n+1}) \to \Omega^{2n+1}\,\mathrm{SO}(2n+1).$$

But, this map is zero on rational homotopy groups, see for example [5, page 9]. So, the composition through the bottom of the diagram to $\mathrm{AlmCont}(M \operatorname{rel} \partial M)$ is zero on rational homotopy. So, since the diagram is commutative, we can go through the top, and use Lemma 5.3 to conclude $\mathrm{Diff}_\partial(\mathbb{D}^{2n+1} \operatorname{rel} \partial\mathbb{D}^{2n+1}) \to \mathrm{Diff}_0(M \operatorname{rel} \partial M) \to \mathrm{Cont}(M \operatorname{rel} \partial M)$ is trivial on rational homotopy groups, as required. $\square$

Corollary 5.5 *If (M, ξ_{OT}) is a compact, cooriented, overtwisted contact manifold of dimension $2n+1$, then $\pi_k\mathcal{C}_0(M, \xi_{\mathrm{OT}} \operatorname{rel} \partial M)$ contains an infinite cyclic subgroup whenever*

$$\pi_k\mathrm{Diff}_0(\mathbb{D}^{2n+1} \operatorname{rel} \partial\mathbb{D}^{2n+1}) \otimes \mathbb{Q} \neq 0, \quad \textit{for } k \leq \phi^{\mathbb{Q}}(\mathbb{D}^{2n}) - 1,\ k \neq 0.$$

Proof By the injectivity result of [6], every nonzero element of

$$\pi_k\mathrm{Diff}_0(\mathbb{D}^{2n+1} \operatorname{rel} \partial\mathbb{D}^{2n+1}) \otimes \mathbb{Q}$$

has nonzero image in $\pi_k\mathrm{Diff}_0(M \operatorname{rel} \partial M)\otimes\mathbb{Q}$, which is then mapped to zero in $\pi_k\,\mathrm{Cont}^{\mathrm{OT}}(M \operatorname{rel} \partial M)\otimes\mathbb{Q}$ by Lemma 5.4. By exactness, there must be a nonzero element in $\pi_k\mathcal{C}_0(M, \xi_{\mathrm{OT}} \operatorname{rel} \partial M) \otimes \mathbb{Q}$. $\square$

6 Further applications

Another immediate application of Theorem 3.1 appears in Engel geometry, where it was recently shown that there is a notion of overtwistedness parallel to contact overtwistedness, and one still gets an h-principle with a fixed overtwisted disk; see [16, Theorem 1.1 and Corollary 1.2]. All of the relevant properties of contact overtwistedness also apply to Engel overtwistedness; the collection of all overtwisted Engel disks is still sufficiently separated, and Theorem 1.1 of [16] gives a strong enough relative h-principle that we can get an h-principle for any number of fixed overtwisted disks. We can apply Theorem 3.1 to conclude that for a 4-manifold M, $\mathcal{E}(M) \to \mathcal{E}^f(M)$ admits a section up to homotopy, where $\mathcal{E}(M), \mathcal{E}^f(M)$ are the spaces of Engel and formal Engel structures on M, respectively. This shows that $\pi_k\mathcal{E}^f(M)$ is a subgroup of $\pi_k\mathcal{E}(M)$ for all k via the map from the semisimplicial realization. Furthermore, using the foliated version of this h-principle from [16, Theorem 6.25], this subgroup agrees with the subgroup found using a certificate of overtwistedness in degrees $k \leq 3$ in [16], using the same arguments used in Theorem 4.4. It was already known that the map $\mathcal{E}(M) \to \mathcal{E}^f(M)$ is surjective on homotopy groups by [3], and one of the main results of [4] shows $\pi_k\mathcal{E}^f(M)$ is a subgroup of $\pi_k\mathcal{E}(M)$ for all k. The natural question this raises is whether the subgroup of $\pi_k\mathcal{E}(M)$ found in [4] using loose Engel structures is the same as the subgroup one gets using semisimplicial realization via overtwisted disks. Understanding this may help to understand how loose and overtwisted Engel structures interact, which is currently poorly understood.

References

[1] **D Bennequin**, *Entrelacements et équations de Pfaff*, from "Third Schnepfenried Geometry Conference, I", Astérisque 107–108, Soc. Math. France, Paris (1983) 87–161 MR Zbl

[2] **M S Borman**, **Y Eliashberg**, **E Murphy**, *Existence and classification of overtwisted contact structures in all dimensions*, Acta Math. 215 (2015) 281–361 MR Zbl

[3] **R Casals**, **J L Pérez**, **Á del Pino**, **F Presas**, *Existence h-principle for Engel structures*, Invent. Math. 210 (2017) 417–451 MR Zbl

[4] **R Casals**, **Á del Pino**, **F Presas**, *Loose Engel structures*, Compos. Math. 156 (2020) 412–434 MR Zbl

[5] **D Crowley**, **T Schick**, **W Steimle**, *The derivative map for diffeomorphism of disks: an example*, Geom. Topol. 27 (2023) 3699–3713 MR Zbl

[6] **J Ebert**, *Diffeomorphisms of odd-dimensional discs, glued into a manifold*, Algebr. Geom. Topol. 23 (2023) 2329–2345 MR Zbl

[7] **J Ebert**, **O Randal-Williams**, *Semisimplicial spaces*, Algebr. Geom. Topol. 19 (2019) 2099–2150 MR Zbl

[8] **E Fernández**, **F Gironella**, *A remark on the contactomorphism group of overtwisted contact spheres*, C. R. Math. Acad. Sci. Paris 358 (2020) 189–196 MR Zbl

[9] **S Galatius**, **O Randal-Williams**, *Stable moduli spaces of high-dimensional manifolds*, Acta Math. 212 (2014) 257–377 MR Zbl

[10] **E Giroux**, **P Massot**, *On the contact mapping class group of Legendrian circle bundles*, Compos. Math. 153 (2017) 294–312 MR Zbl

[11] **T Goodwillie**, **M Krannich**, **A Kupers**, *Stability of concordance embeddings*, Proc. Roy. Soc. Edinburgh Sect. A (online publication March 2023)

[12] **M L Gromov**, *Stable mappings of foliations into manifolds*, Izv. Akad. Nauk SSSR Ser. Mat. 33 (1969) 707–734 MR Zbl In Russian; translated in Math. USSR-Izv. 3 (1969) 671–694

[13] **K Igusa**, *The stability theorem for smooth pseudoisotopies*, K-Theory 2 (1988) vi+355 MR Zbl

[14] **M Krannich**, *A homological approach to pseudoisotopy theory, I*, Invent. Math. 227 (2022) 1093–1167 MR Zbl

[15] **M Krannich**, **O Randal-Williams**, *Diffeomorphisms of discs and the second Weiss derivative of $B\mathrm{Top}(-)$*, preprint (2021) arXiv 2109.03500

[16] **Á del Pino**, **T Vogel**, *The Engel–Lutz twist and overtwisted Engel structures*, Geom. Topol. 24 (2020) 2471–2546 MR Zbl

[17] **N P Strickland**, *The category of CGWH spaces*, preprint (2009) Available at `https://ncatlab.org/nlab/files/StricklandCGHWSpaces.pdf`

[18] **T Vogel**, *Non-loose unknots, overtwisted discs, and the contact mapping class group of S^3*, Geom. Funct. Anal. 28 (2018) 228–288 MR Zbl

Department of Mathematics, University of Toronto
Toronto, ON, Canada

`jacobw.taylor@mail.utoronto.ca`

Received: 29 March 2023 Revised: 18 December 2023

Algebraic & Geometric Topology **25**:1 (2025) 287–299
DOI: 10.2140/agt.2025.25.287
Published: 5 March 2025

Relations amongst twists along Montesinos twins in the 4-sphere

DAVID T GAY
DANIEL HARTMAN

Isotopy classes of diffeomorphisms of the 4-sphere can be described either from a Cerf-theoretic perspective in terms of loops of 5-dimensional handle attaching data, starting and ending with handles in canceling position, or via certain twists along submanifolds analogous to Dehn twists in dimension 2. The subgroup of the smooth mapping class group of the 4-sphere coming from loops of 5-dimensional handles of index 1 and 2 coincides with the subgroup generated by twists along Montesinos twins (pairs of 2-spheres intersecting transversely twice) in which one of the two 2-spheres in the twin is unknotted. We show that this subgroup is in fact trivial or cyclic of order 2.

57K40; 57K45

1 Introduction

By the smooth mapping class group of a smooth manifold X we mean $\pi_0(\mathrm{Diff}^+(X))$, where $\mathrm{Diff}^+(X)$ is the space of orientation-preserving self-diffeomorphisms of X. One way to describe a smooth mapping class is analogous to Dehn twists on surfaces: describe an explicit self-diffeomorphism of some standard neighborhood of some standard object, which is the identity on the boundary of that neighborhood, then implant this diffeomorphism via an embedding of this standard object into the ambient manifold and extend by the identity outside the neighborhood. Smooth mapping classes which are pseudoisotopic to the identity can also be described in a very different way via families of handlebodies, following Cerf [2]: If $\phi\colon X \to X$ is pseudoisotopic to the identity via a pseudoisotopy $\Phi\colon [0,1]\times X \to [0,1]\times X$, let f_t be a 1-parameter family of (generalized) Morse functions interpolating from $f_0 = \pi_{[0,1]}$, projection onto $[0,1]$, to $f_1 = f_0 \circ \Phi$; choosing an associated family of gradient-like vector fields then gives a family of handlebody structures on $X \times [0,1]$. The 1-parameter family of handle-attaching data in X then determines ϕ up to isotopy. Since f_0 and f_1 are Morse functions without critical points, this family starts and ends with canceling handle pairs.

Using the fact that every orientation-preserving diffeomorphism of S^4 is pseudoisotopic to the identity, the first author showed in [3] that every element of $\pi_0(\mathrm{Diff}^+(S^4))$ can be given in this Cerf-theoretic way by a 1-parameter family involving only 2-3-handle pairs, and that under favorable conditions (it is unclear whether these conditions might perhaps always be satisfied) such a family can be traded for a family involving a single 1-2-handle pair. Here we study the subgroup of $\pi_0(\mathrm{Diff}^+(S^4))$ coming from families of 1-2-handle pairs, the “1-2-subgroup”. In [3] the first author gave a countable list of generators for the 1-2-subgroup, explicitly described as Dehn twist-like diffeomorphisms where the embedded object is a Montesinos twin, a pair of spheres intersecting transversely at two points. Here we go back and forth between the family of handlebodies perspective and the Montesinos twists perspective to show that this 1-2-subgroup is actually generated by a single element, and that the square of this element is trivial.

We now develop these two perspectives more carefully:

Definition 1 Given an embedding $f\colon S^1 \times \Sigma \hookrightarrow X$, for some closed oriented surface Σ and some smooth oriented 4-manifold X, the *twist along* f is the isotopy class of diffeomorphisms τ_f obtained by choosing an orientation-preserving embedding $[-1,1] \times S^1 \times \Sigma$ extending f, and performing a right-handed Dehn twist along $[-1,1] \times S^1$ and the identity along Σ. As an element of $\pi_0(\mathrm{Diff}^+(X))$, τ_f is uniquely determined by the isotopy class of the embedding f.

Note that in this definition the twist τ_f is sensitive to the separate orientations of the two factors S^1 and Σ in the embedded 3-dimensional submanifold $f(S^1 \times \Sigma) \subset X$. In particular, if we reverse the orientation of $f(S^1 \times \Sigma)$ by precomposing f with an orientation-reversing diffeomorphism of S^1 then the extension to an embedding of $[-1,1] \times S^1 \times \Sigma$ needs to reverse the $[-1,1]$ direction, but in the end this reverses both factors in the annulus $[-1,1] \times S^1$, so we do not change the meaning of “right-handed Dehn twist”. On the other hand, if we reverse the orientation of $f(S^1 \times \Sigma)$ by precomposing f with an orientation-reversing diffeomorphism of Σ then we again need to reverse the $[-1,1]$ direction in the extension to $[-1,1] \times S^1 \times \Sigma$, but since we did not reverse the orientation of S^1 we did in the end reverse the orientation of the annulus factor $[-1,1] \times S^1$ and right-handed Dehn twists become left-handed Dehn twists. In short, reversing the orientation of the S^1 factor does not change τ_f while reversing the orientation of the Σ factor turns τ_f into τ_f^{-1}.

Definition 2 A *Montesinos twin* in a 4-manifold X is a pair $W = (R, S)$ of embeddings $R, S\colon S^2 \hookrightarrow X$, each with trivial normal bundle, which intersect transversely at two points. A *half-unknotted* Montesinos twin is one in which one of the two 2-spheres is unknotted.

As shown by Montesinos [5; 6] and discussed in [3], the boundary $\partial\nu(W)$ of a neighborhood of a Montesinos twin W in an oriented 4-manifold X is diffeomorphic to T^3, and if $X = S^4$ and we label the factors of T^3 as $S^1_l \times S^1_R \times S^1_S$, this parametrization of $\partial\nu(W)$ is canonically determined by the oriented isotopy class of W up to independent orientation-preserving reparametrizations of S^1_l, S^1_R and S^1_S and ambient isotopy in X. The S^1_l factor is homologically trivial in S^4, while S^1_R is a positive meridian for R and S^1_S is a positive meridian for S.

Definition 3 Given a Montesinos twin W in S^4, the *twist along* W, denoted by τ_W, is the twist along the embedding $S^1_l \times (S^1_R \times S^1_S) \hookrightarrow S^4$, as given in Definition 1. Let $\mathcal{M}$ be the subgroup of $\pi_0(\mathrm{Diff}^+(S^4))$ generated by twists along Montesinos twins. Let $\mathcal{M}_0 \subset \mathcal{M}$ be the subgroup generated by twists along half-unknotted Montesinos twins $W = (R, S)$.

As we will discuss below, $\mathcal{M}_0$ is precisely the 1-2-subgroup of $\pi_0(\mathrm{Diff}^+(S^4))$.

The following is our main result:

Theorem 4 *The group $\mathcal{M}_0$ generated by twists along half-unknotted Montesinos twins is either trivial or cyclic of order* 2.

Besides the Cerf-theoretic perspective identifying $\mathcal{M}_0$ as the 1-2-subgroup, one can also think of $\mathcal{M}_0$ as the simplest class of "twist subgroups" of $\pi_0(\mathrm{Diff}^+(S^4))$, with $\mathcal{M}$, the subgroup generated by twists along arbitrary Montesinos twins, being the next interesting case. Continuing from there, one can consider subgroups generated by twists along general embeddings of $S^1 \times \Sigma_g$ for surfaces of various genus g. It would be interesting to know if Theorem 4 can be generalized to say something about these potentially more complicated subgroups.

Acknowledgments Both authors were supported in their work on this project by National Science Foundation grant DMS-2005554 *Smooth* 4*-manifolds*: 2-, 3-, 5- *and* 6*-dimensional perspectives*.

2 The proof modulo one main calculation

There are two main ingredients in the proof of Theorem 4. The first is part of [3, Lemma 3]:

Lemma 5
$$\tau_{(S,R)} = \tau^{-1}_{(R,S)}.$$

Proof Switching R and S changes the parametrization of the boundary of a tubular neighborhood of $R \cup S$ from $S^1_l \times S^1_R \times S^1_S$ to $S^1_l \times S^1_S \times S^1_R$. However, we are not changing the orientations of R or S, and thus the orientations of the meridians S^1_R and S^1_S do not change. Therefore to keep our parametrization of the boundary of $R \cup S$ correctly oriented, we need to switch the orientation of the longitudinal S^1_l factor. Then, when we turn this into a parametrization of a neighborhood of this 3-torus as $[-1,1] \times S^1_l \times T^2$, the $[-1,1]$ direction does not change orientation (being oriented by the outward normal convention), and thus the annulus $[-1,1] \times S^1_l$ in fact *does* change orientation. Therefore the Dehn twist along this annulus switches from a positive Dehn twist to a negative Dehn twist, and so $\tau_{(S,R)} = \tau^{-1}_{(R,S)}$. □

The second ingredient expands on the connection between twists along Montesinos twins and Cerf-theoretically described diffeomorphisms of S^4 developed in [3], so as to get a full set of generators and some relations for $\mathcal{M}_0$. We now briefly set up the Cerf-theoretic picture in a little more detail.

Let $\mathrm{Emb}(S^1, S^1 \times S^3)$ be the space of embeddings of S^1 into $S^1 \times S^3$, with basepoint taken to be the embedding $S^1 \times \{p\}$ for some $p \in S^3$. (In other words, without further comment, we will only be working in the component of $\mathrm{Emb}(S^1, S^1 \times S^3)$ containing $S^1 \times \{p\}$ and we will take that to be the basepoint for $\pi_1(\mathrm{Emb}(S^1, S^1 \times S^3))$.) There is a homomorphism

$$\mathcal{H}\colon \pi_1(\mathrm{Emb}(S^1, S^1 \times S^3)) \to \pi_0(\mathrm{Diff}^+(S^4)$$

discussed in [3] (where it is called $\mathcal{FH}_1$), which we describe briefly: Given $a \in \pi_1(\mathrm{Emb}(S^1, S^1 \times S^3))$, let $\alpha_t \colon S^1 \hookrightarrow S^1 \times S^3$, with $t \in [0,1]$, be a loop of embeddings representing a, with $\alpha_0 = \alpha_1 = S^1 \times \{p\}$. Extend this to a loop of *framed* embeddings (the fact that we can do this is also explained in [3]). For each t, let Z_t be the 5-dimensional cobordism built by starting with $[0,1] \times S^4$, attaching a 5-dimensional 1-handle along a fixed standard attaching map into $\{1\} \times S^4$, to give an upper boundary canonically identified with $S^1 \times S^3$, and then attaching a 2-handle along the framed circle α_t. Note that Z_0 and Z_1 are the same 5-manifold (ie built exactly the same way), so we can put these cobordisms together to build a 6-manifold W fibering over $S^1 = [0,1]/1 \sim 0$, with fiber over $t \in S^1$ equal to Z_t, so that W itself is a cobordism from $S^1 \times S^4$ to some 4-manifold bundle over S^1. In other words, each Z_t is a cobordism from S^4 to some 4-manifold X_t, and the top boundary of W is a 5-manifold fibering over S^1 with fiber over $t \in S^1$ equal to X_t. Furthermore, since our basepoint is $S^1 \times \{p\}$, the 2-handle at $t = 0 = 1$ cancels the 1-handle, and thus X_0 can be canonically identified with S^4. Hence the top of the cobordism W is in fact an S^4-bundle over S^1 with some potentially interesting monodromy which is determined by the (homotopy class of the) loop of attaching maps α_t. This monodromy, as an element of $\pi_0(\mathrm{Diff}^+(S^4))$, is by definition $\mathcal{H}(a)$.

There is an obvious subgroup of $\pi_1(\mathrm{Emb}(S^1, S^1 \times S^3))$ in the kernel of $\mathcal{H}$, namely the subgroup of homotopy classes represented by embeddings with image equal to $S^1 \times \{p\}$, ie the subgroup corresponding to reparametrizations of the domain S^1 (or "spinning the circle in place"). By multiplying by elements of this subgroup, we can thus always assume that our loops of embeddings $\alpha_t \colon S^1 \hookrightarrow S^1 \times S^3$ have the property that the circle $\{\alpha_t(z) \mid t \in [0,1]\}$, for a fixed $z \in S^1$, is homotopically trivial in $\pi_1(S^1 \times S^3) = \mathbb{Z}$.

The connection between Montesinos twins and loops of circles in $S^1 \times S^3$ is seen as follows: Given a loop $\alpha_t \colon S^1 \hookrightarrow S^1 \times S^3$ as in the preceding paragraph, suppose that the mapped in torus $T \colon S^1 \times S^1 \to S^1 \times S^3$ defined by $T(t,z) = \alpha_t(z)$ is actually an embedding. (Budney and Gabai [1] in fact show that every element of $\pi_1(\mathrm{Emb}(S^1, S^1 \times S^3))$ can be represented by such a loop α_t.) Let C be the basepoint circle $S^1 \times \{p\}$. Note that C lies on T. Surgery along C turns the triple $(S^1 \times S^3, T, C)$ into a triple (S^4, R, S) where R is an embedded S^2 in S^4 obtained by surgering the torus T along the essential simple closed curve C and S is the embedded S^2 which is the cocore of the surgery, or belt sphere of the associated 5-dimensional 2-handle (ie the surgery replaces $S^1 \times B^3$ with $B^2 \times S^2$, and S is $\{0\} \times S^2 \subset B^2 \times S^2$). Furthermore, R and S intersect transversely at two points, namely the two "scars" on R resulting from surgering the torus down to a sphere along C. In other words, $W = (R, S)$ is a Montesinos twin in S^4, which we will call the twin associated to the loop α_t. Note that S is unknotted, since it results from surgery along $S^1 \times \{p\} \subset S^1 \times S^3$, so that W is a half-unknotted Montesinos twin. Conversely, given a Montesinos twin $W = (R, S)$ in S^4, *if we assume that S is unknotted*, then surgering (S^4, R, S) along S yields $(S^1 \times S^3, T, C)$, where $C = S^1 \times \{p\}$ and T is an embedding of $S^1 \times S^1$ which is the trace of a loop of embeddings $\alpha_t \colon S^1 \hookrightarrow S^1 \times S^3$; we will call this the loop of circles associated to the twin W.

The following lemma is implicit in the proof of [3, Theorem 4]:

Lemma 6 *Let $W = (R, S)$ be a half-unknotted Montesinos twin in S^4; for the moment assume that S is unknotted. Let α_t be the loop of circles in $S^1 \times S^3$ associated to W. Then $\tau_W = \mathcal{H}([\alpha_t])$. If, on the other hand, R is unknotted, then let $\overline{\alpha}_t$ be the loop of circles associated to $\overline{W} = (S, R)$. In this case, $\tau_W = \tau_{\overline{W}}^{-1} = \mathcal{H}([\overline{\alpha}_t])^{-1} = \mathcal{H}([\overline{\alpha}_{1-t}])$.*

Note that there are orientation conventions hidden in the above statement. In particular, one needs to understand how the orientations of R and S determine the orientations both of each circle α_t, for each t, and of the t direction in the loop of circles; equivalently, one needs to understand how the orientations of R and S correspond to the orientations of meridian and longitude for the torus $T: S^1 \times S^1 \hookrightarrow S^1 \times S^3$. The truth is that it suffices to know that there exists some orientation convention that makes it correct, but in the end we will not need to nail down that convention to get our proofs correct, because we show that the whole group involved is trivial or cyclic of order 2.

Proof We have to show that, with appropriate orientation conventions, $\tau_W = \mathcal{H}([\alpha_t])$ when $W = (R, S)$ and S is unknotted. The second half of the statement of Lemma 6, for $\overline{W}$, follows directly from Lemma 5.

Given a loop of embeddings $\alpha_t: S^1 \hookrightarrow S^1 \times S^3$, a diffeomorphism representing $\mathcal{H}([\alpha_t])$ can be defined as follows. (Note that this idea goes back to Wall's proof [7] realizing automorphisms of intersection forms of 4-manifolds by diffeomorphisms.) Using the isotopy extension theorem, let $\psi_t: S^1 \times S^3 \to S^1 \times S^3$ be an ambient isotopy starting from $\psi_0 = \mathrm{id}$, for $t \in [0,1]$, such that $\alpha_t = \psi_t \circ \alpha_0$. Since $\alpha_1 = \alpha_0 = S^1 \times \{p\} \subset S^1 \times S^3$ we can assume that ψ_1 is the identity on $S^1 \times U$ for U a 3-ball neighborhood of p. Then surgery along $\alpha_0 = S^1 \times \{p\}$ turns $S^1 \times S^3$ into S^4 in such a way that ψ_1 extends as the identity across the surgered region, and thus ψ_1 can be seen as a self-diffeomorphism of S^4, and the isotopy class of ψ_1 on S^4 is $\mathcal{H}([\alpha_t])$.

When the associated torus $T: S^1 \times S^1 \to S^1 \times S^3$ given by $T(t, b) = \alpha_t(b)$ is an embedding, then there is a standard construction of an explicit ambient isotopy ψ_t supported in a neighborhood $D^2 \times S^1 \times S^1$ of T as follows: Let $(r, \theta) \in [0,1] \times [0, 2\pi]$ be polar coordinates on D^2, with coordinates (a, b) on $S^1 \times S^1$. (We have replaced the original t variable by a because now it represents a spatial coordinate, and t will be used for the time parameter in the isotopy.) Let $f: [0,1] \to [0,1]$ be a smooth nonincreasing function which is 1 on $[0, \epsilon]$ and 0 on $[1-\epsilon, 1]$ for some suitably small positive ϵ. Then

$$\psi_t(r, \theta, a, b) = (r, \theta, a + tf(r), b)$$

is the desired isotopy. The complement of $\{r < \frac{1}{2}\epsilon\}$ in our neighborhood $D^2 \times S^1 \times S^1$ can now be parametrized (and oriented) as $\left[\frac{1}{2}\epsilon, 1\right] \times S^1_\theta \times S^1_a \times S^1_b$. This orientation agrees with the orientation as $\left[\frac{1}{2}\epsilon, 1\right] \times S^1_a \times S^1_b \times S^1_\theta$ and, with respect to this orientation, we see that ψ_1 is a *positive* Dehn twist on $\left[\frac{1}{2}\epsilon, 1\right] \times S^1_a$ crossed with the identity on $S^1_b \times S^1_\theta$.

We now need to see that the three circle parameters S^1_a, S^1_b and S^1_θ translate into S^1_l, S^1_R and S^1_S when we surger along C to turn T into a Montesinos twin (R, S). In particular we need to make sure that S^1_a becomes S^1_l so that the Dehn twist on $\left[\frac{1}{2}\epsilon, 1\right] \times S^1_a$ becomes the Dehn twist on $[-1,1] \times S^1_l$

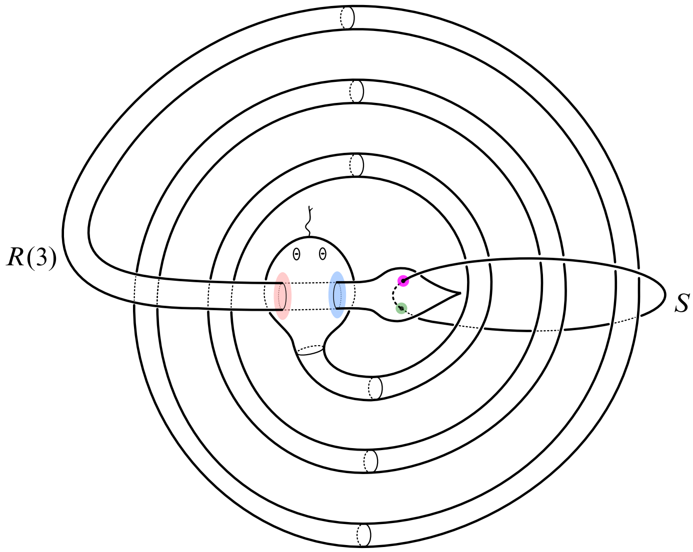


Figure 1: An illustration of $W(3) = (R(3), S)$, with the generalization to $W(i)$ being to wrap i times around instead of three times around. The red and blue disks in $R(3)$ ("ear holes" of the snake) are pushed forward and backwards in time to avoid self-intersection. We only show the equator of S, with the hemispheres lying in the past and future. The two intersection points are colored pink and green.

in our original parametrization of the neighborhood of a Montesinos twin. First of all, C is the circle $\{r = 0, a = 0, b \in S^1_b\}$ in $T = \{r = 0, a \in S^1_a, b \in S^1_b\}$. The circle S^1_θ links T and thus, when we surger T to become the 2-sphere R, S^1_θ becomes the meridian S^1_R. The circle S^1_b essentially *is* the circle C, and thus after surgery becomes the meridian S^1_S to the new 2-sphere S. Finally, in order to see that S^1_a becomes the longitudinal circle S^1_l, we just need to see that S^1_a is homologically trivial in the complement of T. This follows from the fact that S^1_a is homotopically trivial in $S^1 \times S^3$, which we arranged earlier by multiplying by an appropriate element of the domain reparametrization subgroup of $\pi_1(\mathrm{Emb}(S^1, S^1 \times S^3))$. ◻

These facts, together with the fact that $\pi_1(\mathrm{Emb}(S^1, S^1 \times S^3))$ is generated by loops which come from embedded tori, immediately give us the fact that the group of isotopy classes of diffeomorphisms of S^4 coming from loops of circles in $S^1 \times S^3$ agrees with the group generated by twists along half-unknotted Montesinos twins:

Corollary 7 $$\mathcal{H}(\pi_1(\mathrm{Emb}(S^1, S^1 \times S^3))) = \mathcal{M}_0.$$

One of the main results of [3] can be restated (combining Corollary 14 and Theorem 4 of [3] with Corollary 7) as:

Theorem 8 *$\mathcal{M}_0$ is generated by twists $\tau_{W(i)}$ for $i \in \mathbb{N}$, for the Montesinos twins $W(i) = (R(i), S)$ illustrated in* [3, Figures 1, 2 and 3]. *The loops of circles $\alpha(i)_t : S^1 \hookrightarrow S^1 \times S^3$ associated to these twins are described by the embedded tori $T(i)$ illustrated in* [3, Figure 8].

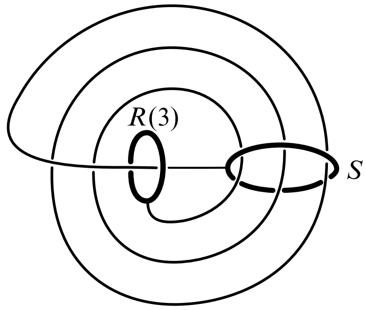


Figure 2: An alternative illustration of $W(3)$, involving two disjoint embedded 2-spheres in S^4 (the two thick circles capped off with hemispheres in past and future) and an arc connecting them. Pushing a finger from one of the spheres out along this arc and then doing a finger move when one encounters the other sphere, creating a pair of transverse intersections, gives $W(3)$. To recover Figure 1, push the finger from $R(3)$ until it meets S. However, this description is more "balanced" between $R(3)$, allowing the user to decide which sphere they prefer to draw as the complicated one.

Figures 1 and 2 reproduce two illustrations of $W(3)$ from [3]; the generalization to $W(i)$ is clear.

An important feature of the twins $W(i) = (R(i), S)$ is that *both* $R(i)$ and S are unknotted. Thus, in addition to the loops of circles $\alpha(i)_t$ associated to $W(i)$, we have loops of circles $\overline{\alpha(i)}_t$ associated to $\overline{W(i)} = (S, R(i))$. Then by Lemma 5, we know that

$$\tau^{-1}_{\overline{W(i)}} = \mathcal{H}([\overline{\alpha(i)}_t]).$$

Our main calculation in this paper is:

Proposition 9 *In $\pi_1(\mathrm{Emb}(S^1, S^1 \times S^3))$ we have $[\overline{\alpha(i)}_t] = n_i[\alpha(1)_t]$ for some integer n_i.*

(We use additive notation for $\pi_1(\mathrm{Emb}(S^1, S^1 \times S^3))$ because Budney and Gabai show that this group is abelian [1].) In fact one can show that $n_i = \pm i$, but we do not need such a precise result, and the result as stated is quicker and easier to prove. We combine the above result with the following observation:

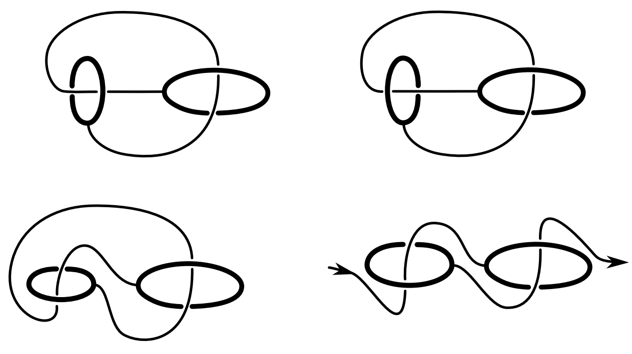


Figure 3: Isotoping $W(1)$ into a symmetric position so as to see that $W(1) = (R(1), S)$ is isotopic to $\overline{W(1)} = (S, R(1))$.

Lemma 10 *The twin $W(1) = (R(1), S)$ is isotopic (taking orientations into account) to $\overline{W(1)} = (S, R(1))$ and thus, by Lemma 5, $\tau_{W(1)} = \tau^{-1}_{\overline{W(1)}}$.*

Proof Figure 3 illustrates $W(1)$ using the "finger move" schematic of Figure 2, and then shows an isotopy to a diagram which is obviously symmetric between the two 2-spheres. □

Note that in [1] the authors discuss "barbell diffeomorphisms", clearly related to Montesinos twists, and give conditions under which barbell diffeomorphisms have order 2. Most likely Lemma 10 is a consequence of [1, Proposition 5.17].

From these three results we get:

Proof of Theorem 4 An immediate corollary of Proposition 9 is that, switching to additive notation for $\pi_0(\mathrm{Diff}^+(S^4))$,

$$-\tau_{W(i)} = \pm n_i \tau_{W(1)}.$$

From Lemma 10, we know that $\tau_{W(1)}$ has order 2 or is trivial. Since $\mathcal{M}_0$ is generated by $\{\tau_{W(i)}, i \in \mathbb{N}\}$, we conclude that $\mathcal{M}_0$ is generated by $\tau_{W(1)}$ and thus is either the trivial group or the cyclic group of order 2. □

The rest of this paper is devoted to proving Proposition 9.

3 Calculating the Budney–Gabai invariants

To prove Proposition 9, we need a picture of the loop of circles $\overline{\alpha(i)}_t$ in $S^1 \times S^3$ associated to the Montesinos twin $\overline{W(i)} = (S, R(i))$ in S^4. To get this, we need to first draw a picture of $\overline{W(i)} = (S, R(i))$ in which $R(i)$ appears as a standard unknotted S^2 and S appears as the interesting half of the twin. Then we can surger along $R(i)$ so as to draw a picture of the resulting embedded torus $\overline{T(i)}$ in $S^1 \times S^3$, from which we can understand the loop of embeddings $\overline{\alpha(i)}_t$. This will then be used to compute the W_2 invariant of $[\overline{\alpha(i)}_t] \in \pi_1(\mathrm{Emb}(S^1, S^1 \times S^3))$ defined in [1]. As the calculation will be sufficient to prove the proposition, we give a summary of Budney and Gabai's W_2 invariant below.

Figure 2 is most useful for performing the isotopy that standardizes $R(i)$ and leaves S looking complicated. To recover Figure 1 from Figure 2, one pushes the finger from the sphere labeled $R(3)$ along the arc until meeting S, and then performing a finger move there. However, one obtains an isotopic Montesinos twin by pushing the figure out along the arc starting from S until it meets $R(3)$ and then performing the finger move; this leaves $R(3)$ still "looking" like an unknot. In fact, we can first perform an isotopy to the diagram in Figure 2 to put $R(3)$ into exactly the position where S was, as in Figure 4. The final step in Figure 4 represents the result of surgering along the (now standardized) $W(3)$ to the torus $\overline{T(3)}$ in $S^1 \times S^3$. Figure 5 illustrates this final torus $\overline{T(3)}$ more explicitly in $S^1 \times S^3$, and this should be compared to Figure 6 (Figure 8 in [3]), which illustrates the original $T(3)$.

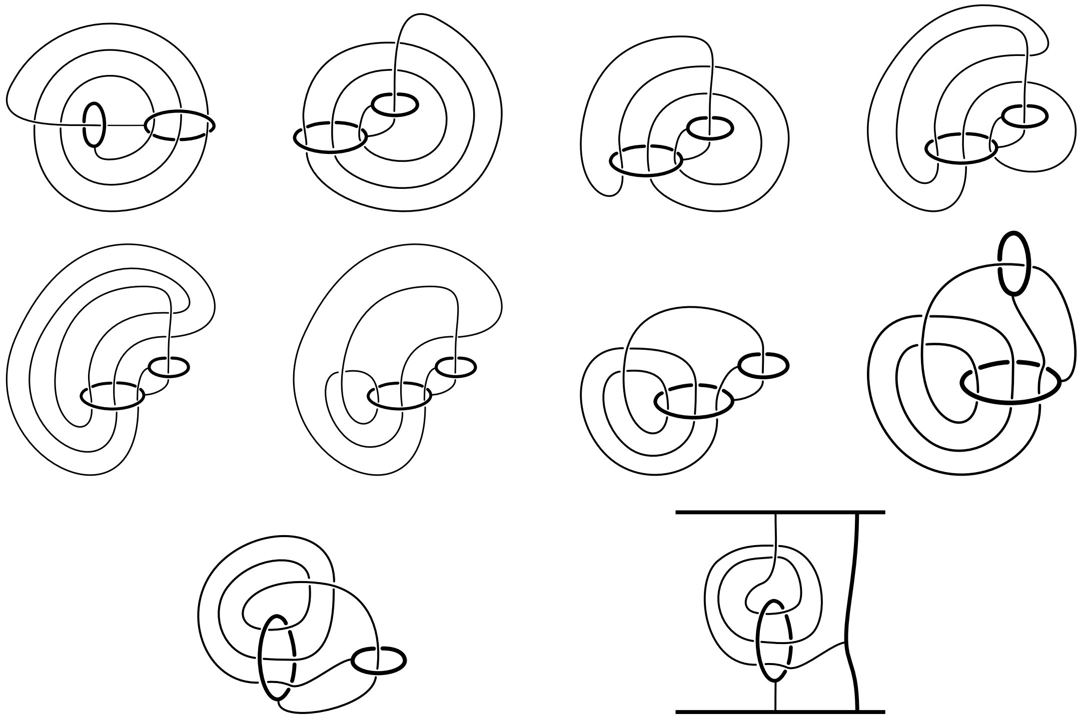


Figure 4: An isotopy of $W(3)$. The final frame should be interpreted as a diagram of an embedded torus in $S^1 \times S^3$, the result of surgering along $R(3)$. The interpretation of this diagram is made clearer in Figure 5.

Now we discuss Budney and Gabai's analysis of $\pi_1(\mathrm{Emb}(S^1, S^1 \times S^3))$ and the aforementioned W_2 invariant.

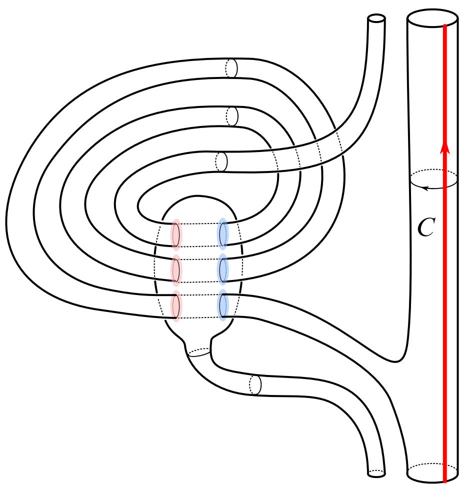


Figure 5: The embedded torus $\overline{T(3)}$ in $S^1 \times S^3$. The top is glued to the bottom, and horizontal slices are S^3's, with the "time" coordinate indicated in red/blue shading, as in Figure 1.

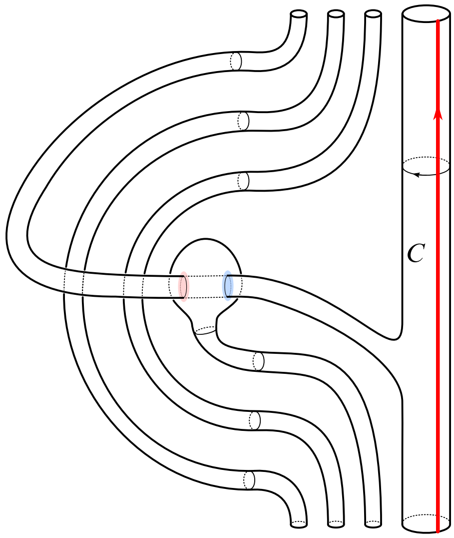


Figure 6: The embedded torus $T(3)$ in $S^1 \times S^3$, the obvious next member of the family of tori described in [1, Figure 4].

In [1, Theorem 2.9] the authors compute π_1 of all path components of $\pi_1(\mathrm{Emb}(S^1, S^1 \times S^3))$. Here we summarize their main result only for the component we care about, the component containing our chosen basepoint $S^1 \times \{p\}$: There is an isomorphism

$$W_1 \times W_2 \colon \pi_1(\mathrm{Emb}(S^1, S^1 \times S^3)) \to \mathbb{Z} \times \Lambda^0, \tag{1}$$

where

$$\Lambda^0 = \mathbb{Z}[x, x^{-1}]/\langle x^n - x^{-n} \ \forall n \in \mathbb{Z}, x^0, x^{-1} \rangle.$$

The isomorphism $W_1 \times W_2$ is the product of two homomorphisms

$$W_1 \colon \pi_1(\mathrm{Emb}(S^1, S^1 \times S^3) \to \mathbb{Z} \quad \text{and} \quad W_2 \colon \pi_1(\mathrm{Emb}(S^1, S^1 \times S^3) \to \Lambda^0.$$

The invariant W_1 detects "spinning the circle in place" as discussed in the proof of Lemma 6. We have assumed already that our loops of embeddings $\alpha_t \colon S^1 \hookrightarrow S^1 \times S^3$ have the property that for a fixed $z \in S^1$, the loop $\{\alpha_t(z), t \in [0, 1]\}$ is homotopically trivial. This directly translates to saying that we can always assume that $W_1([\alpha_t]) = 0$ for any of the loops of circles we will be considering.

Remark 11 As discussed in Lemma 6, we are free to reparametrize the domain of our loops, without affecting $\mathcal{H}$. Considering the isomorphism (1), we get that $\ker(W_2) \subset \ker(\mathcal{H})$.

The important invariant to discuss is thus W_2. We begin by reviewing the definition of W_2 found in [1]. Following the authors' notation, we will denote the two-point configuration space of a manifold M by $C_2(M)$. Let $\mathcal{CC} \subset C_2(S^1 \times S^3)$ be the submanifold of points of the form $((z_1, p), (z_2, p))$, diffeomorphic to $C_2(S^1) \times S^3$, with an orientation coming from the diffeomorphism

$$((z_1, z_2), p) \to ((z_1, p), (z_2, p)).$$

Note that $C_2(S^1)$ may be familiar to many low-dimensional topologists as the pillowcase S^2.

Given any loop of embeddings $\alpha_t : S^1 \hookrightarrow S^1 \times S^3$, the authors define a map $A : S^1 \times C_2(S^1) \to C_2(S^1 \times S^3)$ given by the formula

$$A(t, (z_1, z_2)) = (\alpha_t(z_1), \alpha_t(z_2)).$$

After a slight perturbation if necessary, we may assume A to be transverse to $\mathcal{CC}$, as the set of transverse maps is dense (but not open) [4]. However, to prove homotopy invariance, Budney and Gabai work instead with the Fulton–MacPherson compactification of the configuration spaces. Here transversality is both open and dense, and ultimately is where two of the relations defining Λ originate from. With A transverse to $\mathcal{CC}$, one sees that $A^{-1}(\mathcal{CC})$ is just a finite collection of points. On the set $A^{-1}(\mathcal{CC})$, there is a natural Σ_2 action given by permuting the $C_2(S^1)$ coordinates. On the quotient $A^{-1}(\mathcal{CC})/\Sigma_2$, assign to $[p] \in A^{-1}(\mathcal{CC})/\Sigma_2$ the monomial $\pm x^{k_p}$. Here the sign of the monomial is given by the signed intersection number of $A(p)$ and $\mathcal{CC}$, for some representative p of $[p]$, and the degree k_p of the monomial is given by the following procedure. First we consider the coordinates of the point p: $(t, (z_1, z_2))$. Next, take the path $[z_1, z_2]$ going from z_1 to z_2 in S^1 in the positively oriented direction. As $p \in \mathcal{CC}$, the S^3 coordinate of $\alpha_t(z_1)$ equals the S^3 coordinate of $\alpha_t(z_2)$. Let B_p denote the arc in $S^1 \times S^3$ which connects $\alpha_t(z_2)$ to $\alpha_t(z_1)$ by moving along the S^1 factor in the direction opposite to the orientation, while keeping the S^3 coordinate fixed. With this, we construct a map $K_p : S^1 \to S^1 \times S^3$ by concatenating the arc $\alpha_t([z_1, z_2])$ with the arc B_p. The degree k_p is then given by

$$k_p = \deg(\pi_{S^1} \circ K_p).$$

Alternatively we can calculate k_p first by counting the signed intersection of K_p with $\{z'\} \times S^3$, for a generic choice of $z' \in S^1$. If we choose our S^3 slice away from both $\alpha_t(z_1)$ and $\alpha_t(z_2)$, then k_p can be calculated by counting the signed intersection of the arcs $\alpha_t[z_1, z_2]$ and B_p with the S^3 slice, and adding the result. So

$$k_p = \alpha_t([z_1, z_2]) \cdot (\{z'\} \times S^3) + B_p \cdot (\{z'\} \times S^3),$$

where $\cdot$ is the signed count of transverse intersection points.

Adding the monomials up over all $[p] \in A^{-1}(\mathcal{CC})/\Sigma_2$ gives the formula for W_2:

$$W_2([\alpha_t]) = \sum_{[p] \in A^{-1}(\mathcal{CC})/\Sigma_2} \pm x^{k_p}.$$

It is shown in [1] that W_2 is well defined on homotopy classes of loops when considered as a map to Λ^0. Moreover, none of the choices made affect the result as long as they are made consistently.

Proof of Proposition 9 We begin by recalling some notation. First, $T(i)$ and $\overline{T(i)}$ are all embedded tori in $S^1 \times S^3$ (see Figures 5 and 6). Each embedded torus then corresponds to a loop of embeddings of circles $\alpha(i)_t : S^1 \hookrightarrow S^1 \times S^3$ for $t \in [0, 1]$ and $\overline{\alpha(i)}_t : S^1 \hookrightarrow S^1 \times S^3$ for $t \in [0, 1]$, respectively. Each loop starts at the embedding $t \mapsto (t, p)$, depicted by the red curve C in both Figures 5 and 6. Now in [1], the authors show that $W_2([\alpha(1)_t]) = \pm x^2$. As W_2 is a homomorphism, the proposition will follow once we show that $W_2([\overline{\alpha(i)}_t]) = nx^2$, for some $n \in \mathbb{Z}$.

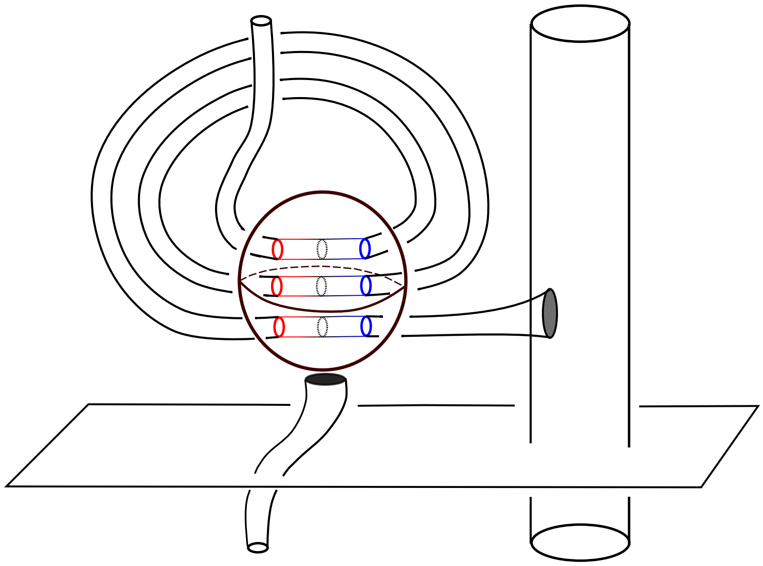


Figure 7: The tubing construction for the torus $\overline{T(3)}$. The general torus $\overline{T(i)}$ is given by using a tube which links the embedded sphere i times before it is attached to the sphere.

To begin our calculation, we note that the torus $\overline{T(i)}$ can be constructed by tubing together the "standard" torus with a 2-sphere as illustrated in Figure 7, with i equal to the number of times the tube spirals around and through the 2-sphere. The ambient space in the figure should be viewed as $[0,1]\times S^3$, with $\{0\}\times S^3$ being identified with $\{1\}\times S^3$ by the identity. Note that all the "spiraling" the tube does happens between $\{0\}\times S^3$ and $\{1\}\times S^3$. So, to calculate W_2, we will compute intersections with the sphere $\{0\}\times S^3$. Note that there are two values of t when $\alpha(i)_t$ intersects $\{0\}\times S^3$ nontransversely, but transversality of A with $\mathcal{CC}$ means that these values of t will not occur in the calculation of W_2.

Supposing (t,z_1,z_2) is a representative for $[p]\in A^{-1}(\mathcal{CC})/\Sigma_2$,

$$k_p=\overline{\alpha(i)}_t([z_1,z_2])\cdot(\{0\}\times S^3)+B_p\cdot(\{0\}\times S^3).$$

As $\overline{\alpha(i)}_t([z_1,z_2])$ is a subarc of the embedded circle $\overline{\alpha(i)}_t(S^1)$, and $\overline{\alpha(i)}_t(S^1)$ has either exactly one positive intersection or exactly two positive intersections and one negative intersection with $\{0\}\times S^3$, the signed intersection count of $\overline{\alpha(i)}_t([z_1,z_2])$ with $\{0\}\times S^3$ lies in the set $\{0,1,-1,2\}$. For the arc B_p, this is a subarc of $S^1\times\{v_p\}$ for some point $v_p\in S^3$, oriented opposite to the given orientation of S^1 in $S^1\times S^3$. As such, the intersection number of B_p with $\{0\}\times S^3$ is either 0 or -1. Putting these together, we get that

$$W_2([\overline{\alpha(i)}_t])=m_{-2}x^{-2}+m_{-1}x^{-1}+m_0x^0+m_1x^1+m_2x^2,$$

for some set of integers m_k for $k=-2,-1,0,1,2$. Finally, by considering the relations for Λ^0, we get

$$W_2([\overline{\alpha(i)}_t])=nx^2,$$

where $n=m_{-2}+m_2$. □

References

[1] **R Budney**, **D Gabai**, *Knotted* 3*-balls in* S^4, preprint (2019) arXiv 1912.09029

[2] **J Cerf**, *La stratification naturelle des espaces de fonctions différentiables réelles et le théorème de la pseudo-isotopie*, Inst. Hautes Études Sci. Publ. Math. 39 (1970) 5–173 MR Zbl

[3] **D T Gay**, *Diffeomorphisms of the* 4*-sphere*, *Cerf theory and Montesinos twins*, preprint (2021) arXiv 2102.12890

[4] **M Golubitsky**, **V Guillemin**, *Stable mappings and their singularities*, Graduate Texts in Math. 14, Springer (1973) MR Zbl

[5] **J M Montesinos**, *On twins in the four-sphere, I*, Quart. J. Math. Oxford Ser. 34 (1983) 171–199 MR Zbl

[6] **J M Montesinos**, *On twins in the four-sphere, II*, Quart. J. Math. Oxford Ser. 35 (1984) 73–83 MR Zbl

[7] **C T C Wall**, *Diffeomorphisms of* 4*-manifolds*, J. London Math. Soc. 39 (1964) 131–140 MR Zbl

Department of Mathematics, University of Georgia
Athens, GA, United States

Department of Mathematics, University of Georgia
Athens, GA, United States

dgay@uga.edu, daniel.hhartman@gmail.com

Received: 4 April 2023 Revised: 10 September 2023

Algebraic & Geometric Topology **25**:1 (2025) 301–327
DOI: 10.2140/agt.2025.25.301
Published: 5 March 2025

Complexity of 3-manifolds obtained by Dehn filling

WILLIAM JACO
JOACHIM HYAM RUBINSTEIN
JONATHAN SPREER
STEPHAN TILLMANN

Let M be a compact 3-manifold with boundary a single torus. We present upper and lower complexity bounds for closed 3-manifolds obtained as even Dehn fillings of M. As an application, we characterise some infinite families of even Dehn fillings of M for which our method determines the complexity of their members up to an additive constant. The constant only depends on the size of a chosen triangulation of M, and the isotopy class of its boundary.

We then show that, given a triangulation $\mathscr{T}$ of M with 2-triangle torus boundary, there exist infinite families of even Dehn fillings of M for which we can determine the complexity of the filled manifolds with a gap between upper and lower bounds of at most $13|\mathscr{T}|+7$. This result is bootstrapped to obtain the gap as a function of the size of an ideal triangulation of the interior of M, or the number of crossings of a knot diagram. We also show how to compute the gap for explicit families of fillings of knot complements in the 3-sphere. The practicability of our approach is demonstrated by determining the complexity up to a gap of at most 10 for several infinite families of even fillings of the figure-eight knot, the pretzel knot $P(-2,3,7)$, and the trefoil.

57K10, 57K31, 57K32, 57Q15

1 Introduction

We define the *complexity* of a triangulable manifold M to be the minimum number of top-dimensional simplices in a semisimplicial triangulation of M. For closed irreducible manifolds in dimension 3 — the focus of this work — this notion coincides for all but three manifolds with Matveev's complexity [23] that was defined in terms of spines. The notion of complexity is an important organising principle when studying manifolds through the lens of low-dimensional topology. For any given $n, d \in \mathbb{N}$ there are only a finite number of d-manifolds of complexity $\leq n$, and systematic census enumeration using triangulations naturally generates all triangulations up to a certain complexity. In this very precise sense, complexity is to manifolds what the crossing number is to knots.

Determining the complexity of a given manifold is a hard problem in general. Before we discuss closed 3-manifolds, note that several results on the complexity of 3-manifolds with boundary exist; see for instance Frigerio, Martelli, and Petronio [9], Ishikawa and Nemoto [12], Jaco, Rubinstein, Spreer, and Tillmann [16], and Rubinstein, Spreer and Tillmann [29] for complexity bounds on ideal triangulations, and Jaco, Johnson, Spreer, and Tillmann [13] for complexity bounds on triangulations with real boundary.

In the closed case, early lower bounds on complexity use an analysis of homology and fundamental groups (see Matveev and Pervova [24] and Pervova and Petronio [27]), or hyperbolic volume computations (see Matveev, Petronio, and Vesnin [25] and Petronio and Vesnin [28]). Bounds in terms of hyperbolic volume are only sharp in very special cases; see Fominykh, Garoufalidis, Goerner, Tarkaev, and Vesnin [8] and Vesnin, Tarkaev, and Fominykh [31]. Cha [4, Corollary 1.11] gave lower bounds in terms of Cheeger–Gromov ρ-invariants. A recent approach developed by Lackenby and Purcell [21] gives complexity bounds for hyperbolic 3-manifolds that fibre over the circle using the monodromy of the bundle. Census enumeration trivially determines the complexity of all manifolds in a given census, and hence a lower bound for all manifolds that do not appear in that census. Currently, this determines the complexity of all closed irreducible orientable 3-manifolds up to complexity 13 (see Matveev and Tarkaev [26]) — an impressive algorithmic and computational achievement.

Upper bounds usually arise from the explicit construction of triangulations, and the difficulty lies in closing the gap between upper and lower bounds. For instance, for the Weber–Seifert dodecahedral space, it is currently only known that its complexity lies between 14 (since it does not appear in the current census) and 23 (by an explicit construction of Burton, Rubinstein, and Tillmann [2]).

Here we build on observations on least-genus surface representatives of $\mathbb{Z}_2$-homology classes to produce new complexity bounds. This is the only approach currently known to provide exact complexity bounds for infinite families of closed 3-manifolds — more precisely, spherical 3-manifolds (see Jaco, Rubinstein, and Tillmann [18; 19]) and 3-manifolds modelled on $\widetilde{\mathrm{SL}_2(\mathbb{R})}$ (see Jaco, Rubinstein, Spreer, and Tillmann [15]). It also certifies complexity for some infinite classes of cusped hyperbolic 3-manifolds [16; 29].

Our new contributions to this line of work are complexity bounds up to a practical additive constant for infinite families of closed 3-manifolds obtained by Dehn filling. We prove:

Theorem 5 *Let M be an orientable compact irreducible 3-manifold with boundary an incompressible torus, and let $\mathscr{T}$ be a triangulation of M with a 2-triangle torus boundary. Then there exist infinite families of even Dehn fillings $M(\alpha_k)$ of M for $\alpha_k \in \mathbb{Q} \cup \{\infty\}$ and $k \geq 0$, such that*

$$2k \leq c(M(\alpha_k)) \leq 2k + 13|\mathscr{T}| + 7.$$

In particular, for each once-cusped hyperbolic 3-manifold M of finite volume, this gives an infinite family of closed hyperbolic 3-manifolds whose volumes converge to the volume of M and whose complexity is known up to an additive constant that only depends on M. We remark that at the time of writing, there is no infinite family of *closed* hyperbolic 3-manifolds for which the complexity is known exactly.

The *gap* in the above bound, denoted by $\mathrm{gap}(M(\alpha_k))$, is the difference between the upper and lower bounds on the complexity of $M(\alpha_k)$. Hence the above theorem provides an infinite family where the gap is $13|\mathscr{T}| + 7$. In particular,

$$\frac{\mathrm{gap}(M(\alpha_k))}{c(M(\alpha_k))} \in O\left(\frac{1}{c(M(\alpha_k))}\right).$$

We extend Theorem 5 to similar statements with input an ideal triangulation (Corollary 6) or a knot diagram (Corollary 7). None of these three results explicitly describes the filling slopes α_k. Knots in the 3-sphere have a canonical framing, and our methods can be used to determine explicit bounds for infinite families of even fillings where the gap is only a function of the number of crossings of a knot projection. A sample result of this form is:

Theorem 8 *Let K be a knot distinct from the unknot, and let D be a reduced diagram of K with n crossings. Moreover, let $M = \mathbb{S}^3 \setminus N(K)$ be the knot exterior of K with the standard framing on ∂M. Let $m_0 = 1401(n-1)$, $n_0 = m_0 2^{7m_0+2}$, and $k > n_0$. Then for the complexity of $M(2k/1)$,*

$$2(k-n_0) \le c(M(2k/1)) \le m_0 + 2k - 1.$$

The proof of Theorem 8 can be adapted to give a bound for other families of even fillings, and those families giving rise to a bound up to an additive constant are easily identified. Since every 3-manifold can be obtained from Dehn filling on a link in the 3-sphere (see Lickorish [22] and Wallace [32]), Theorem 8 can be applied in a quite broad setting. The above result is complementary to similar bounds for integral surgeries obtained by Cha [3; 5].

The reader should think of the theoretical results discussed so far as a flexible toolkit that can be applied to specific families of examples. While Theorem 8 cites a very large constant, this constant is much smaller in practical settings. We present three extended examples, analysing various families of Dehn fillings of the figure-eight knot in Section 5.1, the pretzel knot $P(-2,3,7)$ in Section 5.2, and the trefoil in Section 5.3. In several cases of infinite families of fillings allowing a constant gap, this gap is in the single digits. The goal of this extended list of examples is to demonstrate that, given a knot and very little extra information, we can determine practical upper and lower complexity bounds for infinite families of even Dehn fillings using out-of-the-box software such as Regina [1] or SnapPy [6].

Acknowledgements Jaco is partially supported by the Grayce B Kerr Foundation. Research of Rubinstein, Spreer, and Tillmann is supported in part under the Australian Research Council's Discovery funding scheme (project DP190102259). The main result was conceived whilst the authors were supported through the programme *Research in pairs* by the Mathematisches Forschungsinstitut Oberwolfach in 2017. The authors would like to thank the staff at MFO for an excellent collaboration environment. The authors would also like to thank the referee for very useful remarks that improved the presentation of the paper.

2 Background

We refer to [15] for background and standard definitions, and only recall the following two key definitions: Given a closed 3-manifold M, we define the *complexity* of M to be the minimum number $c(M)$ of tetrahedra in a triangulation of M. The *norm* $\|\phi\|$ of a nontrivial class $\phi \in H^1(M, \mathbb{Z}_2)$ is the negative of the maximal Euler characteristic of a properly embedded surface S, no component of which is a sphere or projective plane, representing the Poincaré dual of ϕ.

2.1 3-manifolds with torus boundary and the Farey tessellation

Let M be an orientable compact irreducible 3-manifold with ∂M consisting of a single incompressible torus boundary component. Let $(\mathfrak{m}, \mathfrak{l})$ be a framing of ∂M. Since ∂M is incompressible and has abelian fundamental group, $\mathrm{im}(\pi_1(\partial M) \to \pi_1(M)) \cong \pi_1(\partial M) \cong H_1(\partial M, \mathbb{Z})$. As is usual for the torus, we freely move between isotopy, homotopy, and homology classes depending on context and most efficient notation. Hence, for an isotopy class of nontrivial simple closed loops on the boundary torus $\alpha \in \mathrm{im}(\pi_1(\partial M) \to \pi_1(M))$, we refer to the nontrivial primitive class $\alpha \in H_1(\partial, \mathbb{Z})$, where $\alpha = \mathfrak{m}^q \mathfrak{l}^p$, as a *slope*, and vice versa. A slope is an *even slope* if it maps to zero in $H_1(M, \mathbb{Z}_2)$.

Proposition 1 [17, Corollary 10] *Let $\alpha \in \mathrm{im}(\pi_1(\partial M) \to \pi_1(M))$ be a slope. There is a properly embedded surface S in M with $[\partial S] = \alpha$ if and only if α is an even slope.*

This motivates the definition of the *norm* of an even slope α in M as

$$\|\alpha\| = \min\{-\chi(S) \mid S \text{ is a properly embedded surface in } M \text{ with } [\partial S] = \alpha\}.$$

We say that S is *taut* for α if S is connected, $[\partial S] = \alpha$, and $\|\alpha\| = -\chi(S)$.

Let $\mathscr{T}$ be a 0-efficient triangulation of M. Then $\mathscr{T}$ has a single vertex, and the induced triangulation $\mathscr{T}_\partial$ of ∂M has exactly two triangles and necessarily contains this vertex. We briefly sketch how the fundamental normal surfaces $\{F_i\}$ of $\mathscr{T}$, together with the dual graph of the Farey tessellation — as an organising principle of boundary slopes on $\mathscr{T}_\partial$ — can be used to compute the slope norm for an arbitrary even slope α of M. We refer to [17, Section 2] for details.

Consider the Farey tessellation $\mathscr{F}$ associated with the framing $(\mathfrak{m}, \mathfrak{l})$ for ∂M; see Figure 1. Each ideal triangle τ corresponds to an isotopy class of 1-vertex triangulations of $\mathscr{T}_\partial$. Its ideal vertices are labelled with the slopes (α, β, γ) of the edges for $\mathscr{T}_\partial$, and each ideal triangle is labelled with its unique even slope, say α, which is referred to as the *even slope* of τ. The base triangle is marked in green, and the *canonical* triangles for the even slopes in yellow. A canonical triangle is characterised by the property that the ideal vertex carrying the even slope lies between the two other ideal vertices on the boundary of the tessellation.

The dual graph to the Farey tessellation $\Gamma(\mathscr{F})$ is an infinite trivalent tree. Travelling across an arc in $\Gamma(\mathscr{F})$ corresponds to flipping an edge in $\mathscr{T}_\partial$ yielding another isotopy class of 2-vertex triangulations of the torus. On the level of the triangulation $\mathscr{T}$ of M, this edge flip is realised by *layering* an extra tetrahedron on top of $\mathscr{T}_\partial$, increasing the size of the triangulation by one.

Every isotopy class of 2-triangle triangulations of ∂M can be realised as the boundary of some triangulation of M, and hence every even slope of ∂M is an edge in some triangulation of M.

Related to this organising principle, there are two measures of distance on $\Gamma(\mathscr{F})$ of interest to us. Let τ and τ' be two ideal triangles of the Farey tessellations. By abuse of notation, we refer to their corresponding nodes in $\Gamma(\mathscr{F})$ by τ and τ' as well. Let α and α' be the even slope labels of τ and τ', respectively. By $d_{\mathscr{F}}(\tau, \tau')$ we denote the length of the unique shortest path in $\Gamma(\mathscr{F})$ between τ and τ'. By $d(\alpha, \alpha')$ we

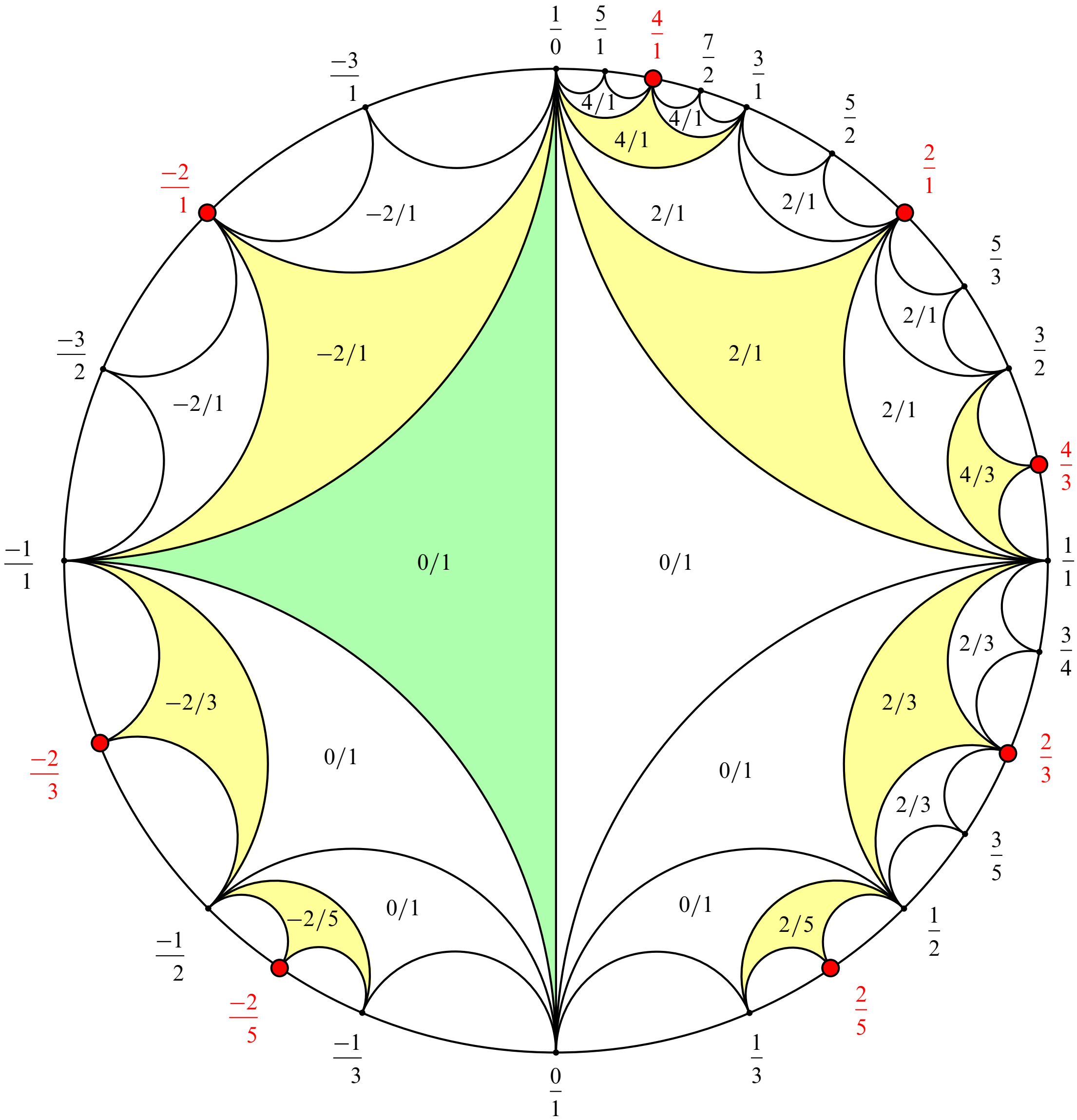


Figure 1: The Farey tessellation.

denote one less than the number of distinct even slope labels we see on the unique shortest path in $\Gamma(\mathscr{F})$ from a triangle labelled α to a triangle labelled α'. Moreover, for α an arbitrary even slope, we define $d([1],\alpha)=\infty$, where $[1]\in \operatorname{im}(\pi_1(\partial M)\to\pi_1(M))$ denotes the trivial loop. By construction, we have $2d(\alpha,\alpha')\le d_{\mathscr{F}}(\tau,\tau')$, and this bound is the best possible, as can be seen by following a path in $\Gamma(\mathscr{F})$ alternating between yellow and white ideal triangles in Figure 1.

In [17] it is shown that for some even slope α, the slope norm of α equals

$$\|\alpha\| = -\chi(S) = \min_{F_i}\{-\chi(F_i)+d([\partial F_i],\alpha)\}, \tag{2-1}$$

where the minimum is taken over all fundamental surfaces F_i of $\mathscr{T}$. Note that it is enough to minimise over the set of incompressible and ∂-incompressible fundamental surfaces of M with connected essential boundary.

Let $M(\alpha)$ be the Dehn filling of M along α. Moreover, let $S \subset M$ be taut for α. Consider the union of S and the meridian disk of the filling torus in $M(\alpha)$, and denote its Poincaré dual by $\phi_\alpha \in H^1(M(\alpha), \mathbb{Z}_2)$. By construction $\|\phi_\alpha\| = \|\alpha\| - 1$.

3 Complexity bounds on even Dehn fillings

In this section we first deduce lower and upper bounds for the complexity of $M(\alpha)$. We then describe infinite families of Dehn fillings for which the gap between these bounds is constant.

3.1 Lower bound

A *balanced lens space* is a lens space M with even fundamental group that satisfies $c(M) = 1 + 2\|\varphi\|$, where φ is a generator for $H^1(M;\mathbb{Z}_2)$. With the setup from Section 2 and the following theorem from [15], we directly obtain a lower bound for the complexity of $M(\alpha)$.

Theorem 2 [15, Corollary 2] *Let M be a closed orientable irreducible connected 3-manifold not homeomorphic with a balanced lens space and suppose that $0 \neq \varphi \in H^1(M;\mathbb{Z}_2)$. Then $c(M) \geq 2 + 2\|\varphi\|$.*

Corollary 3 *Let M be an orientable compact irreducible 3-manifold with boundary an incompressible torus, and let α be an even filling slope of M such that $M(\alpha)$ is not a balanced lens space. Then*

$$c(M(\alpha)) \geq 2\|\alpha\|, \tag{3-1}$$

where $\|\alpha\|$ denotes the slope norm of α in M.

Proof Since $M(\alpha)$ is not a balanced lens space, it follows from Theorem 2 that $c(M(\alpha)) \geq 2 + 2\|\phi_\alpha\| = 2 + 2(\|\alpha\| - 1) = 2\|\alpha\|$. □

3.2 Upper bound

Let M be an orientable compact irreducible 3-manifold with boundary an incompressible torus. Fix a framing $(\mathfrak{m}, \mathfrak{l})$ on ∂M and let $\mathscr{T}$ be a triangulation of M with a 1-vertex 2-triangle torus boundary $\mathscr{T}_\partial$. Let τ be the node in $\Gamma(\mathscr{F})$ corresponding to the isotopy class of $\mathscr{T}_\partial$.

We can turn $\mathscr{T}$ into a triangulation of a Dehn filling of M by *folding* $\mathscr{T}_\partial$ over one of its three boundary edges. That is, the two triangles in $\mathscr{T}_\partial$ are identified in such a way that one obtains a Möbius band. The edge that one folds over becomes the boundary of the Möbius band, and the other two edges are identified; see Figure 2. The kernel of the induced map on fundamental groups from the torus to the Möbius band is generated by the associated filling slope. This can be worked out from the identification of the two edges of $\mathscr{T}_\partial$ by the folding operation as follows:

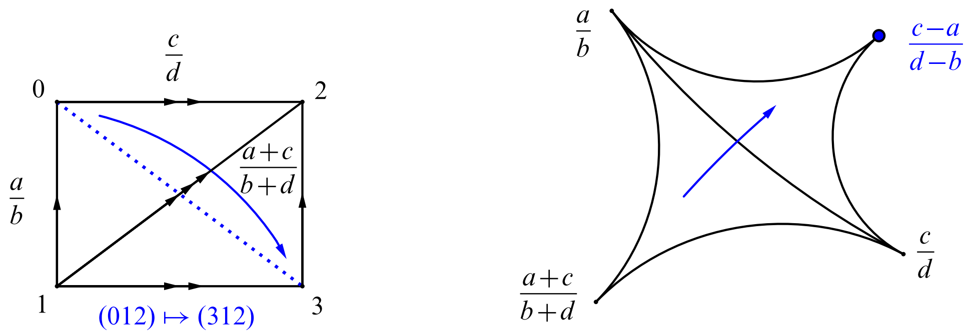


Figure 2: Left: the torus boundary $\mathscr{T}_\partial$ of isotopy class $(a/b, c/d, (a+c)/(b+d))$. The arrow indicates the folding over the diagonal and the dotted line indicates the target filling slope. Right: corresponding ideal triangle(s) in the Farey tessellation. The arrow indicates source and target triangle and the bold vertex indicates the target filling slope.

Suppose we fold over the diagonal edge in Figure 2, left. This yields the filling slope $(c-a)/(d-b)$, which is the opposite diagonal, and hence a triangulation of the manifold $M((c-a)/(d-b))$. Folding over the even edge produces a closed nonorientable surface of the same Euler characteristic as the negative of the current slope norm. This means there are two ways to relate the slope norm of an even boundary slope α to the $\mathbb{Z}_2$-norm of the associated class in the Dehn filled manifold $M(\alpha)$:

(1) layering on an ideal triangle labelled α, thereby adding an additional saddle (decreasing Euler characteristic by one), and then capping off the bounded taut surface with a disk in $M(\alpha)$ (increasing the Euler characteristic by one), or

(2) layering on one ideal triangle before a triangle labelled α, and closing the bounded taut surface by antipodal identification (leaving the Euler characteristic invariant).

Given $\mathscr{T}$ and a target even Dehn filling slope α, we can use the Farey tessellation to work out how to layer on $\mathscr{T}_\partial$ to obtain a triangulation of $M(\alpha)$ via folding: From τ, the node of $\Gamma(\mathscr{F})$ corresponding to the isotopy class of $\mathscr{T}_\partial$, layer on $\mathscr{T}_\partial$ following the unique shortest path from τ to one step before a node labelled α (if τ is already labelled α, perform one layering to obtain an isotopy class of the boundary not labelled α). Denote this target node by τ'. Now folding over the even boundary edge yields a triangulation $\mathscr{T}_\alpha$ of $M(\alpha)$; see Figure 2, left, for $\alpha = (c-a)/(d-b)$.

By construction,

$$c(M(\alpha)) \le |\mathscr{T}_\alpha| = |\mathscr{T}| + d_{\mathscr{F}}(\tau, \tau'). \tag{3-2}$$

Note that this upper bound does not only depend on $|\mathscr{T}|$, but also on the isotopy class of $\mathscr{T}_\partial$ (in (3-2) this information is incorporated in τ). This plays a role in the bound derived in Section 4, and again in Section 5, where we look at different triangulations of the figure-eight knot complement and the pretzel knot $P(-2,3,7)$ to minimise the gap between upper and lower bounds for Dehn fillings of this manifold.

Remark 4 Whenever we want to calculate the norm of an even boundary slope we must work with a 0-efficient triangulation, because this way, for every boundary slope bounding an incompressible and ∂-incompressible surface, a norm-minimising surface with this slope is amongst the fundamental surfaces in the triangulation; see [17, Lemma 13]. However, here and in the following sections we only need a guarantee that for every boundary slope of an incompressible and ∂-incompressible surface, there exists a fundamental normal surface in the triangulation with a single boundary component realising this slope. By virtue of [20, Proposition 3.7 and its corollaries], this is satisfied as soon as the triangulation has a 2-triangle torus boundary.

3.3 Families of filling slopes with constant gap

Let $\mathscr{T}$ be a triangulation of M with 2-triangle torus boundary, let $\{F_i\}$ be the finite set of fundamental normal surfaces of $\mathscr{T}$, and let $\mathscr{S}$ be the (finite) subset of vertices of $\Gamma(\mathscr{F})$ associated with the boundary slopes of those $\{F_i\}$ with a single nontrivial boundary curve in $\mathscr{T}_\partial$. Denote the vertex of $\Gamma(\mathscr{F})$ corresponding to the isotopy class of $\mathscr{T}_\partial$ by τ_0. Choose a framing $(\mathfrak{m}, \mathfrak{l})$ on ∂M such that $\tau_0 = \tau(0/1, 1/0, -1/1)$.

In $\Gamma(\mathscr{F})$, starting at node $\tau_0 = \tau(0/1, 1/0, -1/1)$, follow any infinite path τ_k for $k \geq 0$ in $\Gamma(\mathscr{F})$ where the even slope labels change at every second node. Equivalently, follow a path that alternates between nodes corresponding to white and yellow triangles; see Figure 1.

Let $\tau' \in \mathscr{S}$ be the last node of $\mathscr{S}$ along the path, with even slope label α'. Replace the path by truncating its beginning: start at τ', and remove the portion from τ_0 to τ'. Refer to every even slope α as *admissible* if α is an even slope label on the path and the previous even slope label α'' of a node τ'' on the original path is still on the truncated version of the path. Note that

$$2d(\alpha', \alpha'') \leq d_{\mathscr{F}}(\tau', \tau'') \leq 2d(\alpha', \alpha'') + 1. \tag{3-3}$$

For α admissible, and if $M(\alpha)$ is not a balanced lens space, we have for the difference between upper and lower bounds

$$0 \leq |\mathscr{T}_\alpha| - 2\|\alpha\| = |\mathscr{T}| + d_{\mathscr{F}}(\tau, \tau'') - 2\|\alpha\| \tag{3-4}$$
$$\leq |\mathscr{T}| + d_{\mathscr{F}}(\tau, \tau') + d_{\mathscr{F}}(\tau', \tau'') - 2\|\alpha\| \tag{3-5}$$
$$\leq |\mathscr{T}| + d_{\mathscr{F}}(\tau, \tau') + d_{\mathscr{F}}(\tau', \tau'') - 2d(\alpha', \alpha'') \tag{3-6}$$
$$\leq |\mathscr{T}| + d_{\mathscr{F}}(\tau, \tau') + 1. \tag{3-7}$$

Here (3-4) is the difference between (3-2) and $2\|\alpha\|$. This is nonnegative by virtue of Corollary 3. Equation (3-5) is a simple application of the triangle inequality for $d_{\mathscr{F}}$. Equation (3-6) follows from the setup of the path between τ' and τ'', the definition of $\|\cdot\|$ in (2-1), and the assumption that the slope norm of the slope corresponding to the second even slope label on the truncated path is 0. Finally (3-7) implements the more pessimistic case of (3-3).

Since neither $|\mathscr{T}|$ nor $d_{\mathscr{F}}(\tau, \tau')$ depend on the choice of admissible slope α, this determines the complexity of the infinite family of closed manifolds $\{M(\alpha)\}$ for α admissible, up to a constant.

Note that, if all members of $\{M(\alpha)\}$ are hyperbolic, we can decrease the constant by one, accounting for the fact that the norm of the first slope must be positive. Also note that this bound can be improved by looking at different triangulations $\mathscr{T}$ with different isotopy classes of $\mathscr{T}_\partial$. In particular, the choice of triangulation affects both $|\mathscr{T}|$ and $d_{\mathscr{F}}(\tau, \tau')$.

4 An upper bound for the constant gap

As above, let M be an orientable compact irreducible 3-manifold with boundary an incompressible torus. Moreover, as above, let $\mathscr{T}$ be a triangulation of M with a 2-triangle torus as boundary. In this section we compute upper bounds for $|\mathscr{T}|$ and $d_{\mathscr{F}}(\tau, \tau')$ from Section 3.3, and hence the gap in complexity, for infinite families of Dehn fillings with constant gap of M. Our bounds only depend on $|\mathscr{T}|$ (Theorem 5), the number of tetrahedra in an ideal triangulation $\mathscr{T}'$ of the interior of M (Corollary 6), or the number of crossings of a knot diagram D of a knot $K \subset \mathbb{S}^3$ in the case $M = \mathbb{S}^3 \setminus N(K)$ (Corollary 7).

In addition, we give an improvement of Corollary 7, where we have control over the knot-theoretic framing of ∂M. This allows us to determine constant gaps for explicitly chosen families of even Dehn fillings of knot exteriors only depending on the crossing number of a diagram of a knot (Theorem 8).

Theorem 5 *Let M be an orientable compact irreducible 3-manifold with boundary an incompressible torus, and let $\mathscr{T}$ be a triangulation of M with a 2-triangle torus boundary. Then there exist infinite families of even Dehn fillings $M(\alpha_k)$ of M for $\alpha_k \in \mathbb{Q} \cup \{\infty\}$ and $k \geq 0$, such that*

$$2k \leq c(M(\alpha_k)) \leq 2k + 13|\mathscr{T}| + 7.$$

Proof Since M is a 3-manifold with a single torus boundary component, every incompressible and ∂-incompressible surface in M has one of finitely many boundary slopes [11]. Since the triangulation $\mathscr{T}$ has exactly two boundary triangles, for every boundary slope of an incompressible and ∂-incompressible surface, there exists a fundamental normal surface in $\mathscr{T}$ with a single boundary component realising this slope [20, Proposition 3.7 and its corollaries].

Let $|\mathscr{T}| = n$. By the work of Hass, Lagarias, and Pippenger [10], a fundamental surface F can have at most $n2^{7n+2}$ normal arcs per boundary normal arc type. Choose a framing on M with one edge of $\mathscr{T}_\partial$ following the meridian $\mathfrak{m}$ and one following the longitude $\mathfrak{l}$, such that the isotopy class of $\mathscr{T}_\partial$ is $(0/1, 1/0, -1/1)$. It follows that ∂F intersects each of $\mathfrak{m}$ and $\mathfrak{l}$ at most $2n2^{7n+2}$ times.

Construct an infinite path in the dual of the Farey tessellation $\Gamma(\mathscr{F})$: Starting at node $\tau(0/1, 1/0, -1/1)$ go to a node τ' that we will choose in the course of the proof, which will have associated even slope $\alpha = 2p/q$ with $2p > 2n2^{7n+2} = n2^{7n+3}$. Then proceed away from $\tau(0/1, 1/0, -1/1)$ and τ' with a new even slope in every second node. In the language of Section 3.3, we call the truncated path starting at τ' the admissible path: τ' is the last node on the path possibly still contained in $\mathscr{S} \subset \Gamma(\mathscr{F})$. Denote the associated even slopes of the admissible path by α_k for $k \geq 0$, where $\alpha_0 = \alpha$.

Claim *We have $c(M(\alpha_k)) \geq 2k$ for the complexity of $M(\alpha_k)$.*

Proof of the claim Let τ be a node in $\mathscr{S} \subset \Gamma(\mathscr{F})$, and let $-\chi$ be the smallest negative Euler characteristic of a surface with slope the even slope of τ. Following [17, Algorithm 16], we compute the slope norm of α_k by taking the minimum of $-\chi$ plus the number of even slopes ($\neq \alpha_k$) observed on a path in $\Gamma(\mathscr{F})$ from τ to a node with even slope label α_k, ranging over all nodes $\tau \in \mathscr{S}$. (Note that it would be enough to only consider nodes τ associated to the slope of an incompressible ∂-incompressible surface in M.) By construction this path must pass through τ'. To see this, note that $\Gamma(\mathscr{F})$ is a tree and hence there is a unique shortest path from τ to a node labelled α_k. Assume that this path does not contain τ'. Observe that the unique shortest path from τ to $\tau(0/1, 1/0, -1/1)$ cannot contain τ' because all even slope labels on this path have numerator at most $n2^{7n+3}$. Hence, it follows that the path from $\tau(0/1, 1/0, -1/1)$ to a node labelled α_k (containing τ'), the segment between nodes labelled α_k, the segment from the second node labelled α_k to τ, and the segment from τ to $\tau(0/1, 1/0, -1/1)$ form a (not necessarily simple) cycle in the tree $\Gamma(\mathscr{F})$. This is a contradiction.

However, this implies that we see at least $k+1$ distinct even slopes on the admissible path, and we have $\|\alpha_k\| \geq k$. It then follows from Corollary 3 that $c(M(\alpha_k)) \geq 2k$, provided $M(\alpha_k)$ is not a balanced lens space. But $M(\alpha_k)$ cannot be a balanced lens space because of the nonempty sequence of layerings along the Fibonacci path described below. □

On the other hand, we can triangulate $M(\alpha_k)$ by starting with $\mathscr{T}$ and layering tetrahedra along the shortest path of $\tau(0/1, 1/0, -1/1)$ to τ'. We then need $2k-1$ more tetrahedra to layer onto the boundary to reach a boundary isotopy class that yields a triangulation $\mathscr{T}_{\alpha_k}$ of $M(\alpha_k)$ by folding the even boundary edge. Hence, in order to compute a bound for the gap up to which we can determine the complexity of $M(\alpha_k)$, it remains to bound $d_{\mathscr{F}}(\tau(0/1, 1/0, -1/1), \tau')$; see (3-2).

The shortest path from $\tau(0/1, 1/0, -1/1)$ to some node τ' with even slope coefficients larger than $n2^{7n+3}$ is the following path:

$$\tau(0/1,1/0,-1/1) \quad \tau(1/1,1/0,0/1) \quad \tau(2/1,1/1,1/0) \quad \tau(3/2,2/1,1/1) \quad \tau(5/3,3/2,2/1)$$
$$\cdots \tau(F_l/F_{l-1}, F_{l-1}/F_{l-2}, F_{l-2}/F_{l-3}), \quad \tau(F_{l+1}/F_l, F_l/F_{l-1}, F_{l-1}/F_{l-2}).$$

Here $F_0 = 0$, $F_1 = 1$, and $F_i = F_{i-1} + F_{i-2}$ for $i \geq 2$ is the Fibonacci sequence. As described above, we choose τ' and associated even slope $\alpha = F_{l+1}/F_l$, where F_{l+1} is an even Fibonacci number such that $F_{l+1} > n2^{7n+3}$. By construction, the length of the path from $\tau(0/1, 1/0, -1/1)$ to this τ' is exactly l.

We have $F_i = \lfloor \phi^i/\sqrt{5} + \frac{1}{2} \rfloor$ for $\phi = \frac{1}{2}(1+\sqrt{5}) \approx 1.618$. Observe that $\frac{1}{2}\phi^l \geq \lfloor \phi^i/\sqrt{5} + \frac{1}{2} \rfloor$ for $l \geq 2$. Since $n \geq 1$, we have $n2^{7n+3} \geq 1024$, and $l \geq 2$ can safely be assumed. It follows that we need to bound l so that $\frac{1}{2}\phi^l > n2^{7n+3}$. This translates to

$$l > \frac{1}{\log_2(\phi)}(\log_2(n) + 1 + 7n + 3).$$

Since $n > \log_2(n)$ we can instead compute l to satisfy $l > (8n+4)/\log_2(\phi) \approx 1.4404201(8n+4)$. Since every third Fibonacci number is even, $l = 12n+8$ satisfies the bound.

Altogether, this means we can triangulate $M(\alpha_k)$ by starting with $\mathscr{T}$, layering $12n+8$ tetrahedra on its boundary to obtain a triangulation with boundary isotopy class $(F_{l+1}/F_l, F_l/F_{l-1}, F_{l-1}/F_{l-2})$, followed by layering $2k-1$ additional tetrahedra on its boundary before folding over the boundary.

We thus have the upper bound

$$c(M(\alpha)) \le n + 12n + 8 + 2k - 1 = 13n + 7 + 2k. \qquad \square$$

Corollary 6 *Let M be an orientable compact irreducible 3-manifold with boundary an incompressible torus, and let $\mathscr{T}'$ be an ideal triangulation of the interior of M. Then there exist infinite families of even Dehn fillings $M(\alpha_k)$ of M for $\alpha_k \in \mathbb{Q} \cup \{\infty\}$ and $k \ge 0$, such that*

$$2k \le c(M(\alpha_k)) \le 2k + \tfrac{1}{3}(143|\mathscr{T}'| + 151).$$

Proof Let $n = |\mathscr{T}'|$ be the number of ideal tetrahedra in $\mathscr{T}'$. According to [14, Section 4.4], inflating the ideal vertex of $\mathscr{T}'$ along frame Λ in the vertex link of $\mathscr{T}'$ produces a triangulation $\mathscr{T}$ of the compact core of M with $|\mathscr{T}| = |\mathscr{T}'| + e(\Lambda) + \mathbb{X}(\Lambda) + 2$. Here $e(\Lambda)$ is the number of edges in frame Λ and $\mathbb{X}$ is a correction term accounting for the fact that conflicting diagonals of quadrilateral faces may be introduced in the inflation process, requiring extra tetrahedra to be inserted.

The vertex link of $\mathscr{T}'$ is a triangulated torus with $2n$ vertices (points on edges of $\mathscr{T}'$), $6n$ edges (normal arcs in triangles of $\mathscr{T}'$), and $4n$ triangles (normal triangles in tetrahedra of $\mathscr{T}'$). The frame Λ is a collection of edges of the vertex link with Euler characteristic -1. Hence Λ can have at most $2n+1$ edges, and so $e(\Lambda) \le 2n+1$.

Since edges in Λ are normal arcs in triangles of $\mathscr{T}'$, every triangle $t \subset \mathscr{T}'$ can contain between zero and three edges of the framing. In the case of two or three edges, inflating at t corresponds to adding a triangulated pyramid over a quadrilateral or a triangulated prism over a triangle. The diagonal in the pyramid can be freely chosen, but for the prism only six of the eight combinations of diagonals are possible. As a result, for every such t containing three edges of the frame, we may need an additional tetrahedron to flip a conflicting diagonal. In the worst case this adds another $\mathbb{X}(\Lambda) \le \lfloor \frac{1}{3} e(\Lambda) \rfloor \le \lfloor \frac{1}{3}(2n+1) \rfloor$ tetrahedra to $\mathscr{T}$.

Altogether we have

$$|\mathscr{T}| \le n + 2n + 1 + \lfloor \tfrac{1}{3}(2n+1) \rfloor + 2 \le \lfloor \tfrac{1}{3}(11n+10) \rfloor.$$

Applying Theorem 5 to $\mathscr{T}$ proves the result. $\square$

Corollary 7 *Let K be a knot distinct from the unknot, and let D be a diagram of K with n crossings. Moreover, let $M = \mathbb{S}^3 \setminus N(K)$ be the knot exterior of K.*

Then there exist infinite families of even Dehn fillings $M(\alpha_k)$ of M for $\alpha_k \in \mathbb{Q} \cup \{\infty\}$ and $k \ge 0$, such that

$$2k \le c(M(\alpha_k)) \le 2k + \tfrac{1}{3}(572n + 723).$$

Proof A well-known construction due to Weeks [33, Section 3] produces an ideal triangulation $\mathscr{T}'$ from an n-crossing diagram of a link with one cusp per link component and $4n+4$ tetrahedra. Applying the inflation in the proof of Corollary 6 to $\mathscr{T}'$ hence produces a triangulation $\mathscr{T}$ with a 2-triangle torus boundary with $|\mathscr{T}| \le \lfloor \frac{1}{3}(44n+54) \rfloor$ tetrahedra. Applying Theorem 5 to $\mathscr{T}$ proves the result. □

For the final statement of this section, we say that a diagram D of a knot K is *reduced*, if it does not allow reducing Reidemeister moves of type I or II. We call the pair of essential curves $(\mathfrak{m}_K, \mathfrak{l}_K)$ on ∂M the *knot-theoretic framing* if $\mathfrak{m}_K$ bounds a disk in $N(K)$, and $\mathfrak{l}_K$ intersects $\mathfrak{m}_K$ once and has linking number zero with K in $\mathbb{S}^3$. Determining the knot-theoretic framing first, we can give bounds for explicitly chosen infinite families of Dehn fillings of M. Here we prove this in the special case of filling slopes $2k/1$ for k sufficiently large.

Theorem 8 *Let K be a knot distinct from the unknot, and let D be a reduced diagram of K with n crossings. Moreover, let $M = \mathbb{S}^3 \setminus N(K)$ be the knot exterior of K, and let $m_0 = 1401(n-1)$, $n_0 = m_0 2^{7m_0+2}$, and $k > n_0$.*

Then we have for the complexity of $M(2k/1)$

$$2(k-n_0) \le c(M(2k/1)) \le m_0 + 2k - 1.$$

Remark 9 The focus on fillings $2k/1$ is arbitrary. Using the identical method, we can compute explicit bounds for other families of filling slopes with constant gap (as presented in Section 3.3).

Proof The proof of this statement has the following main steps and ingredients:

(1) Construct a triangulation $\mathscr{T}$ of M with boundary $\mathscr{T}_\partial$ a torus containing $\mathfrak{m}_K$ and $\mathfrak{l}_K$ as simple closed loops of edges meeting in a single vertex.

(2) Turn $\mathscr{T}$ into a triangulation $\mathscr{T}'$ with boundary $\mathscr{T}'_\partial$ a 2-triangle torus of isotopy class $(0/1, 1/0, -1/1)$ with respect to the knot-theoretic framing. In particular, one boundary edge runs along the meridian and one boundary edge runs along the longitude of the knot-theoretic framing of ∂M. This step takes up the bulk of the proof.

(3) As in Theorem 5 invoke Hatcher [11], Jaco and Sedgwick [20], and Hass, Lagarias, and Pippenger [10].

(4) Use the Farey tessellation and the known isotopy class of $\mathscr{T}'_\partial$ to show $\|2k/1\| \ge k - c$ for some constant c. The complexity of $M(2k/1)$ is bounded above by the size of $\mathscr{T}'$ and the length of a path in the dual graph of the Farey tessellation.

The triangulation $\mathscr{T}$ We apply a slightly revised construction of [10, Lemmas 7.1 and 7.2] to D. In [10], the authors first turn D into a maximal planar graph (with crossings as vertices), possibly by introducing extra vertices at bigons of D — which they call *special vertices* — and edges. Since in our case D is reduced, the number of special vertices is bounded above by n itself, and we have for the total number of vertices in the subdivided planar graph $m \le 2n$ (instead of $m \le 5n$ in [10]). The process is illustrated in Figure 3: On the left, add special vertices, giving the planar graph shown in the second step. The third step shows the result of completing to a triangulation. This yields a maximal planar graph, or planar triangulation, with $\le 4n-5$ bounded triangular regions — or triangles. We take the union of these

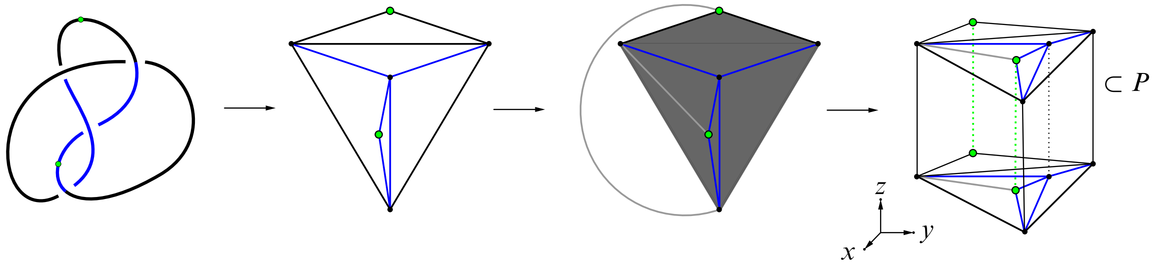


Figure 3: From D to P. The blue lines denote the edges representing K.

triangles cross an interval to obtain a collection of $\leq 4n-5$ triangular prisms, denoted by P. This is shown in the fourth step of Figure 3.

Combining [10, Lemmas 7.1 and 7.2] we only consider one layer of such prisms P (instead of three in [10]) and subdivide them into 14 tetrahedra each (with one vertex in the centre of each quadrilateral, coning over a vertex in the centre of P) to obtain a triangulation P' of P with at most $14(4n-5)=56n-70$ tetrahedra, and at most $2(4n-5)+12=8n+2$ triangles in its boundary $\partial P'$. This is shown in Figure 4. Coning $\partial P'$ to a single point at infinity, this yields a triangulation S of the 3-sphere with $\leq 64n-68<64(n-1)$ tetrahedra.

By construction, S contains the knot K as a simple closed loop L in its 1-skeleton: Follow the top (bottom) edge of a prism for an arc of D from an overcrossing (undercrossing) to an overcrossing (undercrossing). Follow the two edges in a diagonal of a quadrilateral prism face for an arc in D from an overcrossing to an undercrossing, or an undercrossing to an overcrossing, respectively. Whenever we encounter a special vertex, we first follow the appropriate edge of a triangular prism face before following the appropriate diagonal of the next quadrilateral prism face. It follows that the length of L is bounded above by $6n$.

Placing P' into $\mathbb{R}^3$ with the planar triangulations parallel to the xy-plane, and the interval in the z-direction (see Figure 3, right), we can see that D can be recovered from L by projecting a regular neighbourhood of P' in S into the xy-plane from the z-direction.

Removing a small regular neighbourhood of L from S produces either tetrahedra with neighbourhoods of zero, one, two, or three vertices removed, or tetrahedra with the neighbourhood of one edge, and zero or one vertices removed. To see this note that since D is reduced and hence does not admit any reducing

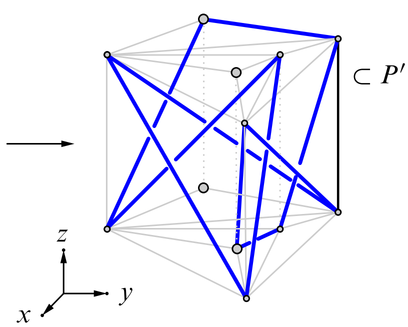


Figure 4: The triangulation P'. Subdivisions are mostly omitted for readability.

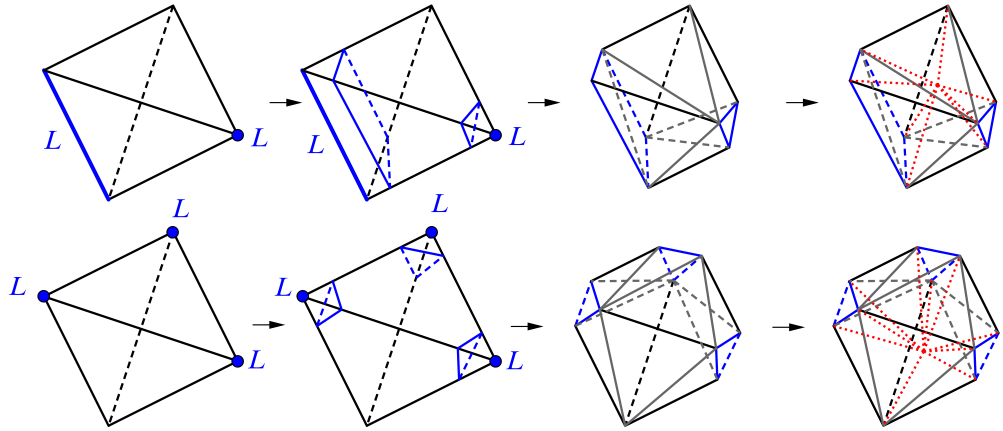


Figure 5: Removing a small neighbourhood of L from a tetrahedron followed by triangulating the resulting truncated tetrahedron. Top row: L meets the tetrahedron in an edge and a vertex. This results in a subdivision into $9+3=12$ tetrahedra. Bottom row: L meets the tetrahedron in three vertices. This results in $13+3=16$ tetrahedra.

Reidemeister II moves, at most one edge per tetrahedron in S is in L, and each tetrahedron in S has exactly one vertex that either lies at the centre of a triangular prism, or at infinity, and hence away from L.

Triangulating the boundary of these truncated tetrahedra produces at most 16 triangles. (See Figure 5 bottom row for the case realising 16 triangles. All other types of truncated tetrahedra can be triangulated with fewer tetrahedra; see for instance Figure 5, top row.) Coning these over a single vertex in its centre produces a triangulation $\mathscr{T}$ of the knot exterior of K with at most $16\cdot 64(n-1)=1024(n-1)$ tetrahedra. Note that at most three triangles per triangulated boundary of a truncated tetrahedron are in the boundary $\mathscr{T}_\partial$ of $\mathscr{T}$; see Figure 5 for some details about constructing $\mathscr{T}$.

Looking at the construction of $\mathscr{T}$ and its boundary, we can identify the geometric meridian $\mathfrak{m}_K$ of the knot exterior as a loop of six edges in the link of a special vertex. If no special vertex exist, we can create one at the beginning of the construction, and since we assume that we have as many special vertices as original vertices in our construction, this does not change our bound. We can also identify the geometric longitude $\mathfrak{l}_K$ as a simple closed path in $\mathscr{T}_\partial$: we simply run along edges in the direction of L, and realise linking number 0 with L by walking around meridian curves at nonspecial vertices as needed. Since $\mathfrak{m}_K$ lives in a neighbourhood of a special vertex, $\mathfrak{m}_K$ and $\mathfrak{l}_K$ are edge-disjoint and meet in a single vertex.

The triangulation $\mathscr{T}'$ In the next step of the construction, we turn $\mathscr{T}$ into a triangulation $\mathscr{T}'$ with a 1-vertex, 2-triangle boundary torus of isotopy class $(0/1, 1/0, -1/1)$. That is, with one of its three boundary edges running parallel to $\mathfrak{m}_K$, and another one running parallel to $\mathfrak{l}_K$.

From our calculations about the number of triangles in the truncated tetrahedra (see above), we conclude that $\mathscr{T}_\partial$ has at most $3\cdot 64(n-1)=192(n-1)$ triangles and hence at most $\frac{9}{2}64(n-1)=288(n-1)$ edges, and, since it is a torus, $96(n-1)$ vertices. In particular, its average vertex degree is 6 and we can always find a vertex v with degree ≤ 6 in $\mathscr{T}_\partial$.

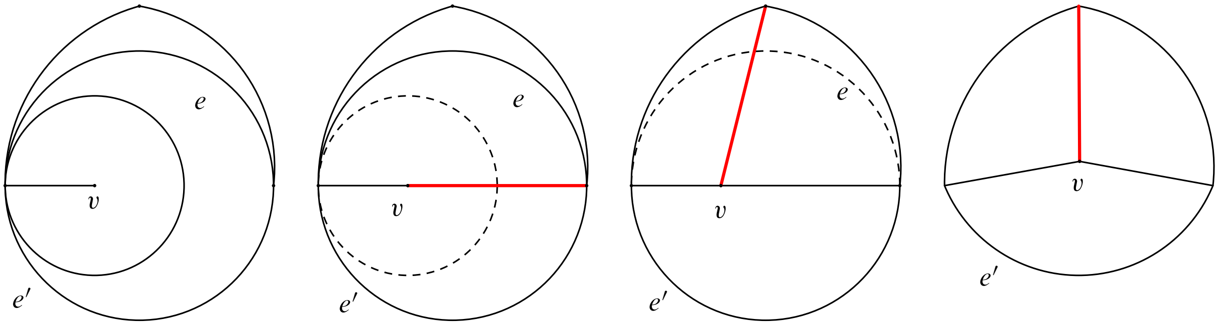


Figure 6: Turning a degree-1 vertex (left) or a degree-2 vertex (second from the left) into a vertex of degree 3 with two or one flips, respectively. Note that edges e and e' must be distinct because $\mathscr{T}_\partial$ is not a 2-sphere.

If v is of degree 1 or 2, we can layer two tetrahedra or one tetrahedron, respectively, onto the triangles adjacent to v, as shown in Figure 6, to turn v into a vertex of degree 3. Note that this is always possible since $\mathscr{T}_\partial$ is not a sphere (and hence the boundary of the two triangles around a vertex of degree 2 must consist of two distinct edges). If $\mathfrak{m}_K$ or $\mathfrak{l}_K$ ran through v (only possible in the case that v initially was of degree 2), we can find a shorter curve on the boundary of the altered triangulation isotopic to the original one.

If v is of degree 4, 5, or 6, we have three main cases:

Case 1 (the simple closed paths following $\mathfrak{m}_K$ and $\mathfrak{l}_K$ do not run through v) Here we have three subcases:

Case 1.1 (all triangles of $\mathscr{T}_\partial$ contain v at most once) We can add one, two, or three tetrahedra, respectively, onto the triangles adjacent to v, as shown in Figure 7, left, to turn v into a vertex of degree 3.

Case 1.2 (there exists a triangle containing v twice, but no triangle contains v three times) At least two triangles contain v twice and, locally, we must have the picture shown in Figure 7, right. Moreover, since $\mathscr{T}_\partial$ is not a sphere, v must be of degree at least 5. Gluing one tetrahedron, as shown in Figure 7, right, decreases the degree by 2, and causes v to have two fewer triangles containing v twice (actually, since the degree of v is at most 6, then all remaining triangles must be distinct).

Case 1.3 (there exists a triangle occurring three times) Then either we have a 1-vertex 2-triangle torus and the simple closed paths following $\mathfrak{m}_K$ and $\mathfrak{l}_K$ pass through v, or the degree of vertex v must be at least 7. Either way this is a contradiction.

Case 2 (one of $\mathfrak{m}_K$ or $\mathfrak{l}_K$ runs through v) Without loss of generality, let $\mathfrak{m}_K$ contain v, and let $\mathfrak{l}_K$ be disjoint from v. Since v is disjoint from $\mathfrak{l}_K$, it follows that $\mathfrak{m}_K$ is of length at least 2, and intersects the triangles adjacent to v in exactly two edges and at least one vertex distinct from v.

Since v has degree at most 6, it occurs in triangles on one side of $\mathfrak{m}_K$ at most five times. Fix one side. It follows from Case 1.3 that no triangle on this side contains v three times. Moreover, if an edge e contains v twice, it cannot be contained in $\mathfrak{m}_K$. Hence we can layer over e as in Case 1.2 to reduce the

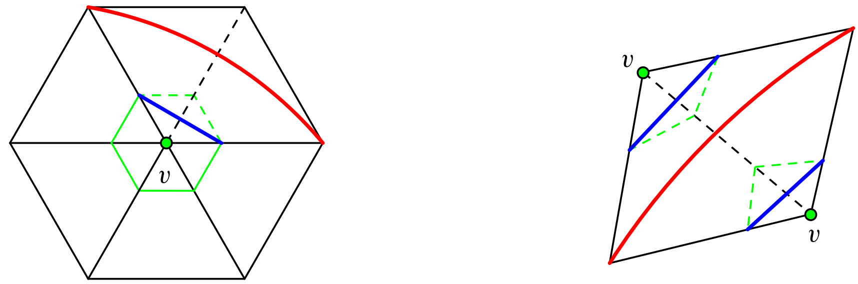


Figure 7: Left: reducing the degree of a boundary vertex v with only distinct triangles around it by one. Right: reducing the degree of a boundary vertex v contained twice in two triangles by two. The layered edge is drawn dashed, and the new boundary edge is drawn in red. Vertex linking normal curves are drawn in green. New arcs of the vertex linking curve are drawn in blue, and old arcs are dashed.

degree of v by 2 without covering an edge contained in $\mathfrak{m}_K$. If no triangle contains v more than once, we proceed as in Case 1.1, noting that we can always avoid covering an edge contained in $\mathfrak{m}_K$ in the process.

Case 3 (both $\mathfrak{m}_K$ and $\mathfrak{l}_K$ run through v) In this case, we do not touch this vertex. It must be of degree at least 6, and hence we can find another vertex of degree at most 6 to perform the above process on.

Altogether, after adding at most three tetrahedra to $\mathscr{T}$ we obtain a triangulation containing a vertex with exactly three distinct triangles around it. Hence, we can glue one additional tetrahedron to the three triangles surrounding this vertex to produce a triangulation with this vertex no longer in its boundary. Note that this is possible whenever $\mathscr{T}_\partial$ has more than one vertex, and that the boundary of this new triangulation is smaller by one vertex, three edges, and two triangles. Moreover, by construction, it still contains simple closed paths of edges running along $\mathfrak{m}_K$ and $\mathfrak{l}_K$ of equal or shorter length.

Iterating this procedure hence necessarily produces a triangulation $\mathscr{T}'$ with only two triangles in its boundary. Since $\mathscr{T}$ has at most $96(n-1)$ vertices in its boundary, the above procedure adds at most $4(96(n-1)-1) = 384(n-1)-4$ extra tetrahedra to $\mathscr{T}$ to produce $\mathscr{T}'$. It follows that $\mathscr{T}'$ contains at most $(384+1024)(n-1)-1 < 1408(n-1) =: m_0$ tetrahedra.

This part of the proof is completely analogous to the proof of Theorem 5. We sketch the argument again for the reader's convenience:

(1) Due to Hatcher, every incompressible and ∂-incompressible surface in M has one of finitely many boundary slopes [11].

(2) Due to Jaco and Sedgwick, for every boundary slope of an incompressible and ∂-incompressible surface, there exists a fundamental normal surface F in $\mathscr{T}'$ with a single boundary component realising this slope [20, Proposition 3.7 and its corollaries].

(3) Due to Hass, Lagarias, and Pippenger [10] F can have at most $n_0 = m_0 2^{7m_0+2}$ normal arcs per boundary normal arc type. Since the isotopy type of the boundary of $\mathscr{T}'$ is $(0/1, 1/0, -1/1)$, it follows that ∂F intersects each of $\mathfrak{m}_K$ and $\mathfrak{l}_K$ at most $2n_0$ times.

Deduce upper and lower bounds for the complexity of $M(2k/1)$ Starting at node $\tau(0/1, 1/0, -1/1)$, consider the following path in the dual of the Farey tessellation $\Gamma(\mathscr{F})$:

$$\tau(0/1,-1/1,1/0) \quad \tau(0/1,1/1,1/0) \quad \tau(2/1,1/1,1/0) \quad \tau(2/1,3/1,1/0) \quad \tau(4/1,3/1,1/0) \\ \cdots \tau((2k-2)/1,(2k-3)/1,1/0) \quad \tau((2k-2)/1,(2k-1)/1,1/0).$$

Layering on top of $\mathscr{T}'_\partial$ along this path and then folding over the even boundary edge produces a triangulation of $M(2k/1)$ with $m_0 + 2k - 1$ tetrahedra.

Let $\tau' = \tau((2n_0-2)/1, (2n_0-1)/1, 1/0)$ be the node corresponding to the isotopy class of the triangulation obtained from $\mathscr{T}'$ by layering $2n_0 - 1$ times on $\mathscr{T}'_\partial$ along the path above. In the language of Section 3.3, call the truncated path starting from τ' the admissible path. By construction, the even slopes of the admissible path are $2k/1$ for $k > n_0$.

Claim *Let $k > n_0$. Then $c(M(2k/1)) \geq 2(k - n_0)$.*

Proof of the claim Let τ be a node in $\Gamma(\mathscr{F})$ associated to the slope of an incompressible ∂-incompressible surface in M, and let $-\chi$ be the negative Euler characteristic of this surface. Following [17], we compute the slope norm of $2k/1$ by taking the minimum of $-\chi$ plus the number of even slopes $2k/1$ for $k > n_0$, observed on a path in $\Gamma(\mathscr{F})$ from τ to node $\tau((2k-2)/1, (2k-1)/1, 1/0)$, ranging over all such nodes τ. By construction, this path must always pass through τ', observing at least $k - n_0$ distinct even slopes (note that $2k/1$ is not one of them). Hence $\|\alpha_k\| \geq k - n_0$. It then follows from Corollary 3 and the fact that $M(2k/1)$ is not a balanced lens space that $c(M(2k/1)) \geq 2(k - n_0)$. □

On the other hand, we can triangulate $M(2k/1)$ by starting with $\mathscr{T}'$ and layering $2k - 1$ tetrahedra along the shortest path of $\tau(0/1, 1/0, -1/1)$ to $\tau((2k-2)/1, (2k-1)/1, 1/0)$. Folding the boundary then produces a triangulation $\mathscr{T}_{2k/1}$ of $M(2k/1)$.

Altogether we have

$$2(k - n_0) \leq c(M(2k/1)) \leq m_0 + 2k - 1$$

for any $k > n_0$. □

It is important to note that while the constant in Theorem 8 is prohibitively large, it can be made quite small in explicit examples. This is mainly due to the following two observations:

(a) boundary edges running parallel to $\mathfrak{m}_K$ and $\mathfrak{l}_K$ seem to be common in small triangulations $\mathscr{T}'$ of the knot exterior, and hence $|\mathscr{T}'|$ is typically very far from the upper bound given in the proof of Theorem 8, and

(b) fundamental normal surfaces often have boundary patterns with far fewer normal arcs than the bound given by Hass, Lagarias, and Pippenger.

We make this precise in Section 5 by providing examples of the actual gap in the cases of the figure-eight knot, the $(2, 3, 7)$-pretzel, and the trefoil.

5 Examples

5.1 Dehn fillings of the figure-eight knot complement

Throughout this section, let M be the complement of the figure-eight knot endowed with the knot-theoretic framing. It is well known (see for instance [30]) that, with respect to this framing, M contains three incompressible ∂-incompressible surfaces: a once-punctured torus with boundary slope $(0,1)$, and two Klein bottles with boundary slopes $(\pm 4, 1)$. Let α be an even boundary slope on ∂M. We are interested in the associated Dehn filling $M(\alpha)$. Note that since the figure-eight knot is amphichiral, $M(\alpha) \cong M(-\alpha)$.

Using a search through the Pachner graph of ideal triangulations of M, truncating and simplifying in every step, we obtain 82 combinatorially inequivalent triangulations of the compact core of M, each with ten tetrahedra and a single vertex contained in their 2-triangle boundaries. Each of them is 0-efficient. Let $\mathscr{T}$ be one of these triangulations, and let S be one of its normal surfaces. The *boundary pattern* (a, b, c) of S records the intersection numbers of S with the three boundary edges of $\mathscr{T}_\partial$.

Following [17], we know that the boundary slopes of both Klein bottles and the punctured torus must appear in the fundamental normal surfaces of $\mathscr{T}$. This allows us to determine the isotopy class of the boundary $\mathscr{T}_\partial$. As a result, the 82 triangulations exhibit four distinct isotopy classes in their boundaries; see Figure 8 for details.

Fix one of the 82 triangulations and denote it by $\mathscr{T}$. Let τ be the ideal triangle (node) in the (dual of the) Farey tessellation corresponding to the isotopy class of $\mathscr{T}_\partial$. Folding over the even boundary edge of $\mathscr{T}_\partial$ realises the even Dehn filling with slope the even slope of the ideal triangle τ' adjacent to τ opposite the even slope vertex of τ; see Section 3.2 and Figure 2 for details. Also, recall that layering over boundary edge e of $\mathscr{T}_\partial$ produces a triangulation with boundary of isotopy class the one corresponding to the adjacent ideal triangle of the Farey tessellation opposite the ideal vertex labelled with the slope of e.

In our example, all incompressible and ∂-incompressible surfaces and their boundary slopes are known, and we will never encounter triangulations of balanced lens spaces. Hence, following the instructions for folding above, obtaining the lower bound for complexity for $M(\pm\alpha)$ is straightforward: it is twice the smallest number of even slopes encountered on the unique shortest path in the dual of the Farey tessellation from one of the slopes $0/1$ and $\pm 4/1$ to a node labelled $\pm\alpha$ (note that the slope norm of all of $0/1$ and $\pm 4/1$ is one).

name	# triangulations $\mathscr{T}$	isotopy class of $\mathscr{T}_\partial$	$(0,1)$ ∂-pattern	$(4,1)$ ∂-pattern	$(-4,1)$ ∂-pattern
class I	24	$(1/0, 1/1, 2/1)$	$(1,1,2)$	$(1,3,2)$	$(1,5,6)$
class II	41	$(1/0, 0/1, 1/1)$	$(1,0,1)$	$(1,4,3)$	$(1,4,5)$
class III	4	$(1/0, 3/1, 4/1)$	$(1,3,4)$	$(1,1,0)$	$(1,7,8)$
class IV	13	$(1/0, 2/1, 3/1)$	$(1,2,3)$	$(1,2,1)$	$(1,6,7)$

Figure 8: The 82 triangulations of the compact core of the figure-eight knot complement with 10 tetrahedra. The boundary patterns and isotopy class triples follow the same order.

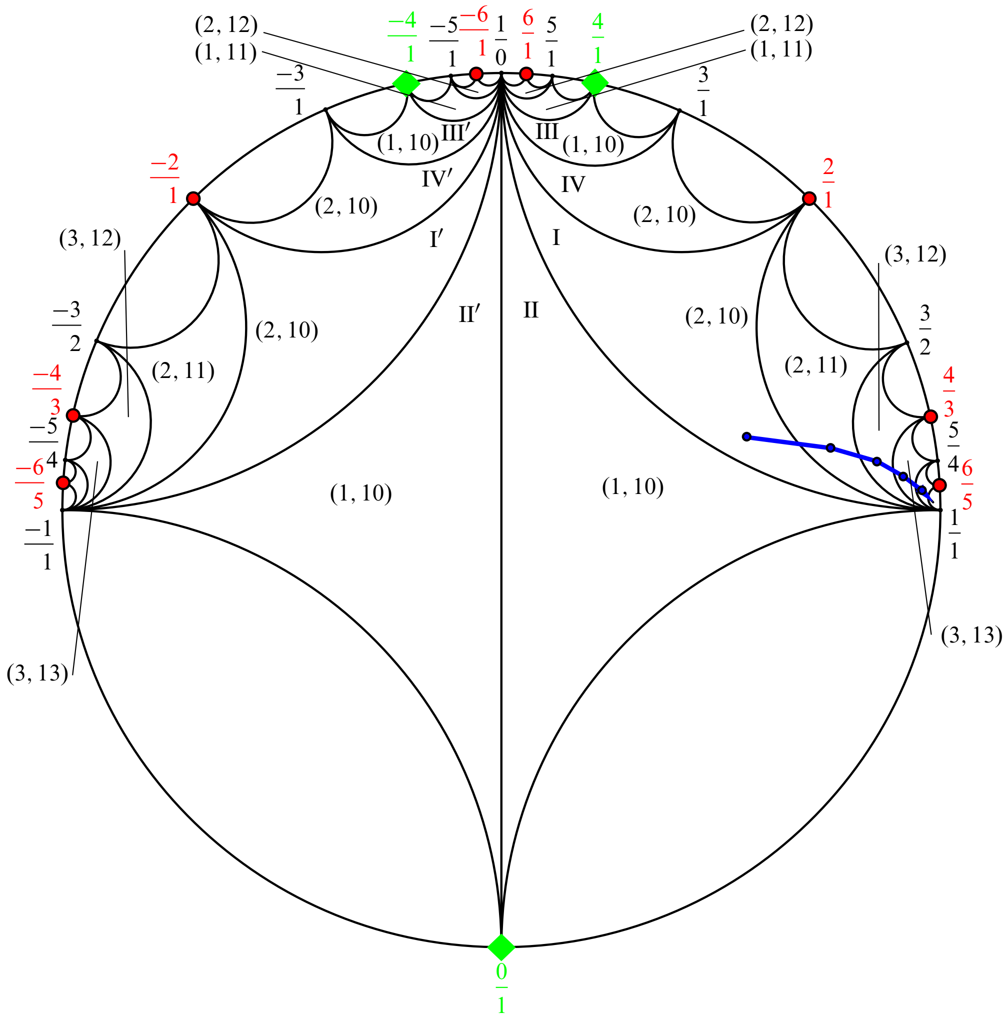


Figure 9: Slope norms and upper bounds per boundary isotopy class of triangulations for the compact core of the figure-eight knot complement. For 10-tetrahedron triangulations, the class numbers I, II, III, and IV are given. Due to the amphichirality of the figure-eight knot, reversing the orientation of the meridian (swapping the left and right of the picture) gives identical slope norms and upper bounds. Boundary slopes of incompressible and ∂-incompressible surfaces are marked in green.

At the same time, a triangulation of $M(\pm\alpha)$ obtained from $\mathscr{T}$ via layering and folding yields the upper bound: it is the size of $\mathscr{T}$ plus the length of the unique shortest path between τ and the node before the first node labelled α. Note that this upper bound depends on the choice of triangulation $\mathscr{T}$.

Figure 9 shows the first few triangles of the Farey tessellation, locating the four isotopy classes of the boundaries of the 82 triangulations. In the following, we use this figure to conveniently obtain lower and upper bounds for infinite families of Dehn fillings of M: Every ideal triangle in Figure 9 is decorated with a pair of numbers (a, b). The first one, a, denotes the slope norm of the even slope at this ideal triangle, and the second one, b, denotes the minimum size of a (known) triangulation with this isotopy type in the boundary. The latter is obtained from layering, starting from the closest triangulation along the unique shortest path in the dual of the Farey tessellation.

As an example, we can fold over the even edge of a triangulation of class I (see Figure 9) to obtain a 10-tetrahedron triangulation of $M(0/1) = \mathbb{T} \times I/\left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. We refer to the base triangle of class I with vertices $(1/0, 1/1, 2/1)$ as the *source triangle*, while we refer to the triangle with vertices $(0/1, 1/1, 1/0)$ as the *target triangle*. The lower bound in complexity for $M(0/1)$ from Corollary 3 is 2 (it is twice the first parameter in the target triangle) and the upper bound is 10, while its actual complexity is 7.

Similarly, we can fold over the even boundary edge of class IV to obtain a 10-tetrahedron triangulation of

$$M(4/1) = \mathrm{SFS}[D : (2,1), (2,1)] \cup_{\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)} \mathrm{SFS}[D : (2,1), (3,1)].$$

Again, the lower bound in complexity for $M(4/1)$ from Corollary 3 is 2 and the upper bound is 10, while its actual complexity is 7.

We now consider class III with boundary isotopy class $(1/0, 3/1, 4/1)$ and layer on its boundary along the path

$$\tau(4/1, 3/1, 1/0) \quad \tau(4/1, 5/1, 1/0) \quad \tau(6/1, 5/1, 1/0) \quad \tau(6/1, 7/1, 1/0) \\ \cdots \tau((2k-2)/1, (2k-3)/1, 1/0) \quad \tau((2k-2)/1, (2k-1)/1, 1/0).$$

Folding over the even boundary edge of slope $(2k-2)/1$ of the resulting triangulation yields $M(2k/1)$. This results in a lower bound in complexity from Corollary 3 of $2k-2$ (note that $M(2k/1)$ for $k \geq 3$ is hyperbolic and hence not a lens space), and an upper bound from the triangulation of $2k+5$ for $k \geq 3$. Experimentally, the actual complexity seems to be $2k+4$ (proven for $k = 3$).

Similarly, we can do this for all infinite paths in the dual of the Farey tessellation that have a new even slope in every second step. Some of these infinite paths are straightforward: For an ideal triangle with even slope label α, pick the odd slope β on the outside (eg $\alpha = 2/1$ and $\beta = 1/1$ in Figure 9). Walk along the infinite path of ideal triangles containing β (the blue path in Figure 9). This way all even slopes of type $\alpha_k = \alpha \oplus 2k\beta$ are encountered, where $\oplus$ denotes Farey addition (in our example, $\alpha_k = (2k+2)/(2k+1)$). For any infinite family of slopes obtained this way, the gap between upper and lower bounds in complexity for $M(\alpha_k)$ can be directly computed from the labels of the starting ideal triangle τ:

Let (a, b) be the label of τ as given in Figure 9. Provided that $M(\alpha_k)$ is not a balanced lens space, the lower and upper bounds in complexity for $M(\alpha_k)$, with $k > 0$, are then given as

$$2(k+a) \leq c(M(\alpha_k)) \leq 2k + b - 1.$$

(See above for details, and note that these bounds are not valid for the starting point $k = 0$ itself.) The gap in complexity for $M(\alpha_k)$, with $k > 0$, is hence $(2k+b-1) - 2(k+a) = b - 2a - 1$.

Remark 10 In this example we only consider infinite families of Dehn fillings of M with a constant gap in complexity. More broadly, we can use the same description and method to produce upper and lower bounds in complexity for families of arbitrary even fillings. The only difference is that the gap is potentially widening. This is due to the lower bound only taking into account new even slope labels on the path of fillings, while the upper bound grows with every step.

5.2 Even Dehn fillings of the pretzel $P(-2, 3, 7)$

In this example we compute lower and upper complexity bounds for even Dehn fillings of the pretzel knot with parameters -2, 3, and 7. Here we only work with information that is known for large collections of knots. In particular, everything we do in this example can be done for many knots in the SnapPy [6] census.

For most of our calculations, software can be used out-of-the box [1; 6]. Some calculations require small scripts or moderate levels of human interaction. For instance, determining the knot-theoretic framing is done using data on exceptional fillings for census knots [7], as well as Regina's capabilities to recognise Seifert fibred spaces [1].

Let M be the complement of the pretzel knot $P(-2, 3, 7)$ (eg the underlying space of triangulation m016 in the SnapPy [6] census). We show how to establish the following bounds of gaps between 6 and 8:

$$k+2 \le c(M(-2(k-1)/(2k-1))) \le 2k+8 \quad \text{for } k \ge 1, \tag{5-1}$$

$$2k \le c(M(-2k/1)) \le 2k+7 \quad \text{for } k \ge 1, \tag{5-2}$$

$$2k+2 \le c(M(-(6k+2)/(2k+1))) \le 2k+10 \quad \text{for } k \ge 1. \tag{5-3}$$

Using the same search through the Pachner graph of ideal triangulations of M as in the previous example, we obtain 93 triangulations of the compact core of M, each 0-efficient, with ten tetrahedra and a single vertex contained in their 2-triangle boundaries.

Looking at the boundary patterns of the Seifert surface, these 93 triangulations split into three classes, as indicated in Figure 10. Since each triangulation $\mathscr{T}$ is 0-efficient, it follows from [17, Theorem 5] that the boundary pattern of the Seifert surface is determined as the boundary pattern of a fundamental orientable normal surface of $\mathscr{T}$ with boundary a single essential curve (and such a surface always exists). Even more, there must be such a surface with maximum Euler characteristic (realising the genus of the knot).

Let $\mathscr{T}$ be the unique triangulation of the pretzel knot exterior with boundary pattern of the Seifert surface $(1, 19, 20)$. This has Regina [1] isomorphism signature `kLvKwIPQcfeghijijjllmgwneflp`.

We first need to determine the knot-theoretic framing. Observe that folding over the even boundary edge of $\mathscr{T}$ yields the lens space $L(18, 5) = M(0/1)$. Moreover, layering over the even boundary edge and then folding back over the resulting degree-1 even boundary edge yields a graph manifold homeomorphic

name	# triangulations $\mathscr{T}$	∂-pattern of Seifert surface
class 1	29	$(1, 17, 18)$
class 2	63	$(1, 19, 18)$
class 3	1	$(1, 19, 20)$

Figure 10: The 93 triangulations of the pretzel knot exterior.

with $M(-2/1)$. Hence the even boundary edge of $\mathscr{T}$ has slope $-2/1$, and the ideal triangle of the Farey tessellation encoding the isotopy class of $\mathscr{T}_\partial$ is adjacent to a triangle with even slope label $0/1$. It follows that the boundary edges of $\mathscr{T}$ have slopes $1/0$, $-2/1$, and $-1/1$, where the order corresponds to the pattern $(1, 19, 20)$.

The triangulation $\mathscr{T}$ has 75 fundamental normal surfaces in standard coordinates. These contain the orientable Seifert surface, and nonorientable surfaces with boundary a single essential curve of eight distinct slopes. Their boundary patterns, boundary slopes, maximum Euler characteristic, and slope norm are summarised in Figure 12. Once the knot-theoretic framing is known, all of this information can be computed directly from the fundamental normal surfaces of $\mathscr{T}$ and the Farey tessellation, following the procedure to compute the slope norm from [17] and from Sections 3.1 and 3.2.

From Figure 11 and extensions into the Farey tessellation we can now directly read off lower and upper bounds for $c(M(\alpha))$, where α is any given even slope α:

(1) Layer on top of $\mathscr{T}_\partial$ along the unique shortest path in the dual of the Farey tessellation from the base triangle $(1/0, -1/1, -2/1)$ labelled $\mathscr{T}$ in Figure 11, to one layering before the target triangle. That is, one layering before the first triangle containing the target slope α as one of its ideal vertices. The result is a triangulation $\mathscr{T}'$ with number of tetrahedra ten plus number of layerings.

(2) Fold over the even boundary edge of $\mathscr{T}'$ to obtain a triangulation of $M(\alpha)$.

(3) The $\mathbb{Z}_2$-norm of the unique $\mathbb{Z}_2$-torsion class of $M(\alpha)$ is one less than the slope norm in the target triangle.

(4) The difference of twice the $\mathbb{Z}_2$-norm plus two (if our triangulation is not a balanced lens space) and the size of $\mathscr{T}'$ (one less than the upper bound recorded in the target triangle) yields the gap up to which we can determine $c(M(\alpha))$.

From the above calculations, we deduce the upper and lower bounds in complexity for infinite families of Dehn fillings of M. In particular, this gives the bounds stated in (5-1)–(5-3).

The above procedure does not work for $M(0/1)$. Here we first need to layer once to obtain a different isotopy class in the boundary, and then fold back over the even edge. In this case, a better gap can be obtained by starting with a triangulation with a different isotopy class in the boundary.

5.3 Even Dehn fillings of the trefoil knot complement

In this section we discuss a nonhyperbolic knot. More specifically, we look at three infinite families of even Dehn fillings of the trefoil knot complement. For each of them we can determine their complexity up to a gap of two.

We start with the 2-tetrahedron triangulation of the right-handed trefoil knot complement M with Regina isomorphism signature `cPcbbbadu`. A search through the Pachner graph yields two triangulations of the compact core of M with four tetrahedra. Their *Regina* isomorphism signatures are `eHLObcdddwun` and

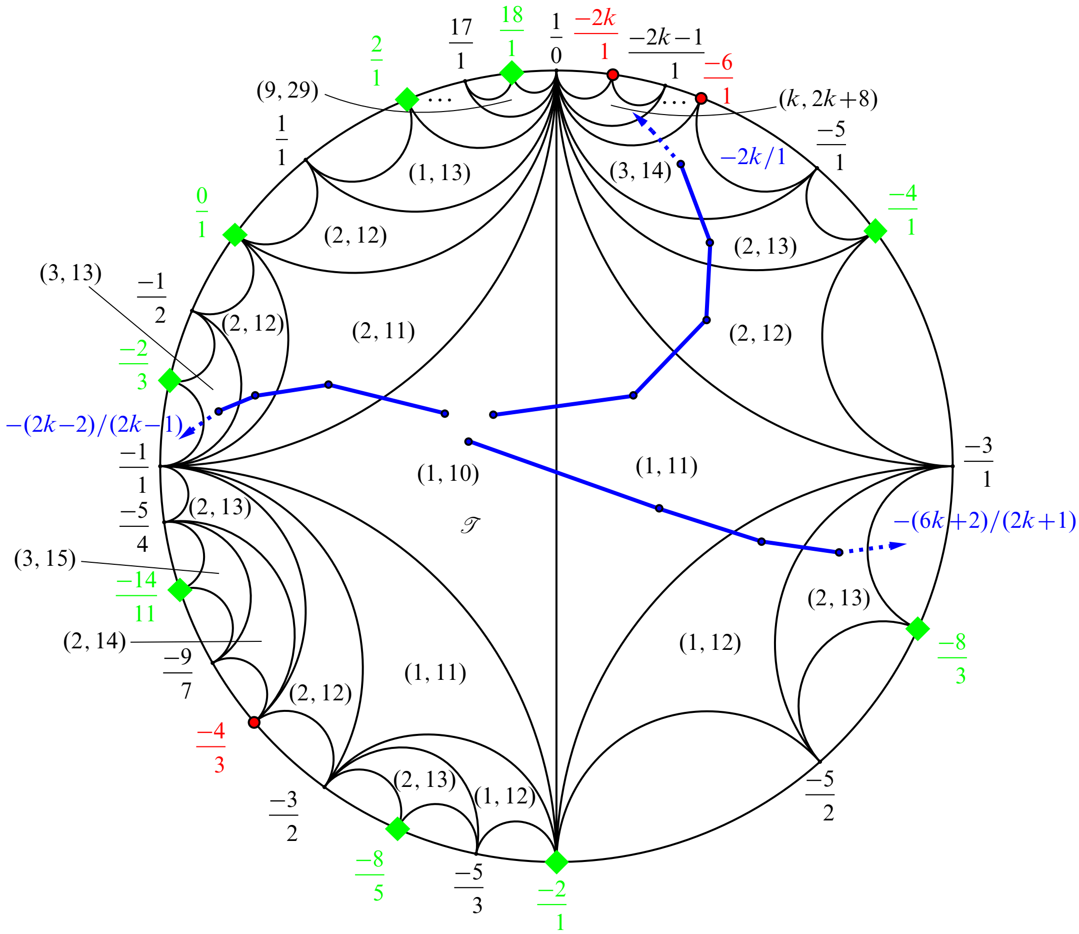


Figure 11: Slope norms and triangulation sizes for the compact core of the pretzel knot as calculated based on triangulation $\mathscr{T}$. Green diamonds indicate (even) slopes of fundamental normal surfaces of $\mathscr{T}$.

`eHLObcdddwuj`, respectively. For the remainder of this section we refer to them as $\mathscr{T}_1$ and $\mathscr{T}_2$. Both triangulations are 0-efficient.

See Figure 13 for consistent choices of framings, and boundary patterns of fundamental normal surfaces for both $\mathscr{T}_1$ and $\mathscr{T}_2$. See Figure 14 for a marked Farey tessellation containing slope norms (as computed via [17, Theorem 5]) and triangulation sizes based on layering on $\mathscr{T}_1$ and $\mathscr{T}_2$, respectively.

Starting at an ideal triangle associated to the isotopy class of the boundary of either $\mathscr{T}_1$ or $\mathscr{T}_2$, there are a total of three infinite paths through the dual graph of the Farey tessellation with a gap of $b-2a-1=2$. As in previous examples, we describe these families in terms of their filling slopes $\alpha_k = \alpha \oplus 2k\beta$. The layerings are determined by starting at one of the two base ideal triangles in Figure 14 and following the path in the dual of the Farey tessellation around the ideal vertex with label $\beta \in \{1/0, -1/1, 1/1\}$. In each step, a line of the Farey tessellation is crossed into a new ideal triangle. To obtain a triangulation

orientable	∂-pattern	χ	∂-slope α	(a,b)
no	$(1,1,0)$	-1	$-2/1$	$(1,10)$
no	$(1,1,2)$	-2	$0/1$	$(2,11)$
no	$(5,3,2)$	-6	$-8/5$	$(2,13)$
no	$(1,3,2)$	-2	$-4/1$	$(2,12)$
no	$(1,3,4)$	-1	$2/1$	$(1,13)$
no	$(3,1,4)$	-3	$-2/3$	$(3,13)$
no	$(3,5,2)$	-4	$-8/3$	$(2,13)$
no	$(11,3,8)$	-9	$-14/11$	$(3,15)$
yes	$(1,19,20)$	-9	$18/1$	$(9,29)$

Figure 12: Boundary pattern, Euler characteristic, and boundary slope of fundamental normal surfaces of triangulation $\mathscr{T}$ of the compact core of the pretzel knot. Rightmost column: tuple (a,b) of slope norm and upper bound for complexity for triangulations with even boundary edge of the given slope.

with boundary of isotopy class the class of the new ideal triangle, we layer over the boundary edge of the existing triangulation with slope the label of the opposite ideal vertex in the old ideal triangle.

- The first family is given by $\alpha_k = (-2/1) \oplus 2k(-1/1)$ for $k \geq 0$. We have for the topological type $M(\alpha_k) = \mathrm{SFS}(\mathbb{S}^2 : (2,1),(3,1),(2k+2,-1))$. The single $\mathbb{Z}_2$-torsion class of $M(\alpha_k)$ has norm k. This leads to $c(M(\alpha_k)) \geq 2k+2$, via the norm, and $c(M(\alpha_k)) \leq 2k+4$, via the layering construction.

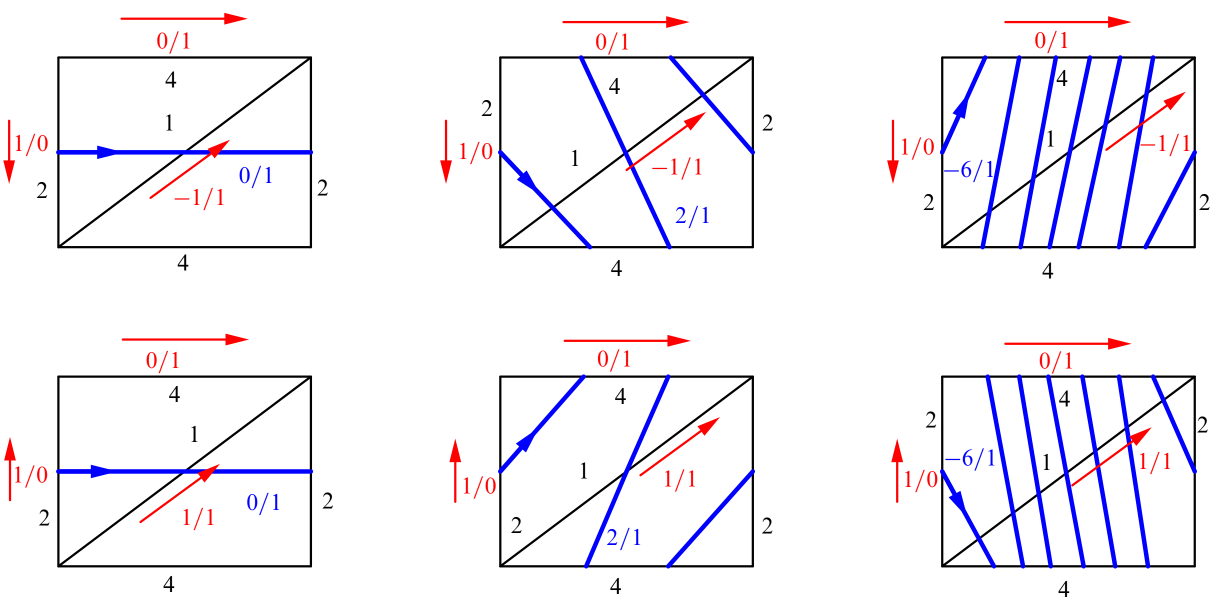


Figure 13: Boundary patterns and choices for framings for $\partial\mathscr{T}_1$ (top) and $\partial\mathscr{T}_2$ (bottom), and triangulations of the compact core of the trefoil knot. The choices for longitude and meridian are topologically equivalent for $\partial\mathscr{T}_1$ and $\partial\mathscr{T}_2$.

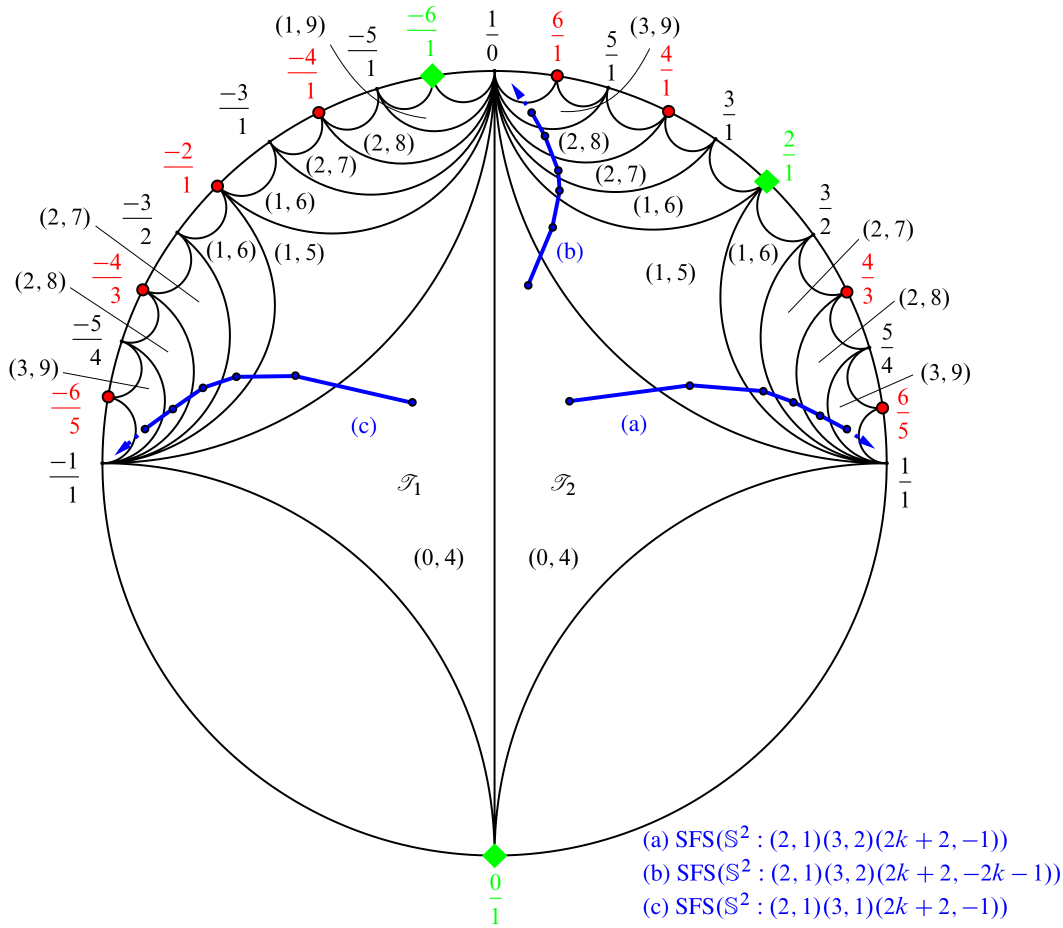


Figure 14: Slope norms and triangulation sizes giving lower and upper bounds for three infinite families of even Dehn fillings of the trefoil complement. Slopes of fundamental normal surfaces in $\mathscr{T}_1$ and $\mathscr{T}_2$ are marked in green. The slope of the Seifert surface is $-6/1$.

- The second family is given by $\alpha_k = (2/1) \oplus 2k(1/0)$ for $k \geq 0$. Here the topological type is $M(\alpha_k) = \mathrm{SFS}(\mathbb{S}^2 : (2,1), (3,2), (2k+2, -2k-1))$, the norm is again k, and we have for complexity $2k+2 \leq c(M(\alpha_k)) \leq 2k+4$.
- The third family is given by $\alpha_k = (2/1) \oplus 2k(1/1)$ for $k \geq 0$. The topological type is $M(\alpha_k) = \mathrm{SFS}(\mathbb{S}^2 : (2,1), (3,2), (2k+2, -1))$, the norm is k, and for complexity, $2k+2 \leq c(M(\alpha_k)) \leq 2k+4$.

In all three cases, the upper bound is conjectured to be the actual complexity.

The three walks in the dual of the Farey tessellation corresponding to the above families are marked in Figure 14. For the first family, we start with triangulation $\mathscr{T}_1$, while for the other two families we start with $\mathscr{T}_2$. Note that family $M(\alpha_k)$ with $\alpha_k = (-2/1) \oplus 2k(1/0)$ has a larger gap due to the Seifert surface being on this path. This reduces the $\mathbb{Z}_2$-norm and hence the lower bound in complexity for subsequent members of the associated infinite family of Dehn fillings.

References

[1] **B A Burton**, **R Budney**, **W Pettersson**, et al., *Regina*: *software for low-dimensional topology* (1999–2021) Available at `https://regina-normal.github.io/`

[2] **B A Burton**, **J H Rubinstein**, **S Tillmann**, *The Weber–Seifert dodecahedral space is non-Haken*, Trans. Amer. Math. Soc. 364 (2012) 911–932 MR Zbl

[3] **J C Cha**, *Complexity of surgery manifolds and Cheeger–Gromov invariants*, Int. Math. Res. Not. 2016 (2016) 5603–5615 MR Zbl

[4] **J C Cha**, *A topological approach to Cheeger–Gromov universal bounds for von Neumann ρ-invariants*, Comm. Pure Appl. Math. 69 (2016) 1154–1209 MR Zbl

[5] **J C Cha**, *Complexities of* 3*-manifolds from triangulations, Heegaard splittings and surgery presentations*, Q. J. Math. 69 (2018) 425–442 MR Zbl

[6] **M Culler**, **N M Dunfield**, **M Goerner**, **J R Weeks**, *SnapPy*, *a computer program for studying the geometry and topology of* 3*-manifolds* Available at `http://snappy.computop.org`

[7] **N M Dunfield**, *A census of exceptional Dehn fillings*, from "Characters in low-dimensional topology", Contemp. Math. 760, Amer. Math. Soc., Providence, RI (2020) 143–155 MR Zbl

[8] **E Fominykh**, **S Garoufalidis**, **M Goerner**, **V Tarkaev**, **A Vesnin**, *A census of tetrahedral hyperbolic manifolds*, Exp. Math. 25 (2016) 466–481 MR Zbl

[9] **R Frigerio**, **B Martelli**, **C Petronio**, *Dehn filling of cusped hyperbolic* 3*-manifolds with geodesic boundary*, J. Differential Geom. 64 (2003) 425–455 MR Zbl

[10] **J Hass**, **J C Lagarias**, **N Pippenger**, *The computational complexity of knot and link problems*, J. ACM 46 (1999) 185–211 MR Zbl

[11] **A E Hatcher**, *On the boundary curves of incompressible surfaces*, Pacific J. Math. 99 (1982) 373–377 MR Zbl

[12] **M Ishikawa**, **K Nemoto**, *Construction of spines of two-bridge link complements and upper bounds of their Matveev complexities*, Hiroshima Math. J. 46 (2016) 149–162 MR Zbl

[13] **W Jaco**, **J Johnson**, **J Spreer**, **S Tillmann**, *Bounds for the genus of a normal surface*, Geom. Topol. 20 (2016) 1625–1671 MR Zbl

[14] **W Jaco**, **J H Rubinstein**, *Inflations of ideal triangulations*, Adv. Math. 267 (2014) 176–224 MR Zbl

[15] **W Jaco**, **J H Rubinstein**, **J Spreer**, **S Tillmann**, *$\mathbb{Z}_2$-Thurston norm and complexity of* 3*-manifolds, II*, Algebr. Geom. Topol. 20 (2020) 503–529 MR Zbl

[16] **W Jaco**, **J H Rubinstein**, **J Spreer**, **S Tillmann**, *On minimal ideal triangulations of cusped hyperbolic* 3*-manifolds*, J. Topol. 13 (2020) 308–342 MR Zbl

[17] **W Jaco**, **J H Rubinstein**, **J Spreer**, **S Tillmann**, *Slope norm and an algorithm to compute the crosscap number*, Algebr. Geom. Topol. 24 (2024) 4307–4351 MR Zbl

[18] **W Jaco**, **J H Rubinstein**, **S Tillmann**, *Minimal triangulations for an infinite family of lens spaces*, J. Topol. 2 (2009) 157–180 MR Zbl

[19] **W Jaco**, **J H Rubinstein**, **S Tillmann**, *$\mathbb{Z}_2$-Thurston norm and complexity of* 3*-manifolds*, Math. Ann. 356 (2013) 1–22 MR Zbl

[20] **W Jaco**, **E Sedgwick**, *Decision problems in the space of Dehn fillings*, Topology 42 (2003) 845–906 MR Zbl

[21] **M Lackenby**, **J S Purcell**, *The triangulation complexity of fibred* 3*-manifolds*, Geom. Topol. 28 (2024) 1727–1828 MR Zbl

[22] **W B R Lickorish**, *A representation of orientable combinatorial* 3*-manifolds*, Ann. of Math. 76 (1962) 531–540 MR Zbl

[23] **S V Matveev**, *Complexity theory of three-dimensional manifolds*, Acta Appl. Math. 19 (1990) 101–130 MR Zbl

[24] **S V Matveev**, **E L Pervova**, *Lower bounds for the complexity of three-dimensional manifolds*, Dokl. Akad. Nauk 378 (2001) 151–152 MR Zbl In Russian.

[25] **S Matveev**, **C Petronio**, **A Vesnin**, *Two-sided asymptotic bounds for the complexity of some closed hyperbolic three-manifolds*, J. Aust. Math. Soc. 86 (2009) 205–219 MR Zbl

[26] **S V Matveev**, **V V Tarkaev**, *Recognition and tabulation of* 3*-manifolds up to complexity* 13, Chebyshevskiĭ Sb. 21 (2020) 290–300 MR Zbl

[27] **E Pervova**, **C Petronio**, *Complexity and T-invariant of abelian and Milnor groups, and complexity of* 3*-manifolds*, Math. Nachr. 281 (2008) 1182–1195 MR Zbl

[28] **C Petronio**, **A Vesnin**, *Two-sided bounds for the complexity of cyclic branched coverings of two-bridge links*, Osaka J. Math. 46 (2009) 1077–1095 MR Zbl

[29] **J H Rubinstein**, **J Spreer**, **S Tillmann**, *A new family of minimal ideal triangulations of cusped hyperbolic* 3*-manifolds*, from "2021–2022 MATRIX annals", MATRIX Book Ser. 5, Springer (2024) 5–28 Zbl

[30] **W P Thurston**, *The geometry and topology of three-manifolds*, lecture notes, Princeton University (1979) Available at `https://url.msp.org/gt3m`

[31] **A Y Vesnin**, **V V Tarkaev**, **E A Fominykh**, *Three-dimensional hyperbolic manifolds with cusps of complexity* 10 *that have maximal volume*, Tr. Inst. Mat. Mekh. 20 (2014) 74–87 MR In Russian; translated in Proc. Steklov Inst. Math. 289 (2015) S227–S239

[32] **A H Wallace**, *Modifications and cobounding manifolds, IV*, J. Math. Mech. 12 (1963) 445–484 MR Zbl

[33] **J Weeks**, *Computation of hyperbolic structures in knot theory*, from "Handbook of knot theory", Elsevier, Amsterdam (2005) 461–480 MR Zbl

WJ: *Department of Mathematics, Oklahoma State University*
Stillwater, OK, United States

JHR: *School of Mathematics and Statistics, The University of Melbourne*
Melbourne VIC, Australia

JS, ST: *School of Mathematics and Statistics, The University of Sydney*
Sydney NSW, Australia

jaco@math.okstate.edu, joachim@unimelb.edu.au, jonathan.spreer@sydney.edu.au,
stephan.tillmann@sydney.edu.au

Received: 28 April 2023 Revised: 6 August 2023

Algebraic & Geometric Topology **25**:1 (2025) 329–344
DOI: 10.2140/agt.2025.25.329
Published: 5 March 2025

The enumeration and classification of prime 20-crossing knots

MORWEN B THISTLETHWAITE

An account is given of the compilation of the 1 847 319 428 prime knots with 20 crossings.

57K10; 57-04

1 Introduction

In the summer of 2018 the author tabulated the knots of 20 crossings. An independent tabulation was made simultaneously by B Burton (personal communication, 2018) using the software *Regina* developed by him and others [3]; as the results of the two tabulations agree, there is some confidence that the results are correct, despite the quantity and complexity of the data.

The knots are listed up to unoriented equivalence, that is to say we regard knot pairs (S^3, K) and (S^3, L) as *equivalent* if there is a homeomorphism of pairs sending (S^3, K) to (S^3, L), and we list one representative of each equivalence class. The issue of determining which knots are amphicheiral or reversible will be addressed as a separate project.

A short historical note: knot tabulations began in earnest in the late nineteenth century with the work of P G Tait [28; 29; 30], T P Kirkman [15] and C N Little [17], Tait having being motivated by the (then current) Kelvin theory of vortex atoms. Initially, as Tait was aware, techniques were not available for distinguishing knot types rigorously; these techniques arrived shortly afterwards with the advent of the fundamental group due to Poincaré [21], whereupon M Dehn [7] and O Schreier [25] initiated the rigorous classification of knots, beginning with torus knots. A fuller account of the history, up to the classification of 16-crossing knots, is given by Hoste, Thistlewaite and Weeks [14], and, to complete the picture, Burton [2] pioneered the classification of knots of 17, 18 and 19 crossings.

A table listing the numbers of prime knots from 3 up to 20 crossings is given in the appendix.

Theorem 1.1 *The number of equivalence classes of prime knots that can be projected with* 20 *crossings, but not with fewer crossings, is* 1 847 319 428*. Of these, all but* 921 *are hyperbolic, the remainder comprising* 915 *satellites of the trefoil knot,* 5 *satellites of the figure-eight knot and the* $(3, 10)$*-torus knot.*

The issue of primality is one that is easy to overlook, but it is important, as one has to guard against "imposter" knots that might be composite in some hidden way and are thus masquerading as prime knots. For this reason a section of this article is devoted to justifying the claim that all listed knots are prime.

Acknowledgements The author would like to thank the referee and the editors for useful suggestions that resulted in a significantly improved exposition.

It is a pleasure to acknowledge the support of Ben Burton in this endeavour; a certain error in a line of the author's earlier code would probably have persisted, had he not been able to compare his results for 17 crossings with Ben's. Historically tables of knots have been error-prone, but the probability of errors is greatly reduced by having independent tabulations to hand.

2 The tabulation

2.1 Obtaining the raw list of knots

For the most part the method is the same as that employed in [14], albeit with some minor differences. Traditionally all tabulations of knots with given crossing-number begin with a listing of the prime alternating knots with that number of crossings. This is one of the few steps in the process that is truly algorithmic: from the solutions of the various Tait conjectures, it is known that a knot with a reduced alternating n-crossing diagram cannot be projected with fewer than n crossings. Also, an alternating knot is guaranteed to be prime if its reduced alternating diagrams have the property that they do not admit a simple closed curve in the projection plane meeting the knot projection transversely in two points on distinct edges of the projection; see Figure 1 and [18]. Furthermore, any two reduced alternating diagrams represent the same link type if and only if one can transform one to the other by means of a finite sequence of *flypes*; see Figure 2.

It is precisely the failure of nonalternating knots to adhere to such desirable properties that renders their classification a challenge.

It is relatively straightforward to write a program that generates all possible reduced alternating diagrams of a given crossing-number n, choosing a representative from each flype equivalence class, although skill is required in devising a program that will run in a reasonable time. This has indeed been accomplished quite dramatically for $n \leq 23$ [22]. The present author has written a program that generates all prime

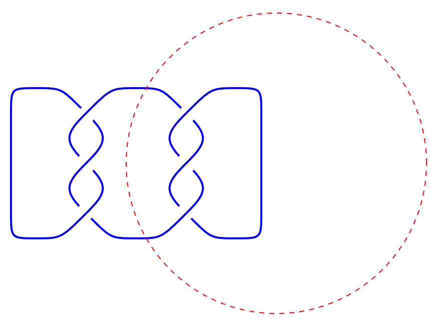


Figure 1: A composite alternating knot.

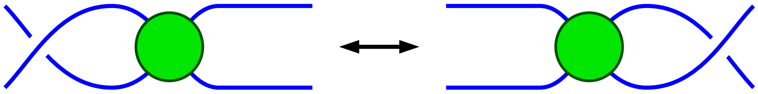


Figure 2: The "flype" transformation.

alternating links with a given number of crossings, and for the case $n = 20$ the number of prime alternating knot types turns out to be 199 631 989.

Once one has the list of alternating diagrams to hand, nonalternating diagrams can be obtained from them by means of crossing switches. It is only necessary to take one alternating diagram from each flype equivalence class, as if alternating diagrams D_1 and D_2 are flype-equivalent, then the diagrams obtained from crossing switches of D_1 are flype-equivalent to those obtained from D_2. Switching all crossings of a knot diagram produces a knot equivalent to the original on account of being its reflected image in the projection plane, so a crude estimate of the number of diagrams to be generated in this way from a single alternating diagram is $2^{19} = 524\,288$. However, it is only on rare occasions that this number is needed, as can be seen from the following observation: if a rational tangle diagram [4] is not alternating, then there exists an isotopy of the tangle that reduces the number of crossings while keeping the four ends of the tangle fixed; see Figure 3. Therefore built into the program is a procedure that detects all nontrivial rational tangle substituents, and then we only allow crossing switches of the "base" alternating knot diagram that keep each of these tangles alternating.

The resulting nonalternating diagrams are subjected to a number of rapid viability tests to check whether the number of crossings can be reduced, and are immediately discarded upon failing any such test. A surviving diagram is then subjected to a different kind of test, specifically to see whether it can be transformed by flypes and passes (Figures 2 and 4) to a diagram whose DT code [8] is lexicographically less. As the size of an equivalence class generated by these moves can be very large, even in the tens of thousands, we declare that the diagram passes the test if it is still lexicographically minimal once some fixed number k of diagrams has been generated by the moves. Smaller values of k will entail larger redundancy, but it makes sense to keep k quite small on account of the time that would be spent on processing a large set of diagrams.

In [14] some diagram moves more "exotic" than flypes and passes were used, but this approach was avoided here as it was deemed unnecessary, quite apart from the increased danger of introducing bugs

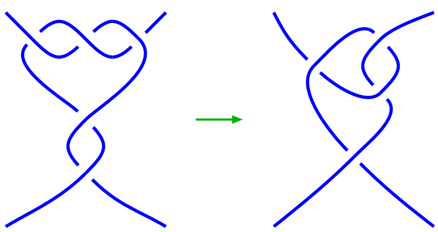


Figure 3: Reducing a nonalternating diagram of a rational tangle.

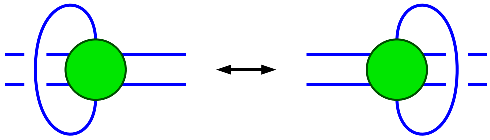


Figure 4: The "pass" transformation on diagrams.

into the program. In practice the value chosen for k was 200, and from experimentation with known tabulations with fewer crossings it was estimated that this resulted in roughly 25% redundancy overall.

2.2 Removing duplicate knots

The next task is to augment the current list of over 2 billion nonalternating diagrams by all tabulated knots with fewer crossings, and then to remove as many duplicates as possible. The extreme difficulty of achieving this "house cleaning" simply by inspecting or manipulating knot diagrams is wonderfully exhibited by the celebrated *Perko pair* [19], a pair of knots with only 10 crossings declared to be inequivalent in Little's 1900 table, this status persisting until 1974 when K Perko finally spotted the equivalence, thereby obtaining the first correct table of 10-crossing knots.

The 1970s also saw R Riley's [23] discovery of a hyperbolic structure on the complement of the figure-eight knot, this being one of the inspirations for W Thurston's breakthrough work on geometric structures on 3-manifolds. This in turn led to J Weeks' extensive program SnapPea and its more recent Python implementation SnapPy [6], one of whose many features is the ability to compute the *canonical cell decomposition* [9; 24] of a hyperbolic 3-manifold with genus-1 cusps.

The preimage of this cell structure in the universal cover can either be seen in the upper half-space model as dual to the Ford domain, or it can be seen by means of a convex hull construction in the Minkowski model. SnapPea performs a very rapid computation of a purported canonical cell decomposition by starting with a known ideal triangulation of the manifold and then implementing a heuristic optimization process that applies combinatorial moves on the triangulation without affecting the underlying topology. Because of inevitable accumulation of roundoff error, the resulting cell decomposition might on occasion not be the canonical one, but nonetheless if two hyperbolic knots produce isomorphic cell decompositions, their respective complements are proved to be homeomorphic, and from the fact that knots are determined by their complements [12], the knots are equivalent. In practice, even at the level of 20 crossings this is an effective way of removing duplicates, which otherwise could be very hard to spot.

Indeed, the current list of over 2 billion 20-crossing nonalternating diagrams was fed through SnapPea's canonical cell decomposition procedure, and the few hundred million diagrams producing duplicate cell decompositions were discarded. During this process approximately a mere 549 491 were declared (with due caution) by SnapPea to be "apparently not hyperbolic", and these were copied to a separate list. The next stage is to try to distinguish by means of invariants the knots in the filtrate. There is no algorithm at work here, as there does not seem to be any way of predicting which invariants will distinguish which knots: we just throw invariants at the knots and hope for the best. However, the method is rigorous, as all computations of invariants are integer based.

2.3 Application of invariants to the list of knots

2.3.1 Description of the invariants The first invariant applied to the remaining diagrams was the Jones polynomial, for which we are fortunate in having the very fast program of [10], which on a single processor of my workstation will process a million 20-crossing knots in $2\frac{1}{2}$ minutes. This partitions the set of diagrams into relatively small equivalence classes, such that within each equivalence class all knots have the same Jones polynomial. The knots in equivalence classes of size 1 are extracted and placed in the "resolved" folder, and the remaining knots are subjected to a sequence of further invariants, the aim being to subject the partition to successive refinements so that eventually all equivalence classes have size 1.

The remaining invariants are classical, and occur in papers of R H Fox [11] and Perko [20]. They rely on the fact that knot groups seem always to have an abundance of subgroups of small index. It follows from the work of Thurston that knot groups are residually finite, but this alone does not explain why the knots in our 20-crossing list are so rich in subgroups of index less than 10. Given a subgroup H of index n of a knot group G, the group G acts transitively by left multiplication on the set of n left cosets of H, giving rise to a transitive permutation representation of degree n of G. Conversely, given a transitive permutation representation of degree n of G, the stabilizers of the n symbols are conjugate subgroups of G of index n. Using the Reidemeister–Schreier rewriting process we can obtain a presentation of such a subgroup H, and by abelianization obtain a finitely generated abelian group, which is essentially the first homology group of the covering space of the knot complement corresponding to H. We can also glue in solid tori to this covering space so that the components of the preimage of a meridian curve are spanned by cross-sectional disks, thus obtaining the first homology group of the so-called *branched covering space*.

The technique is to choose a transitive permutation representation of some group, for example the natural representation of degree 5 of the alternating group A_5, and then find all homomorphisms of the knot group onto that group of permutations, up to composition with inner automorphisms of the image group. The multiset of abelian groups thus obtained is then an invariant of that knot type, and amazingly, together with the Jones polynomial, it was possible in this way to distinguish almost all listed 20-crossing knots from one another and from knots with fewer crossings, using only subgroups of the symmetric group S_7.

Figure 5 contains two examples of this type of invariant applied to a fairly resistant mutant pair of 20-crossing knots.

Each diagram in Figure 5 consists of an upper tangle glued to a lower tangle along four strands; the second diagram can be obtained from the first by excising the upper tangle, rotating it through a half turn in the projection plane and then gluing it back to the lower tangle.

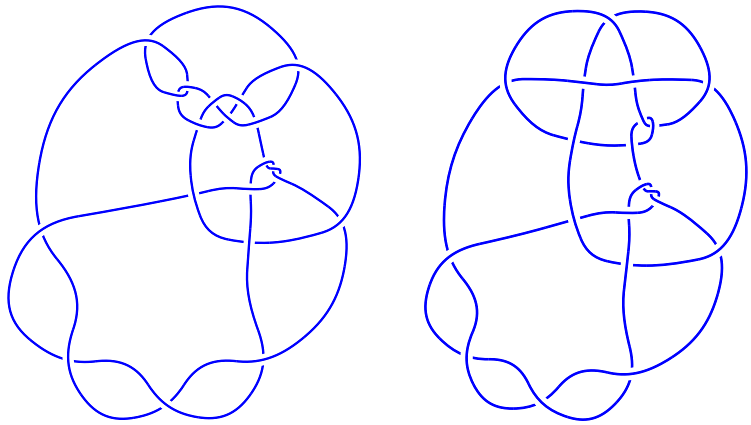


Figure 5: A pair K_1 and K_2 of mutant 20-crossing knots.

Being related by mutancy, these knots cannot be distinguished by the Jones polynomial nor indeed by the HOMFLYPT 2-variable polynomial; also they resisted homology groups associated with permutation representations of degrees 5 and 6. However they did succumb to permutation representations mapping meridians to one of the two conjugacy classes of 7-cycles in the alternating group A_7. For K_1 there were 14 representations, producing homology groups with torsion numbers as follows:

$$[1083964, 14],\ [10873394],\ [117987912],\ [1308356, 2, 2],\ [13423592, 8],\ [155682849, 3],\ [2496669],$$
$$[30245222, 2],\ [353577, 7],\ [477902327],\ [4832310],\ [58290239, 7],\ [8694588],\ [909657, 7].$$

And for K_2 just 13 representations:

$$[10007522,\ 2],\ [1339604,\ 14],\ [20281751],\ [21298634],\ [24072097],\ [2742502,\ 2],\ [304197488,\ 2],$$
$$[40220460],\ [46137,\ 21],\ [4719806],\ [53620280],\ [56118930,\ 3],\ [6282066,\ 2].$$

Since the group A_7 admits an automorphism sending each 7-cycle to its inverse, this conjugacy class cannot be used to detect nonreversibility of knots. It was observed by H Trotter [32] that a more careful choice of target group can be effective for this purpose; indeed this was the first occasion that the existence of nonreversible knots was proved. Later R Hartley [13] used (solvable) groups of functions $x \mapsto ax + b$ (for $a \neq 0$) over finite fields to establish nonreversibility of many knots of up to 10 crossings. For the knots K_1 and K_2 of Figure 5, the sporadic Mathieu simple group M_{11} is effective in showing that they are not reversible. Specifically, we can use the irreducible permutation representation of M_{11} of degree 11, and map meridians to one of the two conjugacy classes of size 990 containing elements of cycle type $(ab)(cdefghij)$. Here are the results, with torsion numbers for each representation enclosed in square brackets as above:

K_1	$[1394030, 2]$
Reverse of K_1	$[287520],\ [65322]$
K_2	$[14118592], [5682, 2]$
Reverse of K_2	$[1598572],\ [4161904]$

It is expected that one can determine reversibility in this way for the list of 20-crossing knots, although it could be very time consuming. Determining amphicheirality is in practice easier, as almost all instances of nonamphicheirality are detected by the Jones polynomial.

2.3.2 Details of the invariants' performance There now follows details of the efficacy of the invariants used for distinguishing the 20-crossing knots from one another and from knots with fewer crossings. The enumeration stage of the classification process generated 2 229 828 372 20-crossing nonalternating diagrams; barring programming error there were no omissions in this list, and the procedure was in three main stages as described below.

Stage 1 The raw list of 20-crossing nonalternating diagrams was augmented by the list of all 352 151 858 hyperbolic knots with fewer than 20 crossings, resulting in an enlarged list containing 2 581 980 230 knots. SnapPea's canonical cell decomposition procedure was applied to each knot in the enlarged list, and the data was sorted so that duplicate cell decompositions became evident. Each cell decomposition was encoded by a string of approximately 300 bytes on average, so the amount of data involved in this step was around 775 GB. As explained in Section 2.2, it is not guaranteed that the cell decompositions output by this procedure are canonical in the sense of [9; 24], but knots with duplicate cell decompositions have homeomorphic complements, so are equivalent owing to the fact that knots are determined by their complements [12].

The canonical cell decomposition procedure declared that 549 491 knots from the list were "apparently not hyperbolic" and these were put in a separate list for further treatment. Aggressive diagram moves revealed that out of these knots 200 were the unknot, 547 611 were composite knots, and a further 482 could be drawn with fewer than 20 crossings. This left a residue of 1198 knots, which on being treated to still more stringent diagram moves were shown to belong to 921 knot types, distinct from one another and distinct from all nonhyperbolic knot types with fewer than 20 crossings. The proof that this list consisted of a single 20-crossing torus knot and 920 20-crossing satellite knots is given in Section 3.

After removing the 549 491 knots declared to be "apparently not hyperbolic" and the knots whose complements had duplicate cell decompositions, the number of knots in the refined list was 1 999 847 149. These were input into the next stage, it being expected that the only duplications were those arising from roundoff error in application of the canonical cell decomposition procedure.

Stage 2 From an accounting point of view this was the easiest stage. The Jones polynomials of the 1 999 847 149 knots output by the previous stage were computed, and the 336 548 774 knots with unique polynomials were extracted and placed in the store of "resolved" knots. The remaining 1 663 298 375 knots were input into Stage 3, which subjected them to the invariants described above, namely first homology groups of branched covering spaces corresponding to permutation representations of the knot groups.

Stage 3 Tables 1 and 2 summarize the results of this stage. The first table uses representations into alternating or symmetric groups of degrees 5 and 6, and the number of unresolved knots was reduced from

degree	cycle type	input	unique	nonunique	runtime
5	$(abcde)$	1 663 298 375	906 980 266	756 318 109	20 days
5	$(abc)(de)$	756 318 109	317 431 388	438 886 721	4 days
5	$(abcd)$	438 886 721	309 112 549	129 774 172	5 days
6	$(abc)(def)$	129 774 172	66 784 736	62 989 436	6 days
6	$(abcde)$	62 989 436	57 189 475	5 799 961	6 days
6	$(abcdef)$	5 799 961	5 071 212	728 749	1 day

Table 1

1 663 298 375 to 728 749. The column labelled "unique" gives the number of knots distinguished from all others and placed into the "resolved" store, and the column labelled "nonunique" gives the number of unresolved knots requiring further treatment. The column labelled "cycle type" gives the cycle type of the conjugacy class to which meridians of the knot group were mapped. The machine used for these computations had 160 GB of memory and 20 processing cores.

The remaining knots were then subjected to permutation representations in various specific groups, as set out in Table 2. Each of these substages took less than a day of runtime.

At this point the list of 17 528 unresolved knots were partitioned into 8755 equivalence classes, where knots within each equivalence class had resisted all invariants applied to date. It was suspected that each of these in fact represented a single knot type, and this was confirmed by a more persistent application of the canonical cell decomposition procedure: SnapPea has a convenient "random retriangulation" feature, and from this a small number of different contenders for canonicity were obtained, amongst which matching cell decompositions were found in each of the outstanding cases.

Surprisingly the last two knots to be distinguished, in the last row of Table 2, were a pair of 14-crossing two-bridged knots, with associated fractions $\frac{505}{192}$ and $\frac{505}{212}$ and respective Conway codes 2111221112 and 2211111122. These are easily distinguished by the fact that they are alternating, and also by their lens space two-fold branched covers, but for some reason they resisted polynomial invariants and the homology invariants of Tables 1 and 2 until the very last step.

group	degree	cycle type	input	unique	nonunique
$\mathrm{PSL}(2,7)$	7	7-cycles	728 749	572 093	156 656
$\mathrm{PSL}(2,11)$	11	11-cycles	156 656	117 446	39 210
$\mathrm{PSL}(2,13)$	14	13-cycles	39 210	15 364	23 846
$\mathrm{PSL}(2,17)$	18	17-cycles	23 846	5 245	18 601
A_7	7	7-cycles	18 601	1 071	17 530
$\mathrm{PSL}(2,19)$	20	19-cycles	17 530	2	17 528

Table 2

This concluded the task of obtaining a list of 20-crossing knots with no omissions or duplications, but it was still necessary to check that there were no "poseur" composite knots in the list. One almost hoped that some would materialize, as such examples would be noteworthy.

3 Establishing primality

A fundamental property of a hyperbolic 3-manifold is that it cannot contain an essential torus. The software *Regina* [3] confirmed that all presumed 1 847 318 507 hyperbolic 20-crossing knots in our list are indeed hyperbolic (ie there are no false positives with respect to hyperbolicity) so all are immediately known to be prime.

A different approach is needed for showing that the 921 (apparently) nonhyperbolic knots are prime. The single torus knot was easily identified, and since torus knots are prime we may restrict our attention to the remaining 920 knots in this list.

Regina [3] was also put to work by Burton to implement normal surface calculations for dealing with these 920 knots, and in this way all but eight were confirmed to be prime. *Regina* was not immediately able to certify primality of the remaining eight knots (personal communication, 2018) and this led the author to an alternative approach based on tangle decompositions, anticipating that both methods could be useful in future tabulations. Indeed the methods could be complementary, for the following reason. All prime satellite knots of up to 20 crossings have minimal diagrams that are tangle sums satisfying the hypothesis of Theorem 3.2, and although this situation is expected to continue for a while, it will not continue indefinitely. At some point a more generic technique for establishing primality will be essential, and this could be provided by programs such as *Regina*.

Figure 6 illustrates one of the eight abovementioned satellite knots. It is visibly a tangle sum of a *companion tangle* on the right (Figure 7, right) with an alternating tangle on the left, and in fact each of these eight satellites admits such a decomposition. From the main result of [18] the alternating tangle summands of these knots have no local knotting (Figure 7, left), so from Theorem 3.2 these eight knots are all prime.

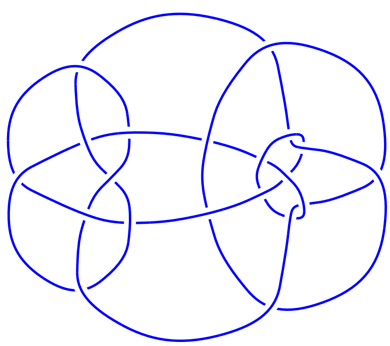


Figure 6: A 20-crossing satellite of the trefoil.

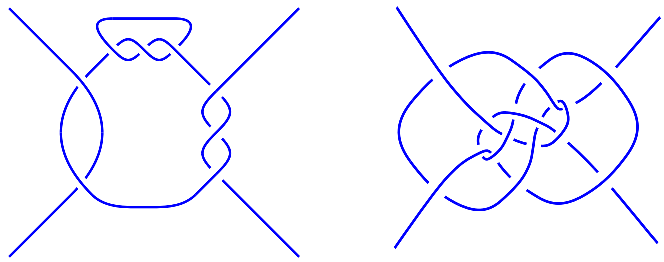


Figure 7: A locally knotted tangle (left) and a companion tangle (right).

We recall the terminology of [16]. In that paper a *tangle* is defined to be a pair (B, T) where B is a 3-ball, ie a manifold with boundary homeomorphic to the standard 3-ball B^3, and T is a proper 1-submanifold of B consisting of two disjoint arcs (naturally we assume that we are in the piecewise linear or smooth category). Thus the boundary of T consists of four points on ∂B. This definition has some obvious generalizations; for example we might allow the number of arcs in T to be greater than 2, but the definition as given is sufficient for our purposes. Tangles (B_1, T_1) and (B_2, T_2) are *equivalent* if there is a homeomorphism of pairs from (B_1, T_1) to (B_2, T_2), and (B, T) is *untangled* or *trivial* if it is equivalent to a product $(D, \{x, y\}) \times I$, where D is a disk and x and y are points in its interior.

In the definition of tangle equivalence given above, for a homeomorphism $h\colon (B, T_1) \to (B, T_2)$ there is no restriction on the effect of h on the boundary 2-sphere of B, other than the requirement that it map ∂T_1 to ∂T_2. For example, any tangle represented as a diagram of a *rational tangle* [4] is equivalent to a tangle where T consists of two parallel line segments, ie it is trivial.

A tangle (B, T) is *locally unknotted* if each 2-sphere in B meeting T transversely in two points bounds in B a ball meeting T in an unknotted spanning arc. Otherwise we say that (B, T) is *locally knotted*; an example is illustrated in Figure 7, left. Observe that if (B, T) is locally unknotted, and there exists a properly embedded disk in B separating the arcs of T, then (B, T) is trivial.

Given tangles (B_1, T_1) and (B_2, T_2), we may glue them together by means of some homeomorphism of (2-sphere, four points) pairs to obtain a link L of one or two components in the 3-sphere. Such a pair (S^3, L) is called a *sum* of the tangles (B_1, T_1) and (B_2, T_2). Given two tangles drawn in the usual way as diagrams, one way of summing them is to join the diagrams by arcs in the projection plane.

If (B, T) contains a 2-sphere S exhibiting local knottedness, with knotted arc α in the ball bounded by S, then there is a well-defined nontrivial knot K obtained by joining the ends of α with an arc in S, and this knot K will persist as a connected summand of any knot formed by summing (B, T) with an arbitrary tangle. Therefore if we can sum (B, T) with a tangle so as to obtain the unknot, (B, T) is locally unknotted. In more complicated situations we have the following effective test for local unknottedness:

Proposition 3.1 *Let (B, T) be a tangle for which there exist tangles (B_1, T_1) and (B_2, T_2), such that summing (B, T) with the (B_i, T_i) in turn produces distinct prime knots K_1 and K_2. Then (B, T) is locally unknotted.*

Proof If (B,T) is locally knotted, there exists a nontrivial knot that is a connected summand of each of the distinct prime knots K_1 and K_2, and this contradicts uniqueness of factorization of knots. □

There is a special type of tangle that pertains to satellite knots. We define a *companion tangle* to be a tangle (B,T) where T consists of two parallel, knotted arcs in B; an example is illustrated in Figure 7, right. Each companion tangle (B,T) contains a properly embedded annulus A in $B-T$ that "follows" the two strands of T in tube-like fashion, ie there is a homeomorphism $h\colon S^1\times I\to A$ such that each section $h(S^1\times\{t\})$ of A bounds a disk in B meeting T transversely in two points. A companion tangle cannot be locally knotted, as it is always possible to sum it with a trivial tangle so as to obtain the unknot; in Figure 7, right, this can be seen by taking two arcs in the projection plane, one joining the two left-hand ends and the other the two right-hand ends.

Let us now consider a knot (S^3,K) constructed as a sum of a companion tangle (B_1,T_1) with a locally unknotted tangle (B_2,T_2). We may form a torus F in the complement of K as the union of the "following" annulus A_1 of (B_1,T_1) described above, with a boundary-parallel annulus A_2 in (B_2,T_2) that "swallows" T_2. Let V be the solid torus containing K that is bounded by F; V is the union of two "halves" $V\cap B_1$ and $V\cap B_2$ glued together along cross-sectional disks D_1 and D_2, both in $\partial B_1=\partial B_2$, and each meeting K in two points. The core Γ of V is a nontrivial knot in S^3, as it is the union of a knotted arc in B_1 that is the core of A_1 with an unknotted arc in B_2 that is the core of A_2.

Any cross-sectional disk of V not meeting K would have to separate the strands of the second tangle (B_2,T_2), but the local unknottedness of (B_2,T_2) would force that tangle to be trivial, and we would be in the situation described above where K is the unknot. On the other hand, if there is no cross-sectional disk of V separating the strands of (B_2,T_2), K is a satellite of Γ and the torus F is incompressible in S^3-K.

The next theorem provides the method for showing that the 920 outstanding knots are prime. It is closely related to results in [26; 16; 5], as explained below; however the full proof is given here as the hypotheses are an exact fit to our situation, and moreover should be applicable to future tabulations with more than 20 crossings.

Theorem 3.2 *Let K be a knot that is a sum of a companion tangle with a locally unknotted tangle. If K is nontrivial, then K is prime.*

Proof We adopt the notation of the preceding discussion: (S^3,K) is a nontrivial knot that is the sum of a companion tangle (B_1,T_1) with a locally unknotted tangle (B_2,T_2), $F=A_1\cup A_2$ is the incompressible torus in S^3-K that "follows" T_1 and "swallows" T_2, and V is the solid torus with boundary F.

Let S be a 2-sphere in S^3 meeting K transversely in two points. Before proceeding further it is useful to observe that each simple closed curve C in $S-K$ is either nullhomotopic in $S-K$ (and hence also nullhomotopic in S^3-K), or else it separates the punctures. In particular, a circle on F that bounds a cross-sectional disk of V cannot lie on S.

We first consider the special case where S is contained in V. Since by hypothesis each of the constituent tangles (B_1, T_1) and (B_2, T_2) is locally unknotted, the conclusion of the theorem holds for S contained in either "half" $V \cap B_i$, and we are motivated to consider the transverse intersection of S with the two disks D_1 and D_2 along which the halves of V are glued together. We may assume that the two points of $S \cap K$ are away from the D_i. The set $S \cap (D_1 \cup D_2)$ is the union of a disjoint collection of circles on S; let C be a circle from this collection that is innermost on S, say without loss of generality $C \subset S \cap D_1$. Then C bounds a disk $\Delta_1 \subset S$ and a disk $\Delta_2 \subset D_1$. The union of the disks Δ_i is an embedded 2-sphere Σ, bounding a ball B' contained in one of the B_i. The number n of points of $\Delta_2 \cap K$ is $0, 1$ or 2, and we consider each of these cases. We can exclude the possibility $n = 2$ summarily, as in this case C would be a simple closed curve on S homotopic in $S^3 - K$ to a meridional curve of $F = \partial V$, a situation ruled out in the previous paragraph.

Suppose that $n = 0$; then Δ_1 meets K in 0 or 2 points. In the latter case, from the hypothesis of local unknottedness applied to Σ, the ball B' would meet K in an unknotted arc; we deduce from this that one of the components of $S^3 - S$ would meet K in this arc, and the conclusion of the theorem would follow. Otherwise the ball B' does not meet K. The circle C might not be innermost on D_1, but nonetheless Δ_1 can be pushed by an isotopy through B', taking with it all components of $S \cap B'$, reducing the number of components of $S \cap (D_1 \cup D_2)$. If $n = 1$, then from the hypothesis of local unknottedness B' meets K in an unknotted arc, so again there is an isotopy that pushes Δ_1 across B', including if necessary another component of $S \cap B'$ meeting K in one point. Here the isotopy will move points of K along the unknotted arc, but can be assumed to fix K setwise. We conclude that there is an isotopy of S into one of the B_i without affecting transversality of $S \cap K$, whence S bounds a ball on one side meeting K in an unknotted arc, and the conclusion of the theorem follows for this special case.

For the remainder of the proof we assume that S has nonempty transverse intersection with $F = \partial V$; the proof will be completed by showing that there an isotopy of S in S^3, maintaining transversality of S with K, that moves S to a 2-sphere contained in V.

Recall that a simple closed curve in the twice-punctured sphere $S - K$ is either nullhomotopic in $S - K$ or is homotopic to a meridian curve of K. The torus F does not contain any simple closed curve of the second type, so each component C of $S \cap F$ is a simple closed curve bounding a disk in $S - K$, and also bounding a disk in F owing to the incompressibility of F. Let us take a component C of $S \cap F$ bounding a disk $\Delta_1 \subset S - K$ whose interior does not meet F; also let Δ_2 be the disk on F bounded by C. Then $\Delta_1 \cup \Delta_2$ is an embedded 2-sphere in $S^3 - K$, and in a manner similar to that of the special case we can perform an isotopy of S that reduces the number of components of $S \cap F$. Repeating the process will eventually move S into V, and the proof of the theorem is complete. $\square$

There is overlap between Theorem 3.2 and results in the literature, most notably H Schubert's paper [26], where the notion of the companionship tree of a knot is introduced, and where it is shown that doubled knots and cabled knots are prime.

In [16] a tangle is called *prime* if it is nontrivial and locally unknotted, and it is proved in that paper that a sum of two prime tangles is a prime link. A companion tangle is prime according to this definition, as it is locally unknotted, and cannot be trivial as its individual strands are knotted arcs. Theorem 3.2 shows that, apart from the obvious single exception, a knot formed as a sum of a companion tangle with a trivial tangle is also prime, thus confirming a special case of the conjecture stated in [16, Section 4].

Theorem 4.4.1 of [5] deals more generally with primality of satellite knots; it includes the hypothesis that the pair (V, K) is locally unknotted, this being the conclusion of the special case dealt with in the proof of Theorem 3.2.

Recall that there are 920 knots in our tabulation that are under examination for primeness. It was suspected that five of these are satellites of the figure-eight knot, and these were easily found in the list; they are all obtained by summing the companion tangle of Figure 7, right, with a 4-crossing rational tangle. Since they are already known to be nontrivial, and rational tangles are certainly locally unknotted, application of Theorem 3.2 shows that they are prime.

Naturally one suspects that the remaining 915 knots are satellites of the trefoil knot. In order to apply Theorem 3.2 we need diagrams of these knots that show each as a sum of a companion tangle with a locally unknotted tangle. Undoubtedly it would be possible to find such diagrams directly; however, a different approach was used here. The list of 199 631 989 prime *alternating* 20-crossing knots provides, up to flype equivalence, all projections of prime nonalternating knots, and an easy search through this list found 434 projections of tangle sums of the required kind. Suitable over- and under-passes were applied to these, resulting in a refined list of 915 knot diagrams that visibly were sums of tangles (B_1, T_1) and (B_2, T_2) with (B_1, T_1) a companion tangle, and matched the tabulated 915 knots.

As the knots were already known to be nontrivial, it remained to check that in each case the tangle (B_2, T_2) was locally unknotted. In all but ten cases verification was immediate, as the diagrams of (B_2, T_2) are either alternating, in which case they are subject to [18, Theorem 1], or they are standard diagrams of arborescent tangles [1].

The ten exceptional cases come in five pairs, each pair consisting of a diagram and its reflection in the projection plane. It was only necessary to check one tangle from each pair, and they are illustrated in Figure 8.

Under mild scrutiny the individual strands of the third, fourth and fifth tangles of Figure 8 are all revealed to be unknotted, so local knots for these tangles are ruled out. One can also notice that the first two tangles are equivalent: there is a homeomorphism that interchanges the lower two tangle ends. Therefore, in order to complete the proof that all 920 satellite knots are prime, we just need to check that the first tangle is locally unknotted. This follows quickly from Proposition 3.1: summing (in the "diagrammatic" sense) with a tangle with no crossings produces a prime 8-crossing knot, and one can obtain a prime 9-crossing knot, also a prime 10-crossing knot by summing with a 2-crossing tangle.

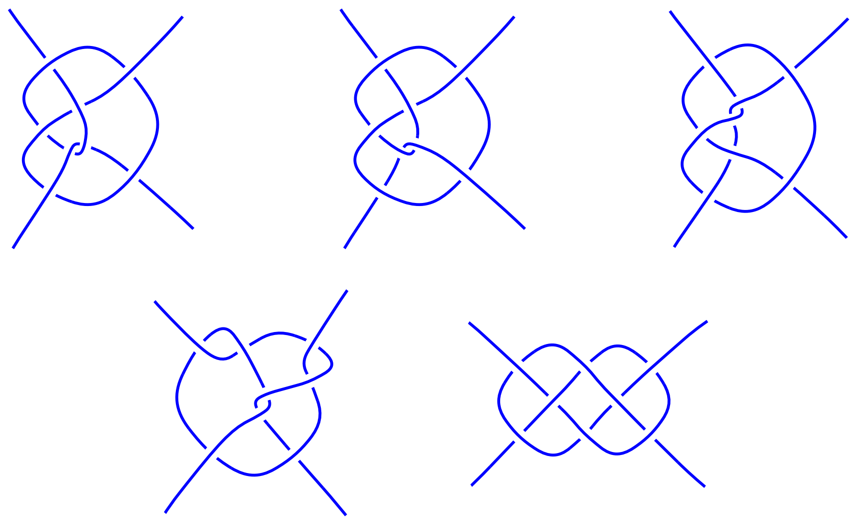


Figure 8

Appendix

A.1 The rate of growth of the number of knots

It is natural to ask whether one can estimate the number of prime knots with a given number of crossings without an actual tabulation. There are very few results in existence on this topic, but the following is known. It is stated for links rather than knots, but it suggests that the number of nonalternating knots grows exponentially faster than that of alternating knots.

Theorem A.3 (i) [27] *Let A_n denote the number of prime alternating link types with n crossings. Then*

$$\lim_{n\to\infty} A_n^{1/n} = \tfrac{1}{40}(101+\sqrt{21001}) \approx 6.1479.$$

(ii) [31] *Let λ be the limit stated in (i). There exists a set $\mathcal{B}$ of prime link types, strictly containing the set of prime alternating link types, such that if B_n is the number of link types in $\mathcal{B}$ with n crossings, then $\lim_{n\to\infty} B_n^{1/n}$ exists and is strictly greater than λ.*

A.2 The number of prime knot types with n crossings, $3 \le n \le 20$

The first correct tabulations of knots of 17, 18 and 19 crossings were produced by Burton [2].

Table 3 gives summary data up to 20 crossings. When reading Table 3, it is worth noting that the only prime alternating nonhyperbolic knots are the $(2,n)$-torus knots (with n necessarily odd) [18]. Thus for even n all n-crossing prime alternating knots are hyperbolic, and for odd n there is a single nonhyperbolic prime alternating knot, namely the $(2,n)$-torus knot. Also it follows that all prime satellite knots are nonalternating.

crossings	alternating knots	nonalternating knots	hyperbolic	torus	satellites
3	1	0	0	1	0
4	1	0	1	0	0
5	2	0	1	1	0
6	3	0	3	0	0
7	7	0	6	1	0
8	18	3	20	1	0
9	41	8	48	1	0
10	123	42	164	1	0
11	367	185	551	1	0
12	1 288	888	2 176	0	0
13	4 878	5 110	9 985	1	2
14	19 536	27 436	46 969	1	2
15	85 263	168 030	253 285	2	6
16	379 799	1 008 906	1 388 694	1	10
17	1 769 979	6 283 414	8 053 363	1	29
18	8 400 285	39 866 181	48 266 380	0	86
19	40 619 385	253 511 073	294 130 212	1	245
20	199 631 989	1 647 687 439	1 847 318 507	1	920

Table 3

References

[1] **F Bonahon**, **L C Siebenmann**, *New geometric splittings of classical knots and the classification and symmetries of arborescent knots*, preprint (1979) `https://dornsife.usc.edu/francis-bonahon/wp-content/uploads/sites/205/2023/06/BonSieb-compressed.pdf`

[2] **B A Burton**, *The next* 350 *million knots*, from "36th International Symposium on Computational Geometry", Leibniz Int. Proc. Inform. 164, Schloss Dagstuhl, Wadern (2020) art. id. 25 MR Zbl

[3] **B Burton**, **R Budney**, **W Pettersson**, et al., *Regina*, *software for low-dimensional topology* (1999–2023) `http://regina-normal.github.io`

[4] **J H Conway**, *An enumeration of knots and links, and some of their algebraic properties*, from "Computational problems in abstract algebra", Pergamon, Oxford (1970) 329–358 MR Zbl

[5] **P R Cromwell**, *Knots and links*, Cambridge Univ. Press (2004) MR Zbl

[6] **M Culler**, **N M Dunfield**, **M Goerner**, **J R Weeks**, *SnapPy*, *a computer program for studying the geometry and topology of* 3*-manifolds* `http://snappy.computop.org`

[7] **M Dehn**, *Die beiden Kleeblattschlingen*, Math. Ann. 75 (1914) 402–413 MR

[8] **C H Dowker**, **M B Thistlethwaite**, *Classification of knot projections*, Topology Appl. 16 (1983) 19–31 MR Zbl

[9] **D B A Epstein**, **R C Penner**, *Euclidean decompositions of noncompact hyperbolic manifolds*, J. Differential Geom. 27 (1988) 67–80 MR Zbl

[10] **B Ewing**, **K C Millett**, *A load balanced algorithm for the calculation of the polynomial knot and link invariants*, from "The mathematical heritage of C F Gauss", World Sci., River Edge, NJ (1991) 225–266 MR Zbl

[11] **R H Fox**, *A quick trip through knot theory*, from "Topology of 3-manifolds and related topics", Prentice-Hall, Englewood Cliffs, NJ (1961) 120–167 MR Zbl

[12] **C M Gordon**, **J Luecke**, *Knots are determined by their complements*, J. Amer. Math. Soc. 2 (1989) 371–415 MR Zbl

[13] **R Hartley**, *Identifying noninvertible knots*, Topology 22 (1983) 137–145 MR Zbl

[14] **J Hoste**, **M Thistlethwaite**, **J Weeks**, *The first* $1,701,936$ *knots*, Math. Intelligencer 20 (1998) 33–48 MR Zbl

[15] **T P Kirkman**, *The enumeration, description and construction of knots of fewer than ten crossings*, Trans. Roy. Soc. Edinburgh 32 (1885) 281–309 Zbl

[16] **W B R Lickorish**, *Prime knots and tangles*, Trans. Amer. Math. Soc. 267 (1981) 321–332 MR Zbl

[17] **C N Little**, *On knots, with a census of order ten*, Trans. Connecticut Acad. Sci. 18 (1885) 374–378 Zbl

[18] **W Menasco**, *Closed incompressible surfaces in alternating knot and link complements*, Topology 23 (1984) 37–44 MR Zbl

[19] **K A Perko, Jr**, *On the classification of knots*, Proc. Amer. Math. Soc. 45 (1974) 262–266 MR Zbl

[20] **K A Perko, Jr**, *On dihedral covering spaces of knots*, Invent. Math. 34 (1976) 77–82 MR Zbl

[21] **H Poincaré**, *Analysis situs*, J. École Polytech. 1 (1895) 1–123 Zbl

[22] **S Rankin**, **O Flint**, **J Schermann**, *Enumerating the prime alternating knots, I*, J. Knot Theory Ramifications 13 (2004) 57–100 MR Zbl

[23] **R Riley**, *Discrete parabolic representations of link groups*, Mathematika 22 (1975) 141–150 MR Zbl

[24] **M Sakuma**, **J R Weeks**, *The generalized tilt formula*, Geom. Dedicata 55 (1995) 115–123 MR Zbl

[25] **O Schreier**, *Über die gruppen* $A^a B^b = 1$, Abh. Math. Sem. Univ. Hamburg 3 (1924) 167–169 MR Zbl

[26] **H Schubert**, *Knoten und Vollringe*, Acta Math. 90 (1953) 131–286 MR Zbl

[27] **C Sundberg**, **M Thistlethwaite**, *The rate of growth of the number of prime alternating links and tangles*, Pacific J. Math. 182 (1998) 329–358 MR Zbl

[28] **P G Tait**, *On knots, I*, Trans. Roy. Soc. Edinburgh 28 (1877) 145–190 Zbl Reprinted in his "Scientific Papers, I", Cambridge Univ. Press (1896) 273–317

[29] **P G Tait**, *On knots, II*, Trans. Roy. Soc. Edinburgh 32 (1884) 327–342 Zbl Reprinted in his "Scientific Papers, I", Cambridge Univ. Press (1896) 318–334

[30] **P G Tait**, *On knots, III*, Trans. Roy. Soc. Edinburgh 32 (1885) 493–506 Zbl Reprinted in his "Scientific Papers, I", Cambridge Univ. Press (1896) 335–347

[31] **M Thistlethwaite**, *On the structure and scarcity of alternating links and tangles*, J. Knot Theory Ramifications 7 (1998) 981–1004 MR Zbl

[32] **H F Trotter**, *Non-invertible knots exist*, Topology 2 (1963) 275–280 MR Zbl

Department of Mathematics, University of Tennessee
Knoxville, TN, United States

mthistle@utk.edu

https://web.math.utk.edu/~morwen

Received: 13 May 2023 Revised: 12 November 2023

Algebraic & Geometric Topology **25**:1 (2025) 345–355
DOI: 10.2140/agt.2025.25.345
Published: 5 March 2025

An exotic presentation of $\mathbb{Z} \times \mathbb{Z}$ and the Andrews–Curtis conjecture

JONATHAN ARIEL BARMAK

We prove that the presentations $\langle x, y \mid [x, y], 1 \rangle$ and $\langle x, y \mid [x, [x, y^{-1}]]^2 y [y^{-1}, x] y^{-1}, [x, [[y^{-1}, x], x]] \rangle$ are not Q^*-equivalent even though their standard complexes have the same simple homotopy type.

20F05, 20F65, 57M05, 57Q10

1 Introduction

A finite presentation $\langle x_1, x_2, \dots, x_n \mid r_1, r_2, \dots, r_m \rangle$ of a group G can be transformed into another presentation of G by performing one of the following:

(i) Changing a relator r_j by $r_j r_i$ for some $i \neq j$.

(ii) Changing a relator r_j by r_j^{-1}.

(iii) Changing a relator r_j by a conjugate $g r_j g^{-1}$ for some g in the free group $F(x_1, x_2, \dots, x_n)$.

(iv) Changing each relator r_j by $\phi(r_j)$, where ϕ is an automorphism of $F(x_1, x_2, \dots, x_n)$.

(v) Adding a generator x_{n+1} and a relator r_{n+1} which coincides with x_{n+1}.

(vi) The inverse of (v), when possible.

Moves (i) to (iii) are called Q-transformations, (i) to (iv) are Q^*-transformations and moves (i) to (vi) are called Q^{**}-transformations. Two finite presentations are said to be Q-equivalent (Q^* or Q^{**}) if one can be obtained from the other by performing a sequence of Q-transformations (Q^* or Q^{**}). The Andrews–Curtis conjecture [1] states that any two presentations of the trivial group with n generators and $m = n$ relators are Q-equivalent. The weak version of the Andrews–Curtis conjecture states that in the same situation the two presentations are just Q^{**}-equivalent. The latter is equivalent to the statement that any contractible finite 2-dimensional CW-complex 3-deforms to a point. The so-called generalized Andrews–Curtis conjecture [13, Section 4.1] says that any two presentations $\mathcal{P}$ and $\mathcal{Q}$ with simple homotopy equivalent standard complexes $K_{\mathcal{P}}$ and $K_{\mathcal{Q}}$ are Q^{**}-equivalent. These three conjectures are open.

Although the original conjecture says that in some cases any Q^{**}-equivalence is a Q-equivalence, there are known examples of presentations which are Q^{**}-equivalent but Q^*-inequivalent. The first were probably those given by Zieschang [24, page 36]: $\langle x, y \mid x^3 y^5 \rangle$ and $\langle x, y \mid x^3 y^3 x^3 y^2 \rangle$, and, with more generality, by McCool and Pietrowski [18]: $\langle x, y \mid x^k y^{pt+1} \rangle$ and $\langle x, y \mid (x^k y^t)^p y \rangle$ for $k, p, t \geq 2$.

In each case the presentations are Q^{**}-equivalent. The authors only prove that they present the same group, but it is easy to translate their proof to the language of Q^{**}-transformations. On the other hand, Whitehead's algorithm can be used to prove that the relators of all these presentations are of minimal length in $\mathbb{F}_2 = F(x, y)$, ie in their orbit under the $\mathrm{Aut}(\mathbb{F}_2)$ action. Since in each example the relators have different length, there is no automorphism of $\mathbb{F}_2$ taking one to the other or its inverse. Thus, these one-relator presentations are not Q^*-equivalent. In fact, the construction of [18] shows that for every $m > 0$ there are m presentations which are pairwise Q^{**}-equivalent and Q^*-inequivalent. By a result of Magnus, the normal closures $N(r)$ and $N(s)$ of two elements r and s in a free group coincide if and only if r is a conjugate of s or its inverse. Thus, these one-relator examples are not Q^*-equivalent for an elementary reason: there is no automorphism of $\mathbb{F}_2$ taking the normal closure of one relator to the normal closure of the other. In [19], Metzler gives an example of a different nature. He shows that the presentations $\langle x, y \mid x^5, y^5, [x, y]\rangle$ and $\langle x, y \mid x^5, y^5, [x^2, y]\rangle$ are Q^{**}-equivalent and not Q^*-equivalent. In this case the normal closures of both sets of relators coincide.

McCool–Pietrowski's family and Metzler's example occur at minimal Euler characteristic (ie the Euler characteristic of the standard complexes of those presentations are minimal among all possible presentations of the same groups), though the one-relator examples can be stabilized (by adding trivial relations) to obtain Q^{**}-equivalent and Q^*-inequivalent presentations (the same argument can be used) arbitrarily far from minimal Euler characteristic.

In this article we construct presentations $\mathcal{P}$ and $\mathcal{Q}$ of $\mathbb{Z} \times \mathbb{Z}$ each with two generators and two relators, having simple homotopy equivalent standard complexes, but such that $\mathcal{P}$ and $\mathcal{Q}$ are not Q^*-equivalent. The normal closures of both relator sets coincide and the Euler characteristic is one above the minimal level.

Non-homotopy-equivalent presentations of the same group G with equal Euler characteristic have been constructed by using stably free nonfree $\mathbb{Z}[G]$-modules. Although every finitely generated projective $\mathbb{Z}[G]$-module is free when $G = \mathbb{Z} \times \mathbb{Z}$, in our construction we need to distinguish Q-equivalence classes, and we use a more subtle idea: an exotic basis of $\mathbb{Z}[G]^2$ which is not obtained from the standard basis by elementary row operations. The basis change matrix, which is not a product of elementary and diagonal matrices, was found by Evans in [9]. The second ingredient of our proof is a new invariant called the winding invariant. The present article was meant to be a section in the paper [3], which introduces this invariant along with applications. We believe that it is better to present this example in a separate article.

There are known examples of presentations $\mathcal{P}_1$ and $\mathcal{P}_2$ which are

(a) not simple homotopy equivalent but homotopy equivalent, or
(b) not homotopy equivalent, but such that the stabilized presentations $\mathcal{P}_1^{(1)}$ and $\mathcal{P}_2^{(1)}$, obtained by adding one trivial relator, become Q-equivalent.

These examples have minimal Euler characteristic. Our presentations $\mathcal{P}$ and $\mathcal{Q}$ are above minimal Euler characteristic and become Q-equivalent after one stabilization.

After the first version of this paper was finished, X Fernández [10] proved that the presentations $\mathscr{P}$ and $\mathscr{Q}$ are in fact Q^{**}-equivalent. Also, at that time we were not aware of Zieschang, McCool–Pietrowski and Metzler's examples. Although the results of the present version and the original are essentially the same, we have added comments suggested by two anonymous referees. The results about the tree of Q-equivalence classes were motivated by one of the referees' comments. We are grateful to the referees for their suggestions.

Projecting to $\mathbb{F}_2'/\mathbb{F}_2''$ Previously known methods for proving that the presentations $\langle x_1, x_2, \dots, x_n \mid r_1, r_2, \dots, r_m\rangle$ and $\langle x_1, x_2, \dots, x_n \mid s_1, s_2, \dots, s_m\rangle$ are not Q or Q^{**}-equivalent were developed by Browning [6], Hog-Angeloni and Metzler [12, Section 2.2] and Borovik, Lubotzky and Myasnikov [5]. As explained by Hog-Angeloni and Metzler in [14, Section 2.2], the idea is to define a homomorphism q from $F(x_1, x_2, \dots, x_n)$ to a test group G^* and prove that the m-tuples $(q(r_1), q(r_2), \dots, q(r_m))$ and $(q(s_1), q(s_2), \dots, q(s_m))$ are not equivalent. The results of Hog-Angeloni and Metzler [13, Theorems 2.3 and 2.4] show that solvable groups are not useful as test groups to distinguish Q^{**}-equivalence. In Borovik, Lubotzky and Myasnikov [5], Browning [7] and Hog-Angeloni and Metzler [14], there are results concerning the use of finite groups to distinguish Q and Q^{**}-equivalences. Our methods appeared as a natural application when studying the winding invariant. However, they can be seen through this perspective as an application of solvable groups to distinguish Q-equivalence. Concretely we use $G^* = \mathbb{F}_2/\mathbb{F}_2''$, the free metabelian group of rank 2.

The presentations $\mathscr{P}$ and $\mathscr{Q}$ present $\mathbb{Z}\times\mathbb{Z}$, so their Whitehead group is trivial. The torsion $\tau(f) \in \mathrm{Wh}(K_{\mathscr{P}})$ of any homotopy equivalence $f\colon K_{\mathscr{Q}} \to K_{\mathscr{P}}$ is therefore trivial. To distinguish Q^*-equivalence classes we will work in $\mathrm{GL}_2(\mathbb{Z}[\pi_1(\mathscr{P})])/\mathrm{GE}_2(\mathbb{Z}[\pi_1(\mathscr{P})])$ instead of $\mathrm{Wh}(K_{\mathscr{P}}) = \mathrm{GL}(\mathbb{Z}[\pi_1(\mathscr{P})])/\mathrm{GE}(\mathbb{Z}[\pi_1(\mathscr{P})])$. We recall the definitions of these concepts.

Given a ring R, denote by $E_n(R)$ the subgroup of $\mathrm{GL}_n(R)$ generated by the elementary matrices. Recall that $E \in \mathrm{GL}_n(R)$ is elementary if all the diagonal coefficients are $1 \in R$ and all the other coefficients but one are $0 \in R$. We call $D_n(R)$ the subgroup of $\mathrm{GL}_n(R)$ of diagonal matrices and $\mathrm{GE}_n(R)$ the subgroup of $\mathrm{GL}_n(R)$ generated by $E_n(R)$ and $D_n(R)$. Note that $D_n(R)$ normalizes $E_n(R)$, so $\mathrm{GE}_n(R) = D_n(R)E_n(R) = E_n(R)D_n(R)$. If R is a Euclidean ring, then $\mathrm{GE}_n(R) = \mathrm{GL}_n(R)$ for every $n \geq 1$. A ring R is said to be generalized Euclidean if $\mathrm{GE}_n(R) = \mathrm{GL}_n(R)$ for every n. It was proved by Bachmuth and Mochizuki [2, Theorem 1] that $R = \mathbb{Z}[X, X^{-1}, Y, Y^{-1}]$ is not generalized Euclidean. Evans [9, Theorem C] gives a concrete example of a 2 by 2 invertible matrix over that ring which is not in $E_2(R)$.

From now on R will denote the ring $\mathbb{Z}[X, X^{-1}, Y, Y^{-1}]$.

Theorem 1 (Evans) *The matrix*

$$\begin{pmatrix} 1-2(X-1)Y^{-1} & 4Y^{-1} \\ -(X-1)^2Y^{-1} & 1+2(X-1)Y^{-1} \end{pmatrix}$$

is in $\mathrm{GL}_2(R)$ *but not in* $\mathrm{GE}_2(R)$.

The commutator subgroup of $\mathbb{F}_2$ is denoted by $[\mathbb{F}_2,\mathbb{F}_2]$ or $\mathbb{F}_2'$. Recall that $w\in\mathbb{F}_2$ lies in $[\mathbb{F}_2,\mathbb{F}_2]$ if and only if the total exponents of x and of y in w are both 0.

Definition 2 Let $w\in\mathbb{F}_2$. Then $w=x_1^{\epsilon_1}x_2^{\epsilon_2}\cdots x_t^{\epsilon_t}$, where $x_i\in\{x,y\}$ and $\epsilon_i\in\{1,-1\}$ for each i. The word w determines a path γ_w in $\mathbb{R}^2$ which begins in $(0,0)$ and is a concatenation of paths $\gamma_1,\gamma_2,\dots,\gamma_l$. The path γ_i moves one unit parallel to the axis x_i and with positive or negative direction depending on the sign ϵ_i. The image of γ_w is contained in the grid $\mathbb{Z}\times\mathbb{R}\cup\mathbb{R}\times\mathbb{Z}$. The ending point of γ_w is (k,l), where k is the total exponent of x in w and l is the total exponent of y. Suppose $w\in[\mathbb{F}_2,\mathbb{F}_2]$, so γ_w finishes in $(0,0)$. For each $(i,j)\in\mathbb{Z}\times\mathbb{Z}$, let $a_{i,j}$ be the winding number $\mathrm{w}\big(\gamma_w,i+\frac{1}{2},j+\frac{1}{2}\big)$ of γ_w around the point $p=\big(i+\frac{1}{2},j+\frac{1}{2}\big)$. Define the *winding invariant* $P_w\in R=\mathbb{Z}[X,X^{-1},Y,Y^{-1}]$ of w to be the Laurent polynomial $P_w=\sum a_{i,j}X^iY^j$.

Our notation for commutators is $[u,v]=uvu^{-1}v^{-1}$. So, for instance, the winding invariant of $[x,y]$ is $P_{[x,y]}=1\in R$ and $P_{[y^{-1},x]}=Y^{-1}$. The path γ_w is just the lift of w to the Cayley graph $\Gamma(\mathbb{Z}\times\mathbb{Z},\{x,y\})=\mathbb{Z}\times\mathbb{R}\cup\mathbb{R}\times\mathbb{Z}$, and the winding invariant $\mathbb{F}_2'\to R$ can be seen as the projection $N\to N/N'$ of the normal closure N of $[x,y]$ in $\mathbb{F}_2$ onto the relation module N/N' of the presentation $\langle x,y\mid[x,y]\rangle$; see [3, Section 8]. If $w\in\mathbb{F}_2'$, then $\psi(\partial w/\partial x)=(1-Y)P_w$ and $\psi(\partial w/\partial y)=(X-1)P_w$; see [3, Section 10]. Here $\partial/\partial x,\partial/\partial y:\mathbb{Z}[\mathbb{F}_2]\to\mathbb{Z}[\mathbb{F}_2]$ are the Fox derivatives, $\psi:\mathbb{Z}[\mathbb{F}_2]\to\mathbb{Z}[\mathbb{F}_2/\mathbb{F}_2']$ is the canonical projection, and $\mathbb{Z}[\mathbb{F}_2/\mathbb{F}_2']$ is identified with R via the isomorphism which maps the class of x to X and the class of y to Y. Thus, P_w is essentially the Alexander polynomial of the group $\langle x,y\mid w\rangle$ (which is only defined up to a multiplication by a unit of R and a change of basis of $\mathbb{Z}^2$). The geometric nature of our definition is useful to understand the intuition behind the algebraic arguments.

The proof of the next result is clear from the definition (see [3, Proposition 7]), the comments above about relation modules or the relation with Fox derivatives.

Lemma 3 *Let $w,w'\in[\mathbb{F}_2,\mathbb{F}_2]$, $u\in\mathbb{F}_2$. Then the following hold*:

(i) $P_{ww'}=P_w+P_{w'}$.

(ii) $P_{w^{-1}}=-P_w$.

(iii) $P_{uwu^{-1}}=X^kY^lP_w$, *where k and l are the total exponents of x and y in u.*

(iv) $P_{[u,w]}=(X^kY^l-1)P_w$.

Call a presentation $\mathcal{P}=\langle x,y\mid r_1,r_2,\dots,r_m\rangle$ *cocommutative* if each relator r_j lies in $[\mathbb{F}_2,\mathbb{F}_2]$.

Remark 4 Every presentation Q^*-equivalent to a cocommutative presentation is also cocommutative. We can associate to a cocommutative presentation $\mathcal{P}$ the vector $\Lambda(\mathcal{P})=(P_{r_1},P_{r_2},\dots,P_{r_m})\in R^m$. This can also be seen as the first column in the Alexander matrix $A\in R^{m\times 2}$ of $\langle x,y\mid r_1,r_2,\dots,r_m\rangle$ divided by $(1-Y)$. The effect on $\Lambda(\mathcal{P})$ of performing a Q-transformation on $\mathcal{P}$ is to change a polynomial P_{r_j} by $P_{r_j}+P_{r_i}$ for certain $i\neq j$, or by $-P_{r_j}$, or by $X^kY^lP_{r_j}$ for certain $k,l\in\mathbb{Z}$. Therefore, if $\mathcal{P}$ is a cocommutative presentation and $\mathcal{Q}$ is Q-equivalent to $\mathcal{P}$, then $\Lambda(\mathcal{Q})^t=E\Lambda(\mathcal{P})^t$ for some $E\in\mathrm{GE}_m(R)$. Here $\Lambda(\mathcal{P})^t,\Lambda(\mathcal{Q})^t\in R^{m\times 1}$ denote the column vectors.

Figure 1: Left, the graphical representation of the curve γ_{r_1} which begins in the black dot. The represented curve and the actual curve are homotopic in the plane with the centers of the squares removed. In the interior of the squares we see the corresponding winding numbers. Right, the curve γ_{r_2}.

If $\mathcal{P}=\langle x,y\mid r_1,r_2,\dots,r_m\rangle$ and $\mathcal{Q}=\langle x,y\mid s_1,s_2,\dots,s_m\rangle$ are cocommutative presentations such that the normal closure of $\{r_1,r_2,\dots,r_m\}$ coincides with the normal closure of $\{s_1,s_2,\dots,s_m\}$, then each s_j is a product of conjugates of r_i's and inverses of r_i's. By Lemma 3, P_{s_j} is an R-linear combination of the P_{r_i}. Therefore $\Lambda(\mathcal{Q})^t=M\Lambda(\mathcal{P})^t$ for certain $M\in R^{m\times m}$. Symmetrically, $\Lambda(\mathcal{P})^t=M'\Lambda(\mathcal{Q})^t$ for some $M'\in R^{m\times m}$.

Let

$$\mathcal{P}=\langle x,y\mid [x,y],1\rangle \quad\text{and}\quad \mathcal{Q}=\langle x,y\mid [x,[x,y^{-1}]]^2y[y^{-1},x]y^{-1},[x,[[y^{-1},x],x]]\rangle.$$

We state the main result of the article.

Theorem 5 *The standard complexes $K_{\mathcal{P}}$ and $K_{\mathcal{Q}}$ are simple homotopy equivalent, while $\mathcal{P}$ and $\mathcal{Q}$ are not Q^*-equivalent. Moreover, the normal closures of both relator sets coincide. Also, these examples occur at Euler characteristic one above the minimal level.*

Since the Whitehead group of $\pi_1(K_{\mathcal{P}})=\mathbb{Z}\times\mathbb{Z}$ is trivial [4], to prove simple homotopy equivalence we only need to prove homotopy equivalence. This can be achieved by standard methods and we postpone this part to the end of the proof.

We begin by proving that $\mathcal{P}$ and $\mathcal{Q}$ are not Q-equivalent. We compute first $\Lambda(\mathcal{Q})$. Let

$$r_1=[x,[x,y^{-1}]]^2y[y^{-1},x]y^{-1} \quad\text{and}\quad r_2=[x,[[y^{-1},x],x]]$$

be the relators of $\mathcal{Q}$. By Lemma 3 the winding invariant of r_1 is $2(X-1)P_{[x,y^{-1}]}+YP_{[y^{-1},x]}$. Since $P_{[y^{-1},x]}=Y^{-1}$ and $P_{[x,y^{-1}]}=-P_{[y^{-1},x]}=-Y^{-1}$, it follows that $P_{r_1}=1-2(X-1)Y^{-1}$; see Figure 1. Similarly, $P_{r_2}=(X-1)P_{[[y^{-1},x],x]}=-(X-1)^2P_{[y^{-1},x]}=-(X-1)^2Y^{-1}$. Thus,

$$\Lambda(\mathcal{Q})=(1-2(X-1)Y^{-1},-(X-1)^2Y^{-1})$$

is the first column of matrix M in Theorem 1.

On the other hand it is easy to see that $\Lambda(\mathcal{P})=(1,0)\in R^2$. If $\mathcal{P}$ and $\mathcal{Q}$ are Q-equivalent, by Remark 4 there exists a matrix $E\in\mathrm{GE}_2(R)$ such that $M\binom{1}{0}=\Lambda(\mathcal{Q})^t=E\Lambda(\mathcal{P})^t=E\binom{1}{0}$. Then $E^{-1}M\binom{1}{0}=\binom{1}{0}$.

Thus $E^{-1}M$ is a matrix of the form

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

for some $A, B \in R$. Moreover, since $E, M \in \mathrm{GL}_2(R)$, $B \in R$ is a unit, so $E_2 E_1 E^{-1} M = \mathrm{Id}$ for the diagonal and elementary matrices

$$E_1 = \begin{pmatrix} 1 & 0 \\ 0 & B^{-1} \end{pmatrix} \quad \text{and} \quad E_2 = \begin{pmatrix} 1 & -A \\ 0 & 1 \end{pmatrix}.$$

Then $M = E E_1^{-1} E_2^{-1} \in \mathrm{GE}_2(R)$, which contradicts Theorem 1. This completes the proof that $\mathcal{P}$ and $\mathcal{Q}$ are not Q-equivalent. The last lines of our proof are implicit in the comments of [8, page 115] about unimodular columns. The matrix M was used by Myasnikov, Myasnikov and Shpilrain in [20, Theorem 1.6] to study Q-transformations of m-tuples in a nonfree group G. As we mentioned above, the idea of our proof naturally appeared when studying applications of the winding invariant. We discovered later that our methods are very similar to those used in the proof of [20, Proposition 5.1].

The fact that $\mathcal{P}$ and $\mathcal{Q}$ are not Q^*-equivalent, either, will follow from the next key lemma.

Lemma 6 *The normal closure N of $\{r_1, r_2\}$ is $[\mathbb{F}_2, \mathbb{F}_2]$.*

Before we give a proof the lemma, we show how to use it to prove that $\mathcal{P}$ and $\mathcal{Q}$ are not Q^*-equivalent. Suppose they are. Then there is an automorphism ϕ of $\mathbb{F}_2$ such that $\phi\mathcal{P} = \langle x, y \mid \phi([x, y]), 1\rangle$ and $\mathcal{Q}$ are Q-equivalent. Since Q-transformations preserve the normal closure of the relators, the normal closure of $\phi([x, y])$ is $[\mathbb{F}_2, \mathbb{F}_2] = N([x, y])$ by Lemma 6. By a well-known result of Magnus, $\phi([x, y])$ is a conjugate of $[x, y]$ or $[x, y]^{-1}$. In any case, $\phi\mathcal{P}$ is Q-equivalent to $\mathcal{P}$, so $\mathcal{P}$ and $\mathcal{Q}$ are Q-equivalent, a contradiction.

Alternatively, that $\mathcal{P}$ and $\mathcal{Q}$ are not Q^*-equivalent follows from the fact that there is no matrix $E \in \mathrm{GE}_2(R)$ such that $E\Lambda(\mathcal{P})^t = \Lambda(\mathcal{Q})^t$, and the following. If $\phi \in \mathrm{Aut}(\mathbb{F}_2)$, then the winding invariant $P_{\phi([x,y])}$ is a unit of R, so there exists $E' \in \mathrm{GE}_2(R)$ such that $E'\Lambda(\mathcal{P})^t = \Lambda(\phi\mathcal{P})^t$; see [3]. Then $\phi\mathcal{P}$ and $\mathcal{Q}$ cannot be Q-equivalent.

Proof of Lemma 6 It is clear that $r_1, r_2 \in [\mathbb{F}_2, \mathbb{F}_2]$, so we only need to show that $[x, y] = 1$ in $\mathbb{F}_2/N$. Let $d = [x, y^{-1}] = xy^{-1}x^{-1}y$. Since

$$1 = r_2 = [x, [d^{-1}, x]] = xd^{-1}xdx^{-1}x^{-1}xd^{-1}x^{-1}d = xd^{-1}xdx^{-1}d^{-1}x^{-1}d$$

in $\mathbb{F}_2/N$, it follows that $xdx^{-1}d^{-1} = dx^{-1}d^{-1}x = x^{-1}(xdx^{-1}d^{-1})x$. Therefore $e = [x, d]$ commutes with x in $\mathbb{F}_2/N$.

On the other hand $1 = r_1 = [x, d]^2 yd^{-1}y^{-1} = e^2 yd^{-1}y^{-1}$, so

$$d = y^{-1}e^2y. \tag{1}$$

Finally, by definition, $e = xdx^{-1}d^{-1} = x(y^{-1}e^2y)x^{-1}d^{-1}$. But since $d = xy^{-1}x^{-1}y$, we have $dy^{-1}x = xy^{-1}$. We use this in the previous equation to obtain

$$e = dy^{-1}xe^2x^{-1}yd^{-1}d^{-1}.$$

We use now that e and x commute to deduce $e = dy^{-1}e^2yd^{-2} = ddd^{-2} = 1$ in $\mathbb{F}_2/N$. The last equality follows from (1). By (1) again we deduce $d = 1$ in $\mathbb{F}_2/N$. Thus, $d = [x, y^{-1}] \in N$, so $[x, y] \in N$. □

To finish the proof of the theorem we need to prove that $K_{\mathcal{P}}$ and $K_{\mathcal{Q}}$ are homotopy equivalent. We have done the most important part already in Lemma 6. It implies that $K_{\mathcal{P}}$ and $K_{\mathcal{Q}}$ have isomorphic fundamental groups with an isomorphism $\pi_1(K_{\mathcal{Q}}) \to \pi_1(K_{\mathcal{P}})$ induced by a map which is the identity on 1-skeletons. Then $K_{\mathcal{P}}$ and $K_{\mathcal{Q}}$ have isomorphic fundamental groups and the same Euler characteristic. In general this does not imply homotopy equivalence, but it does in our case since $\pi_1(K_{\mathcal{P}})$ is free abelian of rank 2. This is explained by Harlander in [11]: Suppose K and L are finite connected 2-dimensional complexes with $\pi_1(K) \cong \pi_1(L) \cong \mathbb{Z}\times\mathbb{Z}$ and $\chi(K) = \chi(L)$. Since $\mathbb{Z}\times\mathbb{Z}$ is aspherical, $H_2(\widetilde{K})$ and $H_2(\widetilde{L})$ are projective R-modules by the generalized Schanuel lemma. By the Quillen–Suslin theorem, these R-modules are free; see Swan's comments [23] on how to go from polynomials to Laurent polynomials. Since $\chi(K) = \chi(L)$, it follows that $H_2(\widetilde{K})$ and $H_2(\widetilde{L})$ have the same rank, so they are isomorphic. Once again, since $\mathbb{Z}\times\mathbb{Z}$ is aspherical, $H^3(\mathbb{Z}\times\mathbb{Z}, H_2(\widetilde{K})) = H^3(\mathbb{Z}\times\mathbb{Z}, H_2(\widetilde{L})) = 0$, so K and L have isomorphic algebraic 2-types. By Mac Lane–Whitehead's theorem [22, Theorem 4.9], K and L are homotopy equivalent. This finishes the proof of Theorem 5.

An alternative and simpler way to see that $K_{\mathcal{P}}$ and $K_{\mathcal{Q}}$ are homotopy equivalent is to show that there exists a homomorphism $\Phi\colon C_2(\widetilde{K}_{\mathcal{Q}}) \to C_2(\widetilde{K}_{\mathcal{P}})$ of R-modules that makes the following diagram commutative

$$\begin{tikzcd}
C_2(\widetilde{K}_{\mathcal{Q}}) \arrow[r, "d_2"] \arrow[d, "\Phi"] & C_1(\widetilde{K}_{\mathcal{Q}}) \arrow[d, equal] \\
C_2(\widetilde{K}_{\mathcal{P}}) \arrow[r, "d_2'"] & C_1(\widetilde{K}_{\mathcal{P}})
\end{tikzcd}$$

A computation of Fox derivatives, or our comments right after Definition 2, show that d_2 has matrix representation

$$\begin{pmatrix} (1-Y)(1-2(X-1)Y^{-1}) & (X-1)^2(1-Y^{-1}) \\ (X-1)(1-2(X-1)Y^{-1}) & -(X-1)^3Y^{-1} \end{pmatrix} = \begin{pmatrix} 1-Y \\ X-1 \end{pmatrix} \Lambda(\mathcal{Q}),$$

and d_2' is represented by

$$\begin{pmatrix} 1-Y & 0 \\ X-1 & 0 \end{pmatrix} = \begin{pmatrix} 1-Y \\ X-1 \end{pmatrix} \Lambda(\mathcal{P}).$$

Therefore the map $\Phi\colon C_2(\widetilde{K}_{\mathcal{Q}}) \to C_2(\widetilde{K}_{\mathcal{P}})$ represented by the transpose of the (invertible) matrix M in Theorem 1 is an isomorphism satisfying $d_2'\Phi = d_2$, and by [22, Theorem 3.9], there is a homotopy equivalence $f\colon K_{\mathcal{Q}} \to K_{\mathcal{P}}$.

The fact that the Euler characteristic of $\mathcal{P}$ and $\mathcal{Q}$ is one level above the minimal follows from the general bound $\chi(K) \geq 1 - \dim H_1(G;\mathbb{Q}) + \dim H_2(G;\mathbb{Q})$ that holds for any finite 2-dimensional complex K with $\pi_1(K) \cong G$, which in turn can be deduced from the fact that there is an epimorphism $H_2(K;\mathbb{Q}) \to H_2(G;\mathbb{Q})$ induced by the inclusion of K into a $K(G,1)$.

The tree of Q-equivalence Let G be a group. If K and L are finite 2-dimensional complexes with fundamental group G, there exist $k, l \geq 0$ such that $K \vee \bigvee_{i=1}^{k} S^2 \simeq L \vee \bigvee_{i=1}^{l} S^2$. In fact for some k, l, one of these complexes 3-deforms into the other [13, page 28]. The tree of homotopy types of finite 2-dimensional complexes with fundamental group G has a directed edge from the homotopy type of K to the homotopy type of $K \vee S^2$. The classification problem consists in understanding this tree for each group G. This has been achieved in few examples (free groups, finite abelian, and few others), and interesting features of the tree have been found in other cases. The complexes of minimal Euler characteristic (minimal among all the 2-complexes with fundamental group G) are roots of this tree, but there can be roots above minimal characteristic. Very recently Nicholson [21] proved that for every $k \geq 0$ there exists a group G and (infinitely many) distinct homotopy types of 2-complexes with fundamental group G and Euler characteristic equal to the minimal plus k. It is an open problem whether there exist 2-complexes X and Y such that $X \vee S^2 \not\simeq Y \vee S^2$ while $X \vee S^2 \vee S^2 \simeq Y \vee S^2 \vee S^2$.

The trees of simple homotopy types and 3-deformation types are defined similarly. If two finite presentations $\mathcal{P}_1 = \langle X \mid R \rangle, \mathcal{P}_2 = \langle X \mid S \rangle$ with same generator set X have their relators with equal normal closure $N(R) = N(S)$, then there exists $k, l \geq 0$ such that the presentations $\mathcal{P}_1^{(k)}$ and $\mathcal{P}_2^{(l)}$ obtained from $\mathcal{P}_1$ and $\mathcal{P}_2$ by adding k and l trivial relators respectively, are Q-equivalent. In fact, k and l can be taken as $|S \setminus R|$ and $|R \setminus S|$, respectively. Given a group G, a finite set X and a normal subgroup $N \trianglelefteq F(X)$ such that $F(X)/N \cong G$, the tree of Q-equivalence classes of finite presentations of G with respect to X and N has a directed edge from the Q-equivalence class of a presentation $\mathcal{P}_1 = \langle X \mid R \rangle$ with $N(R) = N$ to the class of $\mathcal{P}_1^{(1)}$. The example of Metzler of presentations of $\mathbb{Z}_5 \times \mathbb{Z}_5$ with minimal Euler characteristic shows that the tree of Q-equivalence classes of $\mathbb{Z}_5 \times \mathbb{Z}_5$ with respect to $\{x, y\}$, $N(x^5, y^5, [x, y])$ has two roots $\langle x, y \mid R \rangle$ and $\langle x, y \mid S \rangle$ of minimal Euler characteristic, which are adjacent to a same type, since $|R \setminus S| = |S \setminus R| = 1$; see Figure 2. Our example shows that a different situation may happen. Our presentations $\mathcal{P}$ and $\mathcal{Q}$ have Euler characteristic one above the minimal level. In fact $\mathcal{P}$ is not a root, but $\mathcal{Q}$ is. Moreover, the tree of Q-equivalence classes of presentations of $\mathbb{Z} \times \mathbb{Z}$ with respect to $\{x, y\}$ and $\mathbb{F}_2'$

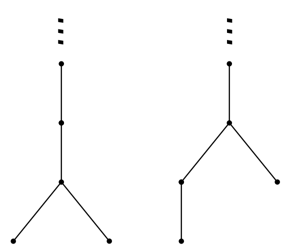


Figure 2: Part of the trees of Q-equivalence classes of Metzler's example and ours.

has a unique class with minimal Euler characteristic, which is $\langle x, y \mid [x, y]\rangle$. We have that $\mathscr{P}^{(1)}$ and $\mathscr{Q}^{(1)}$ are Q-equivalent.

Note that the Andrews–Curtis conjecture says that the tree of Q-equivalence classes of presentations of the trivial group with respect to $\{x_1, x_2, \dots, x_n\}$ and $\mathbb{F}_n$ has a unique class with minimal Euler characteristic.

We mention other known examples, from the perspective of the tree of Q-equivalence. In [16], Lustig proves that the presentations

$$\mathscr{P}_1 = \langle x, y, z \mid y^3, yx^{10}y^{-1}x^{-5}, [x^7, z]\rangle,$$
$$\mathscr{P}_2 = \langle x, y, z \mid y^3, yx^{10}y^{-1}x^{-5}, x^{14}zx^{14}z^{-1}x^{-7}zx^{-21}z^{-1}\rangle,$$

are homotopy equivalent and not simple homotopy equivalent. Moreover, he proves that the normal closures of relators coincide. Thus $\mathscr{P}_1^{(1)}$ is Q-equivalent to $\mathscr{P}_2^{(1)}$.

The twisted presentations of a finite abelian group (see Latiolais [15]) can be used to show that for any k there exists a group G and presentations $\mathscr{P}_1, \mathscr{P}_2, \dots, \mathscr{P}_k$ of G with minimal Euler characteristic whose standard complexes are pairwise not homotopy equivalent but such that $\mathscr{P}_i^{(1)}$ and $\mathscr{P}_j^{(1)}$ are Q-equivalent for every $1 \le i, j \le k$.

In [17] Mannan and Popiel give an example of two presentations $\mathscr{P}_1 = \langle x, y \mid y^2x^{-7}, xyxy^{-1}\rangle$ and $\mathscr{P}_2 = \langle x, y \mid y^2x^{-7}, y^{-1}xyx^2y^{-1}x^{-2}yx^{-3}\rangle$ of the quaternion group Q_{28} with nonisomorphic second homotopy modules for any identification of fundamental groups (and thus not homotopy equivalent). Again they have minimal Euler characteristic. It is proved that both relator sets have equal normal closures. Thus, $\mathscr{P}_1^{(1)}$ and $\mathscr{P}_2^{(1)}$ are Q-equivalent.

Acknowledgements The author is a researcher of CONICET and was partially supported by grant UBACyT 20020160100081BA.

References

[1] **J J Andrews**, **M L Curtis**, *Free groups and handlebodies*, Proc. Amer. Math. Soc. 16 (1965) 192–195 MR Zbl

[2] **S Bachmuth**, **H Y Mochizuki**, *$E_2 \neq \mathrm{SL}_2$ for most Laurent polynomial rings*, Amer. J. Math. 104 (1982) 1181–1189 MR Zbl

[3] **J A Barmak**, *The winding invariant*, preprint (2019) arXiv 1904.10072

[4] **H Bass**, **A Heller**, **R G Swan**, *The Whitehead group of a polynomial extension*, Inst. Hautes Études Sci. Publ. Math. 22 (1964) 61–79 MR Zbl

[5] **A V Borovik**, **A Lubotzky**, **A G Myasnikov**, *The finitary Andrews–Curtis conjecture*, from "Infinite groups: geometric, combinatorial and dynamical aspects" (L Bartholdi, T Ceccherini-Silberstein, T Smirnova-Nagnibeda, A Zuk, editors), Progr. Math. 248, Birkhäuser, Basel (2005) 15–30 MR Zbl

[6] **W J Browning**, *The effect of Curtis–Andrews moves on Jacobian matrices of perfect groups*, unpublished manuscript, Cornell University (1976)

[7] **W J Browning**, *Normal generators of finite groups*, unpublished manuscript (1976)

[8] **K Dennis**, **B Magurn**, **L Vaserstein**, *Generalized Euclidean group rings*, J. Reine Angew. Math. 351 (1984) 113–128 MR Zbl

[9] **M J Evans**, *Primitive elements in the free metabelian group of rank* 3, J. Algebra 220 (1999) 475–491 MR Zbl

[10] **X Fernández**, *Morse theory for group presentations*, Trans. Amer. Math. Soc. 377 (2024) 2495–2523 MR Zbl

[11] **J Harlander**, *A survey of recent progress on some problems in* 2*-dimensional topology*, from "Advances in two-dimensional homotopy and combinatorial group theory" (W Metzler, S Rosebrock, editors), Lond. Math. Soc. Lect. Note Ser. 446, Cambridge Univ. Press (2017) 1–26 MR Zbl

[12] **C Hog-Angeloni**, **W Metzler**, *The Andrews–Curtis conjecture and its generalizations*, from "Two-dimensional homotopy and combinatorial group theory" (C Hog-Angeloni, W Metzler, A J Sieradski, editors), Lond. Math. Soc. Lect. Note Ser. 197, Cambridge Univ. Press (1993) 365–380 MR Zbl

[13] **C Hog-Angeloni**, **W Metzler**, *Geometric aspects of two-dimensional complexes*, from "Two-dimensional homotopy and combinatorial group theory" (C Hog-Angeloni, W Metzler, A J Sieradski, editors), Lond. Math. Soc. Lect. Note Ser. 197, Cambridge Univ. Press (1993) 1–50 MR Zbl

[14] **C Hog-Angeloni**, **W Metzler**, *Further results concerning the Andrews–Curtis conjecture and its generalizations*, from "Advances in two-dimensional homotopy and combinatorial group theory" (W Metzler, S Rosebrock, editors), Lond. Math. Soc. Lect. Note Ser. 446, Cambridge Univ. Press (2018) 27–35 MR Zbl

[15] **M P Latiolais**, *Homotopy and homology classification of* 2*-complexes*, from "Two-dimensional homotopy and combinatorial group theory" (C Hog-Angeloni, W Metzler, A J Sieradski, editors), Lond. Math. Soc. Lect. Note Ser. 197, Cambridge Univ. Press (1993) 97–124 MR Zbl

[16] **M Lustig**, *Nielsen equivalence and simple-homotopy type*, Proc. Lond. Math. Soc. 62 (1991) 537–562 MR Zbl

[17] **W H Mannan**, **T Popiel**, *An exotic presentation of Q_{28}*, Algebr. Geom. Topol. 21 (2021) 2065–2084 MR Zbl

[18] **J McCool**, **A Pietrowski**, *On a conjecture of W Magnus*, from "Word problems: decision problems and the Burnside problem in group theory", Stud. Logic Found. Math. 71, North-Holland, Amsterdam (1973) 453–456 MR Zbl

[19] **W Metzler**, *Über den Homotopietyp zweidimensionaler CW-Komplexe und Elementartransformationen bei Darstellungen von Gruppen durch Erzeugende und definierende Relationen*, J. Reine Angew. Math. 285 (1976) 7–23 MR Zbl

[20] **A D Myasnikov**, **A G Myasnikov**, **V Shpilrain**, *On the Andrews–Curtis equivalence*, from "Combinatorial and geometric group theory", Contemp. Math. 296, Amer. Math. Soc., Providence, RI (2002) 183–198 MR Zbl

[21] **J Nicholson**, *Stably free modules and the unstable classification of* 2*-complexes*, preprint (2021) arXiv 2108.02220

[22] **A J Sieradski**, *Algebraic topology for two-dimensional complexes*, from "Two-dimensional homotopy and combinatorial group theory" (C Hog-Angeloni, W Metzler, A J Sieradski, editors), Lond. Math. Soc. Lect. Note Ser. 197, Cambridge Univ. Press (1993) 51–96 MR Zbl

[23] **R G Swan**, *Projective modules over Laurent polynomial rings*, Trans. Amer. Math. Soc. 237 (1978) 111–120 MR Zbl

[24] **H Zieschang**, *Über die Nielsensche Kürzungsmethode in freien Produkten mit Amalgam*, Invent. Math. 10 (1970) 4–37 MR Zbl

Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires
Buenos Aires, Argentina

Instituto de Investigaciones Matemáticas Luis A Santaló (IMAS), CONICET-Universidad de Buenos Aires
Buenos Aires, Argentina

jbarmak@dm.uba.ar

Received: 24 July 2023 Revised: 5 December 2023

Algebraic & Geometric Topology **25**:1 (2025) 357–397
DOI: 10.2140/agt.2025.25.357
Published: 24 March 2025

Generalizing quasicategories via model structures on simplicial sets

MATT FELLER

We use Cisinski's machinery to construct and study model structures on the category of simplicial sets whose classes of fibrant objects generalize quasicategories. We identify a lifting condition that captures the homotopical behavior of quasicategories without the algebraic aspects and show that there is a model structure whose fibrant objects are precisely those that satisfy this condition. We also identify a localization of this model structure whose fibrant objects satisfy a "special horn lifting" property similar to the one satisfied by quasicategories. This special horn model structure leads to a conjectural characterization of the bijective-on-0-simplices trivial cofibrations of the Joyal model structure. We also discuss how these model structures all relate to one another and to the minimal model structure.

18N50, 55U35; 18N60

1 Introduction

The theory of quasicategories has proven to be a powerful tool across many areas of mathematics, including algebraic geometry, topology, and beyond. The basic idea of a quasicategory is that it is a simplicial set that behaves like a category "up to homotopy". This paper explores how one can generalize this idea, where we have simplicial sets modeling up-to-homotopy versions of structures that are weaker than categories. Our motivating example of such a structure weaker than categories is the 2-Segal sets of Dyckerhoff and Kapranov [5] and Gálvez-Carrillo, Kock and Tonks [8]. In a follow-up paper, we define "quasi-2-Segal sets" that are up-to-homotopy versions of 2-Segal sets, building on the groundwork laid here.

A foundational result of quasicategory theory is the existence of a corresponding model structure on the category of simplicial sets, called the Joyal model structure [14]. A desirable quality of any generalization of quasicategories is therefore the existence of a similar associated model structure. Taking this idea to heart, one could say that within simplicial sets the search for robust generalizations of quasicategories is equivalent to the search for model structures. Hence, the aim of this paper is to dive into the sea of possible model structures and retrieve a few with properties that should prove useful for further study.

1.1 Model structures on simplicial sets

The two most prominent model structures on sSet, the category of simplicial sets, are the Kan–Quillen model structure [15] and the Joyal model structure [14]. In both model structures, all objects are cofibrant, so the well-behaved objects are precisely the fibrant objects. In the Kan–Quillen model structure, the fibrant objects are the Kan complexes which provide a model of spaces/∞-groupoids, and the fibrant objects of the Joyal model structure are the quasicategories which give us a model of $(\infty, 1)$-categories. These model structures are both examples of *Cisinski model structures* on sSet, meaning that they are cofibrantly generated and their cofibrations are precisely the monomorphisms. The Kan–Quillen model structure is a localization of the Joyal model structure in the sense that it has the same cofibrations and its class of fibrant objects (Kan complexes) is contained in the class of Joyal fibrant objects (quasicategories). In general, the process of localizing a model structure to another with the same cofibrations and fewer fibrant objects is well understood; see Hirschhorn [10]. There are other localizations of the Joyal model structure in the literature; see for example Campbell and Lanari [2] and Cisinski [3, Chapter 9].

By starting with the Joyal model structure and localizing, one ends up with fibrant objects that are quasicategories with extra structure. If we instead want to do the opposite and generalize the notion of quasicategory, then we want to "delocalize". The goal of this paper is to lay the groundwork for constructing such delocalizations of the Joyal model structure. Our approach is to focus on the homotopical aspects of quasicategories, constructing various model structures that maintain those aspects but lack a notion of composition. In particular, for morphisms f and g in a quasicategory Q, we consider a homotopy from f to g to be given by a 2-simplex

$$\begin{tikzcd} x \arrow[rr, "f"] \arrow[dr, dashed, "s_0 x"'] & & y \\ & x \arrow[ur, "g"'] & \end{tikzcd}$$

with degenerate edge $0 \to 1$ as indicated. We say that Q is *homotopically behaved*, in the sense that all of the higher invertibility and compositionality we would expect from a good notion of homotopy are satisfied by the 2-simplices of this form, as well as by higher n-simplices whose edge $i \to i+1$ for some $0 \le i \le n-1$ is degenerate. The purpose of this paper is to study Cisinski model structures on sSet whose fibrant objects are homotopically behaved, which we call *homotopically behaved model structures*. Our main result is to construct and describe the homotopically behaved model structure with the smallest possible class of weak equivalences.

Theorem A *There exists a minimal homotopically behaved model structure on* sSet*. The fibrant objects in this model structure are the simplicial sets with lifts of certain modified horn inclusions, which we call* ***J-augmented horn inclusions****.*

We state this theorem in more detail as Theorem 4.33. The terms *homotopically behaved* and *J-augmented horn inclusion* are defined explicitly in Section 1.4.

1.2 2-Segal motivation

Our motivation for considering delocalizations of the Joyal model structure is to construct a "quasi-2-Segal set" model structure, where the fibrant objects satisfy an up-to-homotopy version of the 2-Segal condition introduced in [5] and [8]. Recall that the (strict) Segal (or "1-Segal") condition encodes unique composition; the simplicial sets satisfying this condition (the "1-Segal sets") are equivalent to categories. The simplicial sets satisfying a weakened, up-to-homotopy version of the 1-Segal condition are the quasicategories. The 2-Segal condition is generalization of the 1-Segal condition that encodes partially defined, not necessarily unique composition which is still associative in a particular sense. It is therefore natural to try to extend this generalization from 1-Segal to 2-Segal to the up-to-homotopy setting, ie to look for a robust definition of "quasi-2-Segal sets". A compelling justification for a particular definition would be the existence of a model structure analogous to the Joyal model structure. Since the quasi-2-Segal sets should generalize quasicategories, such a model structure should be a delocalization of the Joyal model structure. In follow-up work [7] we construct such a quasi-2-Segal model structure by localizing our minimal homotopically behaved model structure with respect to maps that encode the 2-Segal condition.

1.3 Cisinski's theory and the minimal model structure

Proving the existence of a model structure from scratch is generally cumbersome and highly technical, but fortunately for our particular situation Cisinski's theory provides a powerful framework for building model structures that requires checking a much more manageable set of conditions. One aspect of this theory is the existence of a minimal model structure, whose class of fibrant objects contains the fibrant objects of every other Cisinski model structure.[1] Therefore, one approach to delocalizing the Joyal model structure is to delocalize all the way back to the minimal model structure, and then localize from there.

As we see in a companion paper [6], we lose a lot by delocalizing all the way down to the minimal model structure. In particular, the main result of that paper is a new characterization of the fibrant objects in the minimal model structure. What we find is that the notion of "homotopy" familiar from quasicategories does not behave well in the fibrant objects of this model structure. In particular, the existence of a homotopy from f to g given by a 2-simplex with degenerate edge need not imply the

[1] One could also call this "maximal", but we prefer "minimal" since the class of weak equivalences is as small as possible, and the fibrant objects have the least structure.

existence a homotopy from g to f. Homotopies of this form also need not compose. One can devise an alternative notion of homotopy which behaves well in the minimal model structure, essentially by demanding that all higher invertibility data be present from the start, but then one sacrifices the simplicity that comes from a homotopy being embodied by a single simplex.

1.4 Homotopically behaved model structures and augmented horns

We denote by $\Delta[n]^*_{i\to i+1}$ the standard n-simplex with the edge $i \to (i+1)$ collapsed to a degeneracy. The idea that an n-simplex with degenerate $i \to (i+1)$ edge is a homotopy of $(n-1)$-simplices can be expressed by the surjective map $\Delta[n]^*_{i\to i+1} \to \Delta[n-1]$ being a weak equivalence. Thus, we introduce the terminology *homotopically behaved* for Cisinski model structures on sSet where each of these surjective maps $\Delta[n]^*_{i\to i+1} \to \Delta[n-1]$ is a weak equivalence. In Section 4 we construct the minimal homotopically behaved model structure, whose fibrant objects have the least possible structure while maintaining the desirable homotopical aspects of quasicategories. Localizing this model structure with respect to the maps from the 2-Segal condition yields a "quasi-2-Segal set" model structure whose fibrant objects must then also have the desirable homotopical aspects of quasicategories.

At the same time, we construct a nontrivial localization of this model structure at $K \to *$, where K is the simplicial set one gets by gluing in a left and right inverse to $\Delta[1]$; see Example 4.5. Although this model structure does not appear to be directly useful for defining quasi-2-Segal sets, it may be of independent interest in understanding the broader picture of model structures on simplicial sets.

The key inspiration for our approach comes from the special outer horn extension property of quasicategories. Recall that a *horn* $\Lambda^i[n]$ is the union of all of the faces of the n-simplex $\Delta[n]$ except for $d_i\Delta[n]$, which we say is *inner* if $0 < i < n$ and is *outer* if $i = 0$ or $i = n$. We refer to these horns as *ordinary horns* to distinguish them from the augmented horns we introduce below. A *quasicategory* is a simplicial set Q such that every inner horn $\Lambda^i[n] \to Q$ extends to an n-simplex $\Delta[n] \to Q$. A quasicategory need not have extensions of an outer horn such as $\Lambda^0[n] \to Q$. However, if the edge $0 \to 1$ is sent to an edge in Q that is invertible in a certain sense, then we say $\Lambda^0[n] \to Q$ is an example of a *special outer horn*, and it turns out that quasicategories do have extensions of all special outer horns.

In a general homotopically behaved model structure, the fibrant objects need not have extensions of ordinary horns. However, the central idea of our approach is to create augmented horns, where we glue a simplicial set onto a particular edge to "invert" it. In Section 3, we see how certain augmented horn inclusions are forced to be weak equivalences in a homotopically behaved model structure. Furthermore, we can characterize the fibrant objects in the minimal homotopically behaved model structure in terms of lifts of J-augmented horn inclusions, as we stated in Theorem A. We denote by J the nerve of the free-living isomorphism, so and define *J-augmented horn inclusions* to be ordinary horn inclusions $\Lambda^j[n] \hookrightarrow \Delta[n]$ with a copy of J glued in along either the $(j-1) \to j$ edge or the $j \to (j+1)$ edge. We see in Corollary 4.28 that the minimal homotopically behaved model structure localizes to the Joyal model structure.

1.5 Special horns

The notion of invertibility in Section 3 comes from attaching a simplicial set I that is contractible in the sense that the map $I \to *$ is a weak equivalence, with our main examples being $I = J$ and $I = K$. In Section 5, we see that this notion does not account for all of the edges we want to consider invertible in the context of quasicategories, and identify a separate class of augmented horn inclusions which we call the *special horn inclusions*. We show that there is a model structure whose fibrant objects are precisely the simplicial sets with special horn inclusions.

Recall that we deemed the localization of the minimal homotopically behaved model structure at $K \to *$ to be unsuitable for our ultimate purposes in defining quasi-2-Segal sets. This special horn model structure is a further localization, and hence is also not suitable. However, this model structure may be of interest with regards to studying the Joyal model structure. In particular, we conjecture that the special horn inclusions, together with the inner horn inclusions, generate the class of bijective-on-0-simplices trivial cofibrations in the Joyal model structure; see Conjecture 5.24.

1.6 Pointwise cylinders

A central element of Cisinski's theory is the *exact cylinder*, which is a functorial choice of simplicial set $E \otimes X$ for each simplicial set X, satisfying certain axioms. Many applications of Cisinski's work use an exact cylinder given by the Cartesian product $E \otimes X = I \times X$ for some simplicial set I with distinct vertices $\{0\} \hookrightarrow I$ and $\{1\} \hookrightarrow I$. For our purposes, the necessary proofs are greatly simplified by instead using an alternative kind of exact cylinder, which we call a *pointwise cylinder*. We introduce pointwise cylinders in Section 4.1. In Section 4.2, we see that the minimal homotopically behaved model structure is also minimal with respect to pointwise cylinders, in the sense that we construct it using a pointwise cylinder in Cisinski's machinery and any other such constructed model structure is a localization of it.

1.7 Organization

In Section 2, we cover basic definitions and notation, and then summarize Cisinski's theory. In Section 3, we define homotopically behaved model structures and augmented horn inclusions. In Section 4 we show that there is a minimal homotopically behaved model structure and that its fibrant objects are the simplicial sets with extensions of certain augmented horns. In Section 5, we define *special horns* as a separate kind of augmented horn, and use Cisinski's machinery to show that there exists a model structure whose fibrant objects are simplicial sets with special horn extensions. In Section 6 we summarize and compare the various model structures constructed in this work.

Acknowledgements

I would like to thank Julie Bergner for her extraordinarily helpful feedback.

The author was partially supported by NSF RTG grant DMS-1839968 and NSF grant DMS-1906281.

2 Background

We recall some basic notions, as well as the necessary aspects of Cisinski's theory.

2.1 Basics of simplicial sets and model structures

Let Δ denote the category whose objects are the finite nonempty ordered sets $[n] = \{0 \leq 1 \leq \cdots \leq n\}$ for $n \geq 0$ and whose morphisms are order-preserving maps. We write $d^i : [n] \to [n+1]$ and $s^i : [n+1] \to [n]$ for the coface and codegeneracy maps, respectively, which generate the morphisms of Δ. A *simplicial set* is a functor $\Delta^{\mathrm{op}} \to \mathrm{Set}$. We denote the category of simplicial sets by sSet and the representable simplicial sets by $\Delta[n]$, except that we often denote $\Delta[0]$ instead by $*$ since it is the terminal object in sSet. Our notation for the i^{th} horn of $\Delta[n]$ is $\Lambda^i[n]$. For more background on simplicial sets, see [9].

We write $f \boxslash g$ if g has the right lifting property with respect to f. The class of morphisms with the right lifting property with respect to a set of maps $\mathscr{A}$ is denoted by $\mathscr{A}^{\boxslash}$, and the class of morphisms f such that $f \boxslash \mathscr{B}$ is denoted by ${}^{\boxslash}\mathscr{B}$. Given a set S of morphisms, the class ${}^{\boxslash}(S^{\boxslash})$ is the closure of S under taking pushouts, transfinite compositions, and retracts. For this reason, we sometimes say that S *generates* the class ${}^{\boxslash}(S^{\boxslash})$.

We restrict our focus to *Cisinski* model structures on sSet, which are cofibrantly generated model structures whose cofibrations are precisely the monomorphisms. A model structure is *cofibrantly generated* if there are sets $\mathscr{I}$ and $\mathscr{J}$ such that $\mathscr{I}$ generates the cofibrations and $\mathscr{J}$ generates the trivial cofibrations. For more background on model categories, see [10] or [11].

2.2 Cisinski's theory

The main result we use from Cisinski is Theorem 2.10, which says that if a set of monomorphisms Λ satisfies certain properties, then Λ characterizes the fibrant objects of some model category via a lifting property. The purpose of this subsection is to recall the background necessary to state this result precisely.

We begin by recalling the notion of a cylinder.

Definition 2.1 [4, Definition 2.4.6] A *cylinder* of a simplicial set X is a factorization

$$X \sqcup X \xhookrightarrow{(\partial_0,\partial_1)} I \otimes X \to X$$

of the canonical fold map $(\mathrm{id}_X, \mathrm{id}_X)$, where the first map is a monomorphism. The maps ∂_ε for $\varepsilon = 0, 1$ pick out copies of X that do not intersect inside of $I \otimes X$.

Beware that the notation $I \otimes X$ in the above definition is purely formal. The simplicial set $I \otimes X$ need not be a monoidal product or tensor of any kind.

Remark 2.2 In an arbitrary model category, we define cylinders similarly, where the first map is required to be a cofibration. In the context of Cisinski's theory, all of the model structures we consider have as cofibrations precisely the class of monomorphisms, justifying this definition.

A *functorial cylinder* is a compatible choice of cylinder for every simplicial set. To make this definition rigorous, we first notice that $X \mapsto X \sqcup X$ and $X \mapsto X$ are endofunctors of sSet, which we denote by $1 \sqcup 1$ and 1, respectively. There is a natural transformation $(\mathrm{id}, \mathrm{id})\colon 1 \sqcup 1 \Rightarrow 1$ whose component at each X is the canonical fold map $(\mathrm{id}_X, \mathrm{id}_X)$.

Definition 2.3 [4, Definition 2.4.8] A *functorial cylinder* is a factorization

$$1 \sqcup 1 \xRightarrow{(\partial_0, \partial_1)} I \otimes - \Rightarrow 1$$

of the natural transformation $(\mathrm{id}, \mathrm{id})$, where each component of the natural transformation

$$(\partial_0, \partial_1)\colon 1 \sqcup 1 \Rightarrow I \otimes -$$

is a monomorphism.

The motivation behind these definitions is to generalize the idea (from the Kan–Quillen model structure on sSet) of $\Delta[1] \times X$ being a cylinder of a simplicial set X, in the sense that a map $X \times \Delta[1] \to Y$ gives a homotopy of maps $X \to Y$. Imposing some further conditions helps maintain the spirit of the original setting, where we imagine $I \otimes X$ as something like a stretched out copy of X.

Definition 2.4 [4, Definition 2.4.8] An *exact cylinder*[2] is a functorial cylinder satisfying the following axioms.

(DH1) The functor $I \otimes -$ commutes with small colimits and preserves monomorphisms.

(DH2) For any monomorphism of simplicial sets $j : A \hookrightarrow B$, the square

$$\begin{tikzcd} A \arrow[r, hook, "j"] \arrow[d, "(\partial_\varepsilon)_A"'] & B \arrow[d, "(\partial_\varepsilon)_B"] \\ I \otimes A \arrow[r, "I \otimes j"'] & I \otimes B \end{tikzcd}$$

is a pullback for each $\varepsilon = 0, 1$.

Remark 2.5 Assuming $I \otimes -$ preserves monomorphisms (as (DH1) calls for), all of the morphisms in the square in (DH2) are monomorphisms. We can interpret the condition that the square be a pullback as saying that the intersection of $I \otimes A \subseteq I \otimes B$ with $B \subseteq I \otimes B$ is precisely A.

Example 2.6 Let I be a cylinder of the terminal simplicial set $\Delta[0]$. In other words, choose a monomorphism of simplicial sets $\Delta[0] \sqcup \Delta[0] \hookrightarrow I$. The functor defined by $I \otimes X = I \times X$ determines an exact cylinder. In particular, when $I = \Delta[1]$, we recover the familiar notion of cylinder from the Kan–Quillen model structure.

Many applications of this theory involve cylinders defined by Cartesian product, but our approach uses a new kind of exact cylinder which we introduce in Section 4.1.

[2] The origin of this definition is [3], where the term is "*donnée homotopique élémentaire*", hence "DH".

Let us follow [4, Remark 2.4.9] and establish some more notation. Given an exact cylinder $I \otimes -$, by taking the pushout of the left and top maps in the square for axiom (DH2), we get the inclusion map

$$(I \otimes A) \cup B \hookrightarrow I \otimes B.$$

To emphasize the dependence of this inclusion on whether $\varepsilon = 0$ or 1 in the inclusion $\partial_\varepsilon \colon \Delta[0] \hookrightarrow I$, we rewrite this map as

$$(I \otimes A) \cup (\{\varepsilon\} \otimes B) \hookrightarrow I \otimes B,$$

thinking of $\{0\}$ and $\{1\}$ as endpoints of $I = I \otimes \Delta[0]$.

In this spirit, we also write ∂I for the union of $\{0\}$ and $\{1\}$ inside of I, and we write $\partial I \otimes X$ as the union of $\{0\} \otimes X$ and $\{1\} \otimes X$ in $I \otimes X$. Then we have a canonical inclusion

$$(I \otimes A) \cup (\partial I \otimes B) \hookrightarrow I \otimes B,$$

arising from a diagram akin to the square in axiom (DH2).

We are now ready for one of the key definitions we need from Cisinski.

Definition 2.7 [4, Definition 2.4.11] Given an exact cylinder $I \otimes -$, we say that a class of morphisms ${}^{\boxslash}(\Lambda^{\boxslash})$ generated by a set Λ of monomorphisms is an $(I \otimes -)$-*anodyne class* if the following conditions hold.

(An1) For each monomorphism of simplicial sets $X \hookrightarrow Y$ and $\varepsilon = 0, 1$, the induced map

$$(I \otimes X) \cup (\{\varepsilon\} \otimes Y) \hookrightarrow I \otimes Y$$

is in ${}^{\boxslash}(\Lambda^{\boxslash})$.

(An2) For each $A \hookrightarrow B$ in ${}^{\boxslash}(\Lambda^{\boxslash})$, the induced map $(I \otimes A) \cup (\partial I \otimes B) \to I \otimes B$ is also in ${}^{\boxslash}(\Lambda^{\boxslash})$.

We can restate each of axioms (An1) and (An2) in a form that is easier to check.

Lemma 2.8 *Let $I \otimes -$ be an exact cylinder and let Λ be a set of monomorphisms. Then axiom* (An1) *is equivalent to* (An1′) *below and axiom* (An2) *is equivalent to* (An2′) *below*:

(An1′) *For each $n \geq 0$ and $\varepsilon = 0, 1$, the map $(I \otimes \partial\Delta[n]) \cup (\{\varepsilon\} \otimes \Delta[n]) \hookrightarrow I \otimes \Delta[n]$ induced by $\partial\Delta[n] \hookrightarrow \Delta[n]$ is in ${}^{\boxslash}(\Lambda^{\boxslash})$.*

(An2′) *For each $A \hookrightarrow B$ in Λ, the induced map $(I \otimes A) \cup (\partial I \otimes B) \to I \otimes B$ is in ${}^{\boxslash}(\Lambda^{\boxslash})$.*

Proof The equivalence (An1′) $\iff$ (An1) follows from the equality of classes

$$\{(I \otimes \partial\Delta[n]) \cup (\{\varepsilon\} \otimes \Delta[n]) \hookrightarrow I \otimes \Delta[n] \mid n \geq 0,\ \varepsilon = 0, 1\}^{\boxslash}$$
$$= \{(I \otimes X) \cup (\{\varepsilon\} \otimes Y) \hookrightarrow I \otimes Y \mid X \hookrightarrow Y \text{ in } \mathrm{sSet},\ \varepsilon = 0, 1\}^{\boxslash},$$

which is a consequence of correspondence (2.4.13.4) of Example 2.4.13 in [4], which ultimately relies on the fact that the boundary inclusions generate the class of monomorphisms. The equivalence of (An2′) and (An2) follows from a similar argument, replacing $\{\varepsilon\}$ with ∂I and arbitrary monomorphisms $X \hookrightarrow Y$ with maps in ${}^{\boxslash}(\Lambda^{\boxslash})$. □

Definition 2.9 Given an exact cylinder $I \otimes -$ and a morphism of simplicial sets $f_0, f_1 \colon A \to X$, we say that an *I-homotopy* from f_0 to f_1 is a map $h \colon I \otimes A \to X$ such that precomposing h with $\{\varepsilon\} \otimes A \hookrightarrow I \otimes A$ yields f_ε for $\varepsilon = 0, 1$. We say that $f, g \colon A \to X$ are *I-homotopic* if there is a finite zigzag of I-homotopies from f to g. Suppressing the dependence on I, we let $[A, X]$ denote the quotient of the set $\mathrm{Hom}(A, X)$ by identifying maps that are I-homotopic.

Theorem 2.10 [4, Theorem 2.4.19] *Given an exact cylinder $I \otimes -$ and a set of monomorphisms Λ such that ${}^{\boxslash}(\Lambda^{\boxslash})$ is an $(I \otimes -)$-anodyne class, there is a cofibrantly generated model structure on* sSet *whose cofibrations are the monomorphisms and whose fibrant objects are the simplicial sets with the right lifting property with respect to Λ. The weak equivalences in this model structure are maps $X \to Y$ such that for every fibrant W the induced map $[Y, W] \to [X, W]$ is a bijection.*

Remark 2.11 It is a theorem of Joyal that if two model structures share the same cofibrations and fibrant objects, then they are the same model structure; see [14, Proposition E.1.10]. Therefore, the weak equivalences described in the above theorem are determined once we know our cofibrations are the monomorphisms and our fibrant objects are those with lifts against Λ. Throughout this paper, we shall implicitly use this fact from Joyal to conclude that when the class of fibrant objects of a Cisinski model structure is contained in the class of fibrant objects in another, the class of weak equivalences of the former model contains the class of weak equivalences of the latter.

3 Homotopically behaved model structures and augmented horns

In this section, we define *homotopically behaved* model structures to be Cisinski model structures where the retract map from an n-simplex with $i \to (i+1)$ edge collapsed to a degeneracy onto the $(n-1)$-simplex is a weak equivalence. This condition captures the idea that a map out of an n-simplex with degenerate $i \to (i+1)$ edge is a homotopy of $(n-1)$-simplices. We show that this condition is equivalent to the condition that certain modified horn inclusions are weak equivalences.

3.1 Augmented horn extensions

In a quasicategory, we can view higher homotopies as simplices with a degenerate edge $i \to i+1$. More precisely, for $n \geq 1$, a homotopy between n-simplices x, y consists of an $(n+1)$-simplex H with some $0 \leq i \leq n$ such that the edge $i \to (i+1)$ is degenerate and $\{d_i H, d_{i+1} H\} = \{x, y\}$.

Before continuing this discussion, let us fix some notation.

Definition 3.1 Given $n \geq 2$, $0 \leq i \leq n-1$, a subcomplex $A \subseteq \Delta[n]$ containing the $i \to (i+1)$ edge of $\Delta[n]$, and a map $\Delta[1] \to X$, let $A^X_{i \to i+1}$ be the pushout

$$\begin{tikzcd} \Delta[1] \arrow[r] \arrow[d, "i \to i+1"'] & X \arrow[d] \\ A \arrow[r] & A^X_{i \to i+1} \end{tikzcd}$$

Example 3.2 To get $\Lambda^1[2]^X_{1 \to 2}$, we attach $\Delta[1] \to X$ along the edge $1 \to 2$, as in the following diagram:

$$\begin{tikzcd} & 1 \arrow[dr, "X"] & \\ 0 \arrow[ur] & & 2 \end{tikzcd}$$

Our main use for these $A^X_{i \to i+1}$ is to discuss modified versions of the ordinary horn inclusions. Given a horn inclusion $\Lambda^j[n] \hookrightarrow \Delta[n]$, the horn $\Lambda^j[n]$ contains the edges $(j-1) \to j$ and $j \to (j+1)$ (excluding the cases $-1 \to 0$ and $n \to n+1$ as there are no such edges). For each $\Delta[1] \to X$ and $i = j-1, j$, we get an induced inclusion $\Lambda^j[n]^X_{i \to i+1} \hookrightarrow \Delta[n]^X_{i \to i+1}$. The modified horn inclusions we study are two special cases of this situation. The first case is when $X = *$, so the induced inclusion $\Lambda^j[n]^*_{i \to i+1} \hookrightarrow \Delta[n]^*_{i \to i+1}$ is the ordinary horn inclusion with the $i \to i+1$ edge collapsed to a degeneracy. The second case is when $\Delta[1] \to X$ is an inclusion, so that the induced map is the ordinary horn inclusion with a copy of X glued in along the $i \to i+1$ edge. Let us set some terminology for these special cases.

Definition 3.3 Fix $n \geq 2$ and $0 \leq i \leq n-1$.

(1) When $X = *$, the terminal simplicial set, we say that $\Lambda^j[n]^*_{i \to i+1}$ is a *pinched horn* for $j = i, i+1$, that $\Delta[n]^*_{i \to i+1}$ is a *pinched n-simplex*, and that the inclusion $\Lambda^j[n]^*_{i \to i+1} \hookrightarrow \Delta[n]^*_{i \to i+1}$ for $j = i, i+1$ is a *pinched horn inclusion*.

(2) When $\Delta[1] \to X$ is an inclusion, we say that $\Lambda^j[n]^X_{i \to i+1}$ is an *X-augmented horn* for $j = i, i+1$, that $\Delta[n]^X_{i \to i+1}$ is an *X-augmented n-simplex*, and that the inclusion $\Lambda^j[n]^X_{i \to i+1} \hookrightarrow \Delta[n]^X_{i \to i+1}$ for $j = i, i+1$ is an *X-augmented horn inclusion*.

Remark 3.4 The notation $A^X_{i \to i+1}$ and terminology "X-augmented" are technically ambiguous because they depend on the map $\Delta[1] \to X$, but the choice of map should be clear from context in all instances in this paper.

Since we think of maps out of a pinched $(n+1)$-simplex $\Delta[n+1]^*_{i \to i+1}$ as homotopies of n-simplices, we can interpret this situation as saying that maps out of a pinched $(n+1)$-simplex $\Delta[n+1]^*_{i \to i+1}$ are equivalent to maps out of the standard n simplex $\Delta[n]$ up to homotopy. This interpretation is encoded more precisely by the surjective map $\Delta[n+1]^*_{i \to i+1} \to \Delta[n]$ being a weak equivalence in the Joyal model structure. As our goal is to construct model structures with fibrant objects that have the same homotopical behavior as quasicategories, a natural starting place is to declare each of these surjective maps $\Delta[n+1]^*_{i\to i+1} \to \Delta[n]$ for $n \geq 1$ and $0 \leq i \leq n$ to be weak equivalences in our model structure. The aim of this subsection is to use 2-out-of-3 arguments to identify other maps that are forced to be weak equivalences in this situation.

Definition 3.5 We say that a Cisinski model structure on the category of simplicial sets is a *homotopically behaved model structure* if the surjective maps $\Delta[m+1]^*_{k\to k+1} \to \Delta[m]$ are weak equivalences for all $m \geq 1$ and $0 \leq k \leq m$.

Our first step is to notice that these maps have sections, which must also be weak equivalences by the 2-out-of-3 property.

Lemma 3.6 *The surjective maps $\Delta[m+1]^*_{k\to k+1} \to \Delta[m]$ are weak equivalences for all $m \geq 1$ and $0 \leq k \leq m$ if and only if the sections $d^k, d^{k+1} \colon \Delta[m] \hookrightarrow \Delta[m+1]^*_{k\to k+1}$ are as well.*

Our next step is to show that pinched horn inclusions are forced to be weak equivalences. Our proof requires first defining *generalized pinched horns*. Given $n \geq 2$, $0 \leq i \leq n-1$, and a subset $S \subseteq \{0,\dots,n\}$ such that $|S| \geq 2$ and exactly one of $i, i+1$ is in S, the generalized horn $\Lambda^S[n] \subseteq \Delta[n]$ is the union of all d_j faces of $\Delta[n]$ for j in S. There exists an $\ell \in S$ not equal to i or $i+1$ and so the d_ℓ face of $\Delta[n]$ contains the $i \to (i+1)$ edge, and therefore $\Lambda^S[n]$ contains the $i \to (i+1)$ edge of $\Delta[n]$, allowing us to apply Definition 3.3.

Definition 3.7 Given $n \geq 2$, $0 \leq i \leq n-1$, and a subset $S \subseteq \{0,\dots,n\}$ such that $|S| \geq 2$ and exactly one of $i, i+1$ is in S, we say that $\Lambda^S[n]^*_{i\to i+1}$ is a *generalized pinched horn* and that $\Lambda^S[n]^*_{i\to i+1} \hookrightarrow \Delta[n]^*_{i\to i+1}$ is a *generalized pinched horn inclusion*. Similarly, given X and an inclusion $\Delta[1] \hookrightarrow X$, we say that $\Lambda^S[n]^X_{i\to i+1}$ is a *generalized X-augmented horn* and that $\Lambda^S[n]^X_{i\to i+1} \hookrightarrow \Delta[n]^*_{i\to i+1}$ is a *generalized X-augmented horn inclusion*.

In addition to using the 2-out-of-3 property, the proofs of the upcoming propositions rely on the fact that every Cisinski model structure is left proper by [10, Proposition 13.1.2]. We state this standard fact as a lemma.

Lemma 3.8 *In a Cisinski model structure, pushouts along inclusions preserve weak equivalences.*

Proposition 3.9 *A Cisinski model structure is homotopically behaved if and only if every generalized pinched horn inclusion is a weak equivalence.*

Proof We first prove the forward implication. By Lemma 3.6, we know that the composite of

$$\Delta[n-1] \hookrightarrow \Lambda^S[n]^*_{i\to i+1} \hookrightarrow \Delta[n]^*_{i\to i+1}$$

is a weak equivalence, so it suffices to show that the map on the left is a weak equivalence by the 2-out-of-3 property. This map on the left is the inclusion of either the d^i or d^{i+1} face, whichever is in S.

For $n = 2$, the map on the left is actually an isomorphism since $\Lambda^S[2]$ is the union of two 1-simplices, one of which gets collapsed to a point to create $\Lambda^S[2]^*_{i\to i+1}$.

For $n \geq 3$, we proceed by induction on $|S|$. For the base case $|S| = 2$, the inclusion $\Delta[n-1] \hookrightarrow \Lambda^S[n]^*_{i\to i+1}$ amounts to gluing in a copy of $\Delta[n-1]^*_{i'\to i'+1}$ along one of the faces of $\Delta[n-1]$. In other words, it is a pushout of the weak equivalence $\Delta[n-2] \hookrightarrow \Delta[n-1]^*_{i'\to i'+1}$ along an inclusion, and so is a weak equivalence.

Now if $|S| \geq 3$, we pick some $j \in S$ not equal to i or $i+1$ and let $S' = S \smallsetminus \{j\}$. By induction we know that the inclusion $\Delta[n] \hookrightarrow \Lambda^{S'}[n+1]^*_{i\to i+1}$ is a weak equivalence, so it suffices to show that the inclusion $\Lambda^{S'}[n+1]^*_{i\to i+1} \hookrightarrow \Lambda^S[n+1]^*_{i\to i+1}$ is as well. But this latter inclusion is itself a pushout of a pinched generalized horn whose subset of indices is of size one less than $|S|$, and so is a weak equivalence.

We now turn to the reverse implication, so let us assume that every generalized pinched horn inclusion is a weak equivalence. In particular, we can pick S such that $|S| = 2$, and consider the composite

$$\Delta[n-1] \hookrightarrow \Lambda^S[n]^*_{i\to i+1} \hookrightarrow \Delta[n]^*_{i\to i+1}.$$

By Lemma 3.6, it suffices to show that every such composite map is a weak equivalence, and therefore (by the 2-out-of-3 property) to show that the map on the left is a weak equivalence. We proceed by induction on n. As we saw in the proof of the forward implication, in the base case $n = 2$ the map on the left is an isomorphism. For $n \geq 3$, the map on the left is a pushout of $\Delta[n-2] \hookrightarrow \Delta[n-1]^*_{i\to i+1}$. □

We can now go one step further and show that certain generalized augmented horn inclusions must also be weak equivalences in a homotopically behaved model structure.

Proposition 3.10 *Given a simplicial set I and an inclusion $\Delta[1] \hookrightarrow I$, if the map $I \to *$ is a weak equivalence in a given Cisinski model structure, then the model structure is homotopically behaved if and only if every generalized I-augmented horn inclusion is a weak equivalence in that model structure.*

Proof In the diagram

$$\begin{tikzcd}
I \arrow[r, "\sim"] \arrow[d, hook] & * \arrow[d, hook] \\
\Lambda^S[n]^I_{i\to i+1} \arrow[r, "\sim"] \arrow[d, hook] & \Lambda^S[n]^*_{i\to i+1} \arrow[d, hook] \\
\Delta[n]^I_{i\to i+1} \arrow[r, "\sim"] & \Delta[n]^*_{i\to i+1}
\end{tikzcd}$$

the horizontal maps are all weak equivalences by Lemma 3.8. The bottom-left vertical map is a weak equivalence if and only if the bottom-right vertical map is by the 2-out-of-3 property, which is a weak equivalence if and only if the model structure is homotopically behaved by Proposition 3.9. □

Let us recall the definition of J, the key example of a simplicial set that is weakly equivalent to $*$ in every Cisinski model structure on sSet.

Definition 3.11 Let $\mathbb{I}$ be the category with two objects and precisely one morphism in every hom-set, sometimes called the *free-living isomorphism*. We denote the nerve of the free-living isomorphism by $J = N(\mathbb{I})$.

Corollary 3.12 *A Cisinski model structure is homotopically behaved if and only if every generalized J-augmented horn inclusion is a weak equivalence.*

Proof Observe that $J \to *$ has the right lifting property with respect to all monomorphisms, and so is weak equivalence in every Cisinski model structure. Therefore the hypothesis of Proposition 3.10 is satisfied for $\Delta[1] \hookrightarrow J$. $\square$

We can further strengthen this statement once we prove the following lemma.

Proposition 3.13 *Given a simplicial set I and $\Delta[1] \hookrightarrow I$, every generalized I-augmented horn inclusion can be realized as a sequence of pushouts of I-augmented horn inclusions.*

Proof We emulate Joyal's proof for generalized inner horns, [14, Proposition 2.12(iv)]. To show that every generalized I-augmented horn inclusion $\Lambda^S[n]^X_{i\to i+1} \hookrightarrow \Delta[n]^*_{i\to i+1}$ is a pushout of I-augmented horn inclusions, we proceed by induction on $k = n - |S|$. The base case, $k = 0$, is immediate because then $|S| = n$ so $\Lambda^S[n]^X_{i\to i+1} \hookrightarrow \Delta[n]^*_{i\to i+1}$ is itself an I-augmented horn inclusion. For $k \geq 1$, we pick ℓ in $\{0, \dots, n\} \smallsetminus (S \cup \{i, i+1\})$ and let $T = S \cup \{\ell\}$. Then we have

$$\Lambda^S[n]^X_{i\to i+1} \hookrightarrow \Lambda^T[n]^X_{i\to i+1} \hookrightarrow \Delta[n]^*_{i\to i+1}.$$

The right map has $n - |T| < n - |S| = k$. The left map is a pushout of a generalized I-augmented horn inclusion with indexing set S' with the same size as S, so that $n - 1 - |S'| = n - 1 - |S| < k$. Therefore, both of these maps are sequences of pushouts of I-augmented horn inclusions by the inductive hypothesis, meaning the composite is as well. $\square$

Corollary 3.14 *Let $\mathcal{M}$ be a Cisinski model structure on* sSet*. Given a simplicial set I and an inclusion $\Delta[1] \hookrightarrow I$, if the map $I \to *$ is a weak equivalence in the model structure $\mathcal{M}$, then $\mathcal{M}$ is homotopically behaved if and only if every I-augmented horn inclusion is a weak equivalence in $\mathcal{M}$. In particular, the model structure $\mathcal{M}$ is homotopically behaved if and only if every J-augmented horn inclusion is a weak equivalence in $\mathcal{M}$.*

Proof By Proposition 3.13, generalized I-augmented horn inclusions are weak equivalences if and only if I-augmented horn inclusions are. Apply this observation to Proposition 3.10 and Corollary 3.12. $\square$

We conclude this subsection by comparing the lifting properties of fibrant objects in a homotopically behaved model structure to those of quasicategories.

Definition 3.15 Let $h\colon \mathrm{sSet} \to \mathrm{Cat}$ be the left adjoint of the nerve functor. We say an edge in a simplicial set X is a *categorical preisomorphism* if it becomes an isomorphism in the category $h(X)$.

Remark 3.16 The functor h freely builds a category out of a simplicial set X where the set of objects of hX is the set of 0-simplices X_0, and the set of morphisms is generated by the 1-simplices with 2-simplices witnessing composition. We discuss h in more detail in Section 5. The key takeaway at the moment is that the universal property of the unit $X \to hX$ implies that an edge e of X is a categorical preisomorphism precisely if every map from X to the nerve of a category sends e to an isomorphism.

An intuitive justification for why augmented horn inclusions are weak equivalences in a homotopically behaved model structure comes from recalling the special outer horn lifting property of quasicategories.

Proposition 3.17 [12, Theorem 1.3] *If Q is a quasicategory, then for every $n \geq 2$ and every $u : \Lambda^0[n] \to Q$ such that the edge $0 \to 1$ in $\Lambda^0[n]$ is sent to a categorical preisomorphism by u, we have a lift*

$$\begin{tikzcd} \Lambda^0[n] \arrow[r, "u"] \arrow[d, hook] & X \\ \Delta[n] \arrow[ur, dotted] & \end{tikzcd}$$

Similarly, for every $n \geq 2$ and every $v : \Lambda^n[n] \to Q$ such that the edge $n-1 \to n$ in $\Lambda^n[n]$ is sent to a categorical preisomorphism by v, we have an extension of v along $\Lambda^n[n] \hookrightarrow \Delta[n]$.

We can interpret this result is as follows: even though quasicategories do not necessarily have lifts of all outer horns, if we know that the $0 \to 1$ edge of the horn $\Lambda^0[n] \to X$ or the $(n-1) \to n$ edge of the horn $\Lambda^n[n] \to X$ horn is "invertible" in X in a certain sense, then we do get a lift.

This same intuition applies to I-augmented horn inclusions in homotopically behaved model structures where $I \to *$ is a weak equivalence. Since I is weakly equivalent to a point, it makes sense to think of all of the edges of I as invertible. Therefore, a map $\Lambda^j[n]^I_{i \to i+1} \to X$ (where $j = i$ or $i+1$) is a horn in X where we view the $i \to (i+1)$ edge as invertible. The I-augmented horn inclusions being weak equivalences in this model structure implies that if X is fibrant, then we get a lift $\Delta[n]^I_{i \to i+1} \to X$ extending that horn.

3.2 Augmented triangulations

The goal of this subsection is to address a complication arising from the discussion above. To explain, let us first make the following definition.

Definition 3.18 Given Z and an inclusion $\iota : \Delta[1] \hookrightarrow Z$, we say that an edge $e : \Delta[1] \to X$ in an arbitrary simplicial set X is an *Z-edge* if e factors through ι.

Given a simplicial set I and $\Delta[1] \hookrightarrow I$ and a homotopically behaved model structure with $I \to *$ a weak equivalence, we have seen above how it makes sense to view I-edges in arbitrary simplicial sets as invertible. But then any good notion of "invertible edges" should satisfy a *simplicial* 2*-out-of-*3 property:

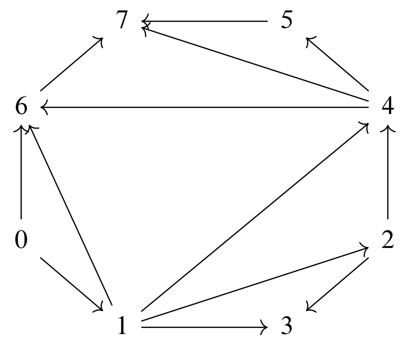


Figure 1

if two edges of a 2-simplex $\Delta[2] \to X$ are invertible, then so is the third edge. The complication is that in an arbitrary simplicial set, the set of I-edges do not necessarily satisfy the simplicial 2-out-of-3 property (unless $I = \Delta[1]$). As a minimal counterexample, one can simply take $\Delta[2]$ itself and glue in a copy of I along two of its nondegenerate edges. The takeaway is that no single I can be used to identify which edges we want to view as invertible in an arbitrary simplicial set (except for the special case when $I = \Delta[1]$).

To address this concern, let us characterize the edges that we want to be invertible even if they are not I-edges themselves. We begin by defining *unordered triangulations*.

Definition 3.19 Given $n \geq 2$ and a regular $(n+1)$-gon with vertices labeled 0 through n (in no particular order), we say an *unordered triangulation* $\mathcal{T}$ is a decomposition of this $(n+1)$-gon into 2-simplices such that every 0-simplex corresponds to a unique vertex of the $(n+1)$-gon and such that the 1-simplices point from lower numbers to higher numbers.

Figure 1 shows an example of an unordered triangulation of the octagon.

Example 3.20 For $n = 2$, there is only one unordered triangulation of the triangle, the standard 2-simplex itself. For $n = 3$, there are precisely six unordered triangulations of the square:

$$\begin{tikzcd} 3 & 2 \arrow[l] \\ 0 \arrow[u] \arrow[r] \arrow[ur] & 1 \arrow[u] \end{tikzcd} \qquad \begin{tikzcd} 3 & 2 \arrow[l] \\ 0 \arrow[u] \arrow[r] & 1 \arrow[u] \arrow[ul] \end{tikzcd}$$

$$\begin{tikzcd} 2 \arrow[r] & 3 \\ 0 \arrow[u] \arrow[r] \arrow[ur] & 1 \arrow[u] \end{tikzcd} \qquad \begin{tikzcd} 2 \arrow[r] & 3 \\ 0 \arrow[u] \arrow[r] & 1 \arrow[u] \arrow[ul] \end{tikzcd} \qquad \begin{tikzcd} 3 & 1 \arrow[l] \arrow[d] \\ 0 \arrow[u] \arrow[r] & 2 \arrow[ul] \end{tikzcd} \qquad \begin{tikzcd} 3 & 1 \arrow[l] \arrow[d] \\ 0 \arrow[u] \arrow[r] \arrow[ur] & 2 \end{tikzcd}$$

Remark 3.21 For those familiar with 2-Segal objects, we note that these unordered triangulations are similar to the triangulations used to define the 2-Segal condition, except that in the 2-Segal definition one

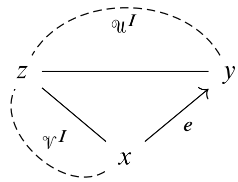


Figure 2

requires the vertices of the $(n+1)$-gon be cyclically ordered with the exception of the $0 \to n$ edge. The first two triangulations in Example 3.20 are the triangulations of the square used in the 2-Segal condition.

We now define *augmented unordered triangulations* that characterize edges that we want to view as invertible.

Definition 3.22 Given a simplicial set I and $\Delta[1] \hookrightarrow I$, an *$I$-augmented unordered triangulation* $\mathcal{T}^I$ is an unordered triangulation $\mathcal{T}$ with a copy of I glued in along all but one of the outer edges. We say that $\mathcal{T}^I$ has *size* n if there are $n+1$ outer edges (and so n copies of I glued in). We consider I itself to be an I-augmented unordered triangulation of size 1. We say that an edge $\Delta[1] \to X$ in an arbitrary simplicial set is an *almost-I-edge* if it is a $\mathcal{T}^I$-edge for some $\mathcal{T}^I$.

In the above definition, all but one of the outer edges being invertible (since they are I-edges) means that we should consider the remaining outer edge to be invertible as well by iterated simplicial 2-out-of-3 arguments. The idea is that if we consider I-edges invertible, then an edge $e\colon \Delta[1] \to X$ in an arbitrary simplicial set is forced to be invertible by iterated application of the simplicial 2-out-of-3 property precisely if it is an almost-I-edge. Figure 2 indicates the inductive argument affirming this intuition, which we spell out in the following propositions.

Proposition 3.23 *Almost-I-edges satisfy the simplicial 2-out-of-3 property. More precisely, if $\Delta[2] \to X$ is a 2-simplex in an arbitrary simplicial set where two of the faces are almost-I-edges, then so is the third.*

Proof Let $\mathcal{U}^I$ and $\mathcal{V}^I$ be I-augmented unordered triangulations that the given edges factor through. Call the remaining edge e. Then we can define $\mathcal{T}^I$ by gluing $\mathcal{U}^I$ and $\mathcal{V}^I$ to the appropriate faces of $\Delta[2]$ as in Figure 2, making the remaining edge a $\mathcal{T}^I$-edge. □

Proposition 3.24 *If the map $I \to *$ is a weak equivalence in a given homotopically behaved model structure, then so is every $\mathcal{T}^I \to *$ and every generalized $\mathcal{T}^I$-augmented horn inclusion.*

Proof We proceed by induction on the size of $\mathcal{T}^I$. The base case of size 1, where $\mathcal{T}^I = I$, is covered by Proposition 3.10.

For $\mathcal{T}^I$ of size bigger than 1, denote by $x \xrightarrow{e} y$ the outer edge of $\mathcal{T}^I$ without a copy of I. Then we can break down $\mathcal{T}^I$ into the 2-simplex of which e is a face plus $\mathcal{U}^I$ and $\mathcal{V}^I$ of smaller size, as in Figure 2.

By the inductive hypothesis, we know that $\mathcal{U}^I \to *$ and $\mathcal{V}^I \to *$ are weak equivalences, so the inclusion of the point z into $\mathcal{U}^I$ and $\mathcal{V}^I$ is as well, and so taking the pushout we see that $* \hookrightarrow \mathcal{U}^I \cup \mathcal{V}^I$ is as well. But then $\mathcal{U}^I \cup \mathcal{V}^I \hookrightarrow \mathcal{T}^I$ is a pushout of a $\mathcal{U}^I$-augmented 2-horn inclusion (or a $\mathcal{V}^I$-augmented 2-horn inclusion). Thus we see that $* \to \mathcal{T}^I$ is a weak equivalence and hence $\mathcal{T}^I \to *$ is also by the 2-out-of-3 property. By applying Proposition 3.10 to $\mathcal{T}^I$, we see that all generalized $\mathcal{T}^I$-augmented horn inclusions are weak equivalences. □

Definition 3.25 We say that a (generalized) $\mathcal{T}^I$-augmented horn inclusion is a (*generalized*) *almost-I-augmented horn inclusion*.

Remark 3.26 Given a simplicial set I and $\Delta[1] \hookrightarrow I$, there are countably many I-augmented unordered triangulations $\mathcal{T}^I$, up to isomorphism. Thus, there are countably many almost-I-augmented horn inclusions up to isomorphism. More generally, given any countable set of inclusions $\{\Delta[1] \hookrightarrow I_r\}_{r \geq 1}$, the set of all almost-$I_r$-augmented horn inclusions for varying $r \geq 1$ is still countable.

We have thus shown that these almost-I-augmented horn inclusions are forced to be weak equivalences in a homotopically behaved model structure where $I \to *$ is a weak equivalence. Our next task is to show that we can apply Cisinski's machinery to this class of maps to get a model structure for certain I.

4 Minimal homotopically behaved model structures

In this section we apply Cisinski's machinery to produce model structures whose fibrant objects are those with lifts of particular augmented horns. We do so using the new concept of a pointwise exact cylinder. We then show that these model structures are "minimal" in a certain sense, both with respect to being homotopically behaved and with respect to the chosen exact cylinders.

4.1 Pointwise cylinders

Our only example of an exact cylinder given above was of the form $X \mapsto I \times X$ for some simplicial set I. In this subsection, we describe a slightly more complex kind of exact cylinder which we use to construct our model structure.

Let sk_0 denote the endofunctor of sSet that sends a simplicial set X to its 0-skeleton (the simplicial set that has the same 0-simplices as X but no nondegenerate higher simplices). Given a monomorphism $\iota\colon \Delta[1] \hookrightarrow I$, we let $\iota \odot X$ be the pushout

$$\begin{tikzcd} \Delta[1] \times \mathrm{sk}_0 X \arrow[r, hook] \arrow[d, hook, "\iota \times \mathrm{sk}_0 X"'] & \Delta[1] \times X \arrow[d, hook] \\ I \times \mathrm{sk}_0 X \arrow[r, hook] & \iota \odot X \end{tikzcd}$$

In other words, the simplicial set $\iota \odot X$ is $\Delta[1] \times X$ with a copy of I glued in along $\Delta[1] \times \{x\}$ for each 0-simplex x of X.

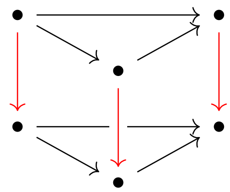


Figure 3

Example 4.1 When $X = \Delta[2]$, we glue in a copy of I along the three red edges of $\Delta[1]\times\Delta[2]$ depicted in Figure 3.

Proposition 4.2 *The above description of $\iota \odot -$ defines an exact cylinder.*

Proof Functoriality follows from $\iota \odot -$ being a pushout of the functors $I \times \mathrm{sk}_0(-)$, $\Delta[1]\times \mathrm{sk}_0(-)$, and $\Delta[1]\times -$. These three functors preserve monomorphisms and small colimits, so their pushout does as well, meaning that $\iota \odot -$ satisfies axiom (DH1) from Definition 2.4. By Remark 2.5, to check axiom (DH2) we simply observe that for any inclusion of simplicial sets $A \hookrightarrow B$, the simplices of $\iota \odot B$ that are in both $\iota \odot A$ and $\{\varepsilon\} \odot B$ are precisely those in $\{\varepsilon\} \odot A$. □

Definition 4.3 We call the exact cylinder $\iota \odot -$ the *pointwise cylinder* for the inclusion $\iota\colon \Delta[1] \hookrightarrow I$.

Example 4.4 The functor $(\mathrm{id}_{\Delta[1]}) \odot -$ is simply the functor $\Delta[1]\times -$, since we do not glue anything extra onto the vertical edges in this case.

Example 4.5 Let P be the pushout on the left below

$$\begin{tikzcd}
\Delta[1] \arrow[r, "s^0"] \arrow[d, "d^1"'] & \Delta[0] \arrow[d] \\
\Delta[2] \arrow[r, "p"'] & P
\end{tikzcd}
\qquad
\begin{tikzcd}
\Delta[1] \arrow[r, "d^0"] \arrow[d, "d^2"'] & \Delta[2] \arrow[r, "p"] & P \arrow[dd, "i_2"] \\
\Delta[2] \arrow[d, "p"'] & & \\
P \arrow[rr, "i_0"'] & & K
\end{tikzcd}$$

and then let K be the pushout on the right, with $\kappa\colon \Delta[1] \hookrightarrow K$ being the diagonal composite of the right square. We can view the simplicial set K as

$$\begin{tikzcd}
b \arrow[r, dotted] \arrow[d, "f"'] & b \arrow[d, "g"] \\
a \arrow[r, dotted] \arrow[ur, "\kappa"] & a
\end{tikzcd}$$

where the dotted arrows indicate degenerate edges. We can think of K as the edge κ with a left inverse g and a right inverse f glued in. We will use the pointwise cylinder $\kappa \odot -$ in later sections.

Example 4.6 Recall that J is the nerve of the free-living isomorphism $\mathbb{I}$. Through a slight abuse of notation, we use $J \odot -$ to denote the pointwise cylinder for the inclusion $\Delta[1] \hookrightarrow J$.

4.2 Constructing the model structures

The rest of this section is devoted to proving the existence of two homotopically behaved model structures. One is the minimal homotopically behaved model structure, whose fibrant objects are precisely the simplicial sets with lifts of J-augmented horn inclusions. The other is a localization of this model structure at $K \to *$ (where K is the simplicial set from Example 4.5), whose fibrant objects are precisely the simplicial sets with lifts of K-augmented horn inclusions. The key to this result is the following proposition; the existence of our desired model structures then follows from Cisinski's Theorem 2.10.

Proposition 4.7 (1) *The almost-J-augmented horn inclusions together with the map $\{0\} \hookrightarrow J$ generate a $(J \odot -)$-anodyne class.*

(2) *The almost-K-augmented horn inclusions together with the maps $\{\varepsilon\} \hookrightarrow K$ for $\varepsilon = 0, 1$ generate a $(\kappa \odot -)$-anodyne class.*

Recall that we defined a $(I \otimes -)$-anodyne class in Definition 2.7, but in Lemma 2.8 we reformulated the axioms to be easier to check. Thus, proving this proposition amounts to verifying the axioms from Lemma 2.8. It turns out that (An1′) and all but one case of (An2′) can be proved just as easily for arbitrary $\Delta[1] \hookrightarrow I$ in the place of $\Delta[1] \hookrightarrow J$ or $\kappa\colon \Delta[1] \hookrightarrow K$. Let us give a name to the $\Delta[1] \hookrightarrow I$ such that the remaining case of (An2′) is satisfied.

Definition 4.8 We say that an inclusion $\Delta[1] \hookrightarrow I$ is *anodyne-ready* if the maps

$$(\partial I \odot I) \cup (I \odot \{\varepsilon\}) \hookrightarrow I \odot I$$

for $\varepsilon = 0, 1$ are a sequence of pushouts of almost-I-augmented horn inclusions.

We can thus break down the proof of Proposition 4.7 into the following two pieces.

Proposition 4.9 *If $\iota\colon \Delta[1] \hookrightarrow I$ is anodyne-ready, then the almost-I-augmented horn inclusions together with the maps $\{\varepsilon\} \hookrightarrow I$ for $\varepsilon = 0, 1$ generate an $(\iota \odot -)$-anodyne class.*

Proposition 4.10 *The inclusions $\Delta[1] \hookrightarrow J$ and $\kappa\colon \Delta[1] \hookrightarrow K$ are anodyne-ready.*

We prove Proposition 4.9 in Section 4.3 and prove Proposition 4.10 in Section 4.4. We discuss the resulting model structures we get from Cisinski's theory in Section 4.5.

Remark 4.11 In our arguments below, we view the simplicial set $\Delta[1] \times \Delta[n]$ as the nerve of the poset $[1] \times [n]$

$$\begin{array}{ccccccccc}
(0,0) & \longrightarrow & (0,1) & \longrightarrow & \cdots & \longrightarrow & (0,n-1) & \longrightarrow & (0,n) \\
\downarrow & & \downarrow & & & & \downarrow & & \downarrow \\
(1,0) & \longrightarrow & (1,1) & \longrightarrow & \cdots & \longrightarrow & (1,n-1) & \longrightarrow & (1,n)
\end{array}$$

and so we view simplices of $\Delta[1] \times \Delta[n]$ as paths in this poset. We depict the first coordinate vertically.

4.3 Proving Proposition 4.9

Given I and $\Delta[1] \hookrightarrow I$, let $A(I)$ denote the set of I-augmented horn inclusions plus the maps $\{\varepsilon\} \hookrightarrow I$. The inclusion $\Delta[1] \hookrightarrow I$ is anodyne-ready precisely if $A(I)$ partially satisfies axiom (An2′). In this subsection, we justify this terminology by showing that $A(I)$ satisfies axiom (An1′) and the rest of (An2′) for arbitrary I and $\Delta[1] \hookrightarrow I$, making $\Delta[1] \hookrightarrow I$ being anodyne ready precisely the missing piece for $A(I)$ to generate an $(I \odot -)$-anodyne class.

We begin by checking that our set of maps $A(I)$ satisfies axiom (An1′).

Lemma 4.12 *Given a simplicial set I and $\Delta[1] \hookrightarrow I$, the maps $(\{\varepsilon\} \odot \Delta[n]) \cup (I \odot \partial\Delta[n]) \hookrightarrow I \odot \Delta[n]$ for $n \geq 1$ and $\varepsilon = 0, 1$ can be realized as a sequence of pushouts of I-augmented horn inclusions.*

Proof We prove the case $\varepsilon = 0$; the argument for $\varepsilon = 1$ is similar. We begin by identifying which simplices of $I \odot \Delta[n]$ are neither in $\{0\} \odot \Delta[n]$ nor in $I \odot \partial\Delta[n]$. Because $n \geq 1$, the extra copies of I glued in along $\Delta[1] \times \mathrm{sk}_0(\Delta[n])$ are already present in $I \odot \partial\Delta[n]$, and so all of the simplices not in the domain of our inclusion must be contained in $\Delta[1] \times \Delta[n] \subseteq I \odot \Delta[n]$.

For $0 \leq j \leq n$, let P_j be the $(n+1)$-simplex corresponding to the path

$$\begin{tikzcd}
(0,0) \arrow[r] & \cdots \arrow[r] & (0,j-1) \arrow[r] & (0,j) \arrow[d] \\
& & & (1,j) \arrow[r] & (1,j+1) \arrow[r] & \cdots \arrow[r] & (1,n)
\end{tikzcd}$$

and for $0 \leq j \leq n-1$, let Q^j_{j+1} be the n-simplex corresponding to the path

$$\begin{tikzcd}
(0,0) \arrow[r] & \cdots \arrow[r] & (0,j-1) \arrow[r] & (0,j) \arrow[dr] \\
& & & & (1,j+1) \arrow[r] & \cdots \arrow[r] & (1,n)
\end{tikzcd}$$

Let Q_0 denote the n-simplex $(1,0) \to \cdots \to (1,n)$. These simplices are precisely the simplices of $\Delta[1] \times \Delta[n]$ that are not contained in $\{0\} \times \Delta[n]$ or in $\Delta[1] \times \partial\Delta[n]$, since any other simplex either avoids both vertices $(0,j)$ and $(1,j)$ for some $0 \leq j \leq n$ and so is in $\Delta[1] \times \partial\Delta[n]$, or is the n-simplex $(0,0) \to \cdots \to (0,n)$ that is contained in $\{0\} \times \Delta[n]$. The diagram in Figure 4 shows how these simplices fit together, with an arrow indicating that one simplex is a face of another. It remains to describe the process by which we glue in each of these simplices via an I-augmented horn pushout. The red arrows indicate which pairs of simplices are attached at the same step of this process, and conversely a black arrow indicates that the simplices are glued in at different steps.

Let us spell out this process explicitly. To attach these simplices via I-augmented horn pushouts, we begin with P_n, whose only missing face is its d_n face, the n-simplex Q^{n-1}_n. This horn pushout can be realized as an I-augmented horn pushout because the $n \to n+1$ edge of P_n is the vertical edge

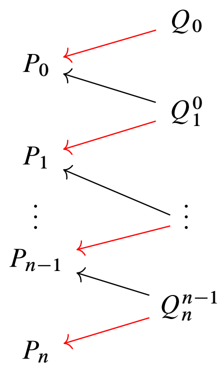


Figure 4

$(0,n) \to (1,n)$, which is an I-edge. We continue inductively, gluing in each P_j together with its d_j face for $j = n-1, n-2, \dots, 2, 1, 0$. The red arrows in the diagram highlight the inclusion of the d_j face into each $(n+1)$-simplex P_j. $\square$

We record a consequence of this lemma for later use.

Corollary 4.13 *Given a simplicial set I and $\Delta[1] \hookrightarrow I$ and any bijective-on-0-simplices inclusion $A \hookrightarrow B$, the maps $(\{\varepsilon\} \odot B) \cup (I \odot A) \hookrightarrow I \odot B$ for $\varepsilon = 0, 1$ can be realized as a sequence of pushouts of I-augmented horn inclusions.*

Proof Since $A \hookrightarrow B$ is bijective on 0-simplices, it can be witnessed as a sequence of pushouts of boundary inclusions $\partial\Delta[n] \hookrightarrow \Delta[n]$ for $n \geq 1$, and so $(\{\varepsilon\} \odot B) \cup (I \odot A) \hookrightarrow I \odot B$ can be witnessed as a sequence of pushouts of the maps $(\{\varepsilon\} \odot \Delta[n]) \cup (I \odot \partial\Delta[n]) \hookrightarrow I \odot \Delta[n]$ for $n \geq 1$. $\square$

Having shown that $A(I)$ satisfies (An1$'$), we turn to proving that part of (An2$'$) is satisfied, which follows from the following more general lemma (by setting $I' = \mathcal{T}^I$).

Lemma 4.14 *Fix $\Delta[1] \hookrightarrow I'$. For all $n \geq 2$ and $0 \leq i \leq n$, if $A \hookrightarrow B$ is a pushout along*

$$\big((\{0\} \sqcup \{1\}) \times \Delta[n]\big) \cup \big(\Delta[1] \times \Lambda^i[n]\big) \hookrightarrow \Delta[1] \times \Delta[n]$$

such that either the $(\varepsilon, i-1) \to (\varepsilon, i)$ edges or the $(\varepsilon, i) \to (\varepsilon, i+1)$ edges (for $\varepsilon = 0, 1$) are sent to I'-edges in A, then $A \hookrightarrow B$ is a finite composite of pushouts of I'-augmented horn inclusions.

Proof We begin by identifying which simplices of $\Delta[1] \times \Delta[n]$ are not in $\Delta[1] \times \Lambda^i[n]$ or $(\{0\} \sqcup \{1\}) \times \Delta[n]$. A simplex is in $\Delta[1] \times \Lambda^i[n]$ if there is some $0 \leq j \leq n$ with $j \neq i$ such that the simplex avoids both of the vertices $(0, j)$ and $(1, j)$. A simplex is in $(\{0\} \sqcup \{1\}) \times \Delta[n]$ if its vertices are all 0 or all 1 in the first coordinate.

Let us fix notation for each of the simplices that avoid satisfying both of these criteria. First, we let P_j and Q^j_{j+1} be defined as in the proof of Lemma 4.12. For $0 \leq i, j \leq n$ such that $j \neq i$, let $R(i)_j$ be the

n-face of P_j that skips the (ε, i) vertex (where $\varepsilon = 0$ if $j > i$ and $\varepsilon = 1$ if $j < i$). For $0 \le i \le n$ and $0 \le j \le n-1$ such that $j \ne i-1, i$, let $S(i)^j_{j+1}$ be the $(n-1)$-face of Q^j_{j+1} that skips the (ε, i) vertex (where $\varepsilon = 0$ if $j > i$ and $\varepsilon = 1$ if $j < i-1$). Let $S(i)^{i-1}_{i+1}$ be the $(n-1)$-simplex corresponding to the path

$$\begin{tikzcd}[column sep=small]
(0,0) \arrow[r] & \cdots \arrow[r] & (0,i-1) \arrow[drr] & & & \\
& & & & (1,i+1) \arrow[r] & \cdots \arrow[r] & (1,n)
\end{tikzcd}$$

Figure 5 shows how these simplices fit together, with an arrow indicating that one simplex is a face of another. As in Figure 4 from the proof of Lemma 4.12, the different-colored arrows indicate which pairs of simplices are attached at the same step of the process below, while black arrows indicate that the simplices are glued in at different steps.

To describe $A \hookrightarrow B$ as a sequence of pushouts of I'-augmented horn inclusions, we consider the case where the $(\varepsilon, i) \to (\varepsilon, i+1)$ edges are sent to I'-edges in A. (The other case is similar.) For $0 \le j \le i-1$ and $i+1 \le j \le n-1$, the only face of Q^j_{j+1} that is not already in A is its d_i face $S(i)^j_{j+1}$. When $0 \le j \le i-1$, the $(1,i) \to (1,i+1)$ edge is in Q^j_{j+1}, and when $i+1 \le j \le n-1$, the $(0,i) \to (0,i+1)$ edge is in Q^j_{j+1}, in both cases corresponding to the $i \to (i+1)$ edge of the n-simplex Q^j_{j+1}. We also have $S(i)^{i-1}_{i+1}$ as the d_i face of Q^{i-1}_i, where the edge $(1,i) \to (1,i+1)$ is the $i \to (i+1)$ edge of Q^{i-1}_i. We can therefore glue in every Q simplex (along with its d_i face) except for Q^i_{i+1} as the first steps in our sequence of I'-augmented horn pushouts. The pairs of simplices glued in at this step are indicated by the blue arrows in the diagram. Then for $0 \le j \le i-1$ the d_{i+1} face of P_j is $R(i)_j$ and the $i+1 \to i+2$ edge is $(1,i) \to (1,i+1)$, and for $i+1 \le j \le n-1$ the d_i face of P_j is $R(i)_j$ and the $i \to (i+1)$ edge is $(0,i) \to (0,i+1)$, so we can now glue in every P simplex (along with its missing d_i or d_{i+1} face) except for P_i as the next steps in our sequence of pushouts. These steps are indicated by the red arrows. All that is left is P_i and its d_{i+1} face Q^i_{i+1}, and since the $i+1 \to i+2$ edge of P_i is $(1,i) \to (1,i+1)$, these remaining simplices are glued in via an I'-augmented horn pushout as well, indicated by the green arrow. □

By applying Lemma 4.14 when $I' = \mathcal{T}^I$, we get the following corollary which says that $A(I)$ satisfies part of (An2′).

Corollary 4.15 *Given a simplicial set I and $\Delta[1] \hookrightarrow I$, for every almost-$I$-augmented horn inclusion*

$$\Lambda^j[n]^{\mathcal{T}^I}_{i \to i+1} \hookrightarrow \Delta[n]^{\mathcal{T}^I}_{i \to i+1}$$

(where $j = i$ or $i+1$), the map

$$(\partial J \odot \Delta[n]^{\mathcal{T}^I}_{i \to i+1}) \cup (I \odot \Lambda^j[n]^{\mathcal{T}^I}_{i \to i+1}) \hookrightarrow I \odot \Delta[n]^{\mathcal{T}^I}_{i \to i+1}$$

is a finite composite of pushouts of almost-I-augmented horn inclusions.

Proof This map satisfies the hypotheses of Lemma 4.14 when $I' = \mathcal{T}^I$. □

$$\begin{tikzcd}
P_0 & & R(i)_0 \arrow[ll, red] & \\
& Q_1^0 \arrow[ul] \arrow[dl] & & S(i)_1^0 \arrow[ll, blue] \arrow[ul] \arrow[dl] \\
P_1 & & R(i)_1 \arrow[ll, red] & \\
\vdots & \vdots \arrow[ul] \arrow[dl] & \vdots & \vdots \arrow[ul] \arrow[dl] \\
P_{i-2} & & R(i)_{i-2} \arrow[ll, red] & \\
& Q_{i-1}^{i-2} \arrow[ul] \arrow[dl] & & S(i)_{i-1}^{i-2} \arrow[ll, blue] \arrow[ul] \arrow[dl] \\
P_{i-1} & & R(i)_{i-1} \arrow[ll, red] & \\
& Q_i^{i-1} \arrow[ul] \arrow[dl] & & \\
P_i & & & S(i)_{i+1}^{i-1} \arrow[ull, blue] \arrow[uul] \arrow[dll] \arrow[ddl] \\
& Q_{i+1}^i \arrow[ul, green] \arrow[dl] & & \\
P_{i+1} & & R(i)_{i+1} \arrow[ll, red] & \\
& Q_{i+2}^{i+1} \arrow[ul] \arrow[dl] & & S(i)_{i+2}^{i+1} \arrow[ll, blue] \arrow[ul] \arrow[dl] \\
P_{i+2} & & R(i)_{i+2} \arrow[ll, red] & \\
\vdots & \vdots \arrow[ul] \arrow[dl] & \vdots & \vdots \arrow[ul] \arrow[dl] \\
P_{n-1} & & R(i)_{n-1} \arrow[ll, red] & \\
& Q_n^{n-1} \arrow[ul] \arrow[dl] & & S(i)_n^{n-1} \arrow[ll, blue] \arrow[ul] \arrow[dl] \\
P_n & & R(i)_n \arrow[ll, red] &
\end{tikzcd}$$

Figure 5

We have now assembled the necessary ingredients to prove Proposition 4.9, which says that $\Delta[1] \hookrightarrow I$ being anodyne-ready is indeed the missing piece for the set of maps $A(I)$ to generate an anodyne class.

Proof of Proposition 4.9 To check that axiom (An1′) is satisfied, we note that the map

$$(\{\varepsilon\} \odot \Delta[0]) \cup (I \odot \partial\Delta[0]) \hookrightarrow I \odot \Delta[0]$$

for $\varepsilon = 0, 1$ is isomorphic to $\{\varepsilon\} \hookrightarrow I$, which takes care of the $n = 0$ case. For $n \geq 1$ we apply Lemma 4.12.

To check that axiom (An2′) is satisfied, we use Corollary 4.15 to account for all of the almost-I-augmented horn inclusions. The remaining maps $\{\varepsilon\} \hookrightarrow I$ are accounted for because we assumed $\Delta[1] \hookrightarrow I$ to be anodyne-ready. □

4.4 Proving Proposition 4.10

We now turn to showing that $\Delta[1] \hookrightarrow J$ and $\Delta[1] \hookrightarrow K$ are anodyne-ready. An important ingredient of the proof will be the following observation about when we can upgrade ordinary horn inclusions to augmented horn inclusions.

Lemma 4.16 *Given a simplicial set I and $\Delta[1] \hookrightarrow I$ and $k \geq 2$, if $\Lambda^j[k] \to X$ is a k-horn in some simplicial set such that all of the edges of $\Lambda^j[k]$ are sent to almost-I-edges, then all of the edges of $\Delta[k]$ are sent to almost-I-edges in $\Delta[k] \sqcup_{\Lambda^j[k]} X$. Furthermore, the inclusion $X \hookrightarrow \Delta[k] \sqcup_{\Lambda^j[k]} X$ can be witnessed as a pushout of an almost-I-augmented horn inclusion.*

Proof For $k > 2$, there are no new edges added in the pushout, so the first claim is immediate. For $k = 2$, since the two edges of $\Lambda^j[2]$ are sent to almost-I-edges, the new edge is also an almost-I-edge in the pushout $X \sqcup_{\Lambda^j[2]} \Delta[2]$ by applying the simplicial 2-out-of-3 property from Proposition 3.23. In each of these cases, the inclusion $X \hookrightarrow \Delta[k] \sqcup_{\Lambda^j[k]} X$ can be upgraded to an almost-I-augmented horn pushout because every edge of the horn in X is an almost-I-edge. □

Iterated application of Lemma 4.16 yields the following corollary.

Corollary 4.17 *Given a simplicial set I and $\Delta[1] \hookrightarrow I$, if $A \hookrightarrow B$ is a sequence of pushouts of k-horns for varying $k \geq 2$, and $A \to X$ is a map such that all of the edges of A are sent to almost-I-edges, then all of the edges of B are sent to almost-I-edges in $B \sqcup_A X$. Furthermore, the inclusion $X \hookrightarrow B \sqcup_A X$ can be witnessed as a sequence of pushouts of almost-I-augmented horn inclusions.*

Another ingredient to the proof of Proposition 4.10 is the following observation that the given inclusions are themselves a sequence of pushouts of ordinary k-horns.

Lemma 4.18 *The inclusions $\Delta[1] \hookrightarrow J$ and $\kappa\colon \Delta[1] \hookrightarrow K$ are obtained via a sequence of pushouts of k-horn inclusions for $k \geq 2$.*

Proof We first consider $\Delta[1] \hookrightarrow J$. Recall that J has precisely two nondegenerate n-simplices for all $n \geq 0$. Furthermore, for each n-simplex, only the d_0 and d_n faces are nondegenerate. We may therefore inductively build J from $\Delta[1]$ as follows. For the base case, we observe that $\Delta[1] \subseteq J$ contains both 0-simplices and one of the nondegenerate 1-simplices. Now, assuming for $n \geq 1$ that we have glued in all $(n-1)$-simplices and exactly one of the nondegenerate n-simplices, we may glue in one of the $(n+1)$-simplices along an $(n+1)$-horn (where $n+1 \geq 2$) since it is missing exactly one of its faces. We then have all of the n-simplices and exactly one of the nondegenerate $(n+1)$-simplices.

Now let us consider $\kappa\colon \Delta[1] \hookrightarrow K$. We build K out of $\Delta[1]$ by pushouts along outer 2-horns where the $0 \to 2$ edge is degenerate. □

$$\begin{tikzcd}
P_0 & & & \\
& Q_1^0 \arrow[ul] \arrow[dl] & & \\
P_1 & & R(0)_1 \arrow[ll] & \\
& Q_2^1 \arrow[ul] \arrow[dl] & & S(0)_2^1 \arrow[ul] \arrow[ll] \arrow[dl] \\
P_2 & & R(0)_2 \arrow[ll] & \\
& Q_3^2 \arrow[ul] \arrow[dl] & & S(0)_3^2 \arrow[ul] \arrow[ll] \arrow[dl] \\
\vdots & & \vdots &
\end{tikzcd}$$

Figure 6

Combining Lemma 4.18 and Corollary 4.17 yields the following corollary.

Corollary 4.19 *Given a simplicial set X, if an edge $\Delta[1] \to X$ is an almost-J-edge, then the inclusion $X \hookrightarrow J \sqcup_{\Delta[1]} X$ can be witnessed as a pushout of almost-J-augmented horns. Similarly, if $\Delta[1] \to X$ is an almost-K-edge, then the inclusion $X \hookrightarrow K \sqcup_{\Delta[1]} X$ can be witnessed as a pushout of almost-K-augmented horns.*

The last ingredient to the proof of Proposition 4.10 is the observation that each inclusion

$$(\partial\Delta[1] \times I) \cup (\Delta[1] \times \{\varepsilon\}) \hookrightarrow (\Delta[1] \times I)$$

for $I = J, K$ and $\varepsilon = 0, 1$ is also obtained via a sequence of pushouts of ordinary k-horns. The following two lemmas handle the two cases $I = J$ and $I = K$ separately.

Lemma 4.20 *The inclusion $(\partial\Delta[1] \times J) \cup (\Delta[1] \times \{0\}) \hookrightarrow \Delta[1] \times J$ is a sequence of pushouts of k-horn inclusions for varying $k \geq 2$.*

Proof For each $\ell \geq 0$, let B_ℓ denote the nondegenerate ℓ-simplex of J whose initial vertex is 0. Each B_ℓ contains all m-simplices for $m < \ell$, so $J = \bigcup_{\ell \geq 0} B_\ell$. Notice that $B_0 = \{0\}$, so $(\partial\Delta[1] \times J) \cup (\Delta[1] \times \{0\})$ is precisely $(\partial\Delta[1] \times J) \cup (\Delta[1] \times B_0)$. It therefore suffices to show that each inclusion

$$(\partial\Delta[1] \times J) \cup (\Delta[1] \times B_\ell) \hookrightarrow (\partial\Delta[1] \times J) \cup (\Delta[1] \times B_{\ell+1})$$

for all $\ell \geq 0$ is a sequence of pushouts of $(\ell+1)$-horns and $(\ell+2)$-horns, with $(\ell+1)$-horns only necessary when $\ell \geq 1$. Recalling the notation from Figure 5 in the proof of Lemma 4.14, we depict in Figure 6 the simplices of $\Delta[1] \times B_{\ell+1}$ that are not in $\Delta[1] \times B_\ell$. We can proceed by gluing in the Q^i_{i+1} simplices together with the $S(0)^i_{i+1}$ simplices for each $i \geq 1$ via pushouts of $(\ell+1)$-horns, indicated by the blue arrows. Note that this first step is only necessary if $\ell \geq 1$. Then we can glue in via an $(\ell+2)$-horn the P_0

$$\begin{tikzcd}
(0,b) \arrow[r, "0\times f"] & (0,a) \arrow[r, "0\times g"] \arrow[d, "\Delta[1]\times a"] & (0,b) \arrow[r, "0\times h"] & (0,a) \arrow[d, "\Delta[1]\times a"] \\
(1,b) \arrow[r, "1\times f"'] & (1,a) \arrow[r, "1\times g"'] & (1,b) \arrow[r, "1\times h"'] & (1,a)
\end{tikzcd}$$

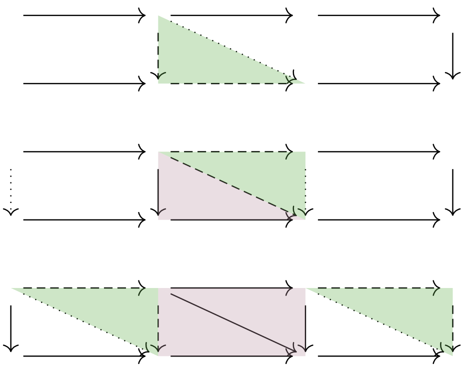


Figure 7

simplex with Q_1^0, as indicated by the red arrow. Finally, we glue each P_i together with $R(0)_i$ for each $i \geq 1$, also via $(\ell+2)$-horns, as indicated by the green arrows. □

Lemma 4.21 *Given $\varepsilon = 0, 1$, the inclusion $(\partial\Delta[1] \times K) \cup (\Delta[1] \times \{\varepsilon\}) \hookrightarrow \Delta[1] \times K$ is a sequence of pushouts of 2-horn and 3-horn inclusions.*

Proof We start by including all of the missing 1-simplices, which we can do working in $\Delta[1] \times \mathrm{sk}_1 K$. For this proof, let us rename the 0 and 1 vertices of K to a and b, respectively, and let f, g, and h denote the nondegenerate edges of K. The 1-simplices of $\Delta[1] \times K$ that we start with can then be pictured as in the first diagram of Figure 7. The latter three diagrams in Figure 7 show the order in which we can glue in four 2-horns to end up with all of the 1-simplices of $\Delta[1] \times K$.

From here, we must glue in the missing 3-simplices as well as the remaining 2-simplices. First, we attach the 3-simplices outlined by the edges

$$\begin{tikzcd}
{} \arrow[d, "\Delta[1]\times a"'] & & \\
{} \arrow[r, "1\times g"'] & {} \arrow[r, "1\times h"'] & {}
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
{} \arrow[r, "0\times g"] & {} \arrow[r, "0\times h"] & {} \arrow[d, "\Delta[1]\times a"] \\
& & {}
\end{tikzcd}$$

The 3-simplex on the left is only missing its d_1 face, and the 3-simplex on the right is only missing its d_2 face (which are different 2-simplices), so we can attach them both via 3-horn extensions. Having done so, we can then glue in the 3-simplex

$$\begin{tikzcd}
{} \arrow[r, "0\times g"] & {} \arrow[d, "\Delta[1]\times b"] & \\
& {} \arrow[r, "1\times h"'] & {}
\end{tikzcd}$$

that is now only missing its d_0 face (which was left empty in the diagram above). We now have all of the nondegenerate 3-simplices that involve the h edges. For those involving the f edges, the order

$$
\begin{tikzcd}
{} \arrow[r, "0\times f"] & {} \arrow[r, "0\times g"] & {} \arrow[d, "\Delta[1]\times b"] \\
& & {}
\end{tikzcd}
\qquad
\begin{tikzcd}
{} \arrow[r, "0\times f"] & {} \arrow[d, "\Delta[1]\times a"] & \\
& {} \arrow[r, "1\times g"'] & {}
\end{tikzcd}
\qquad
\begin{tikzcd}
{} \arrow[d, "\Delta[1]\times b"'] & & \\
{} \arrow[r, "1\times f"'] & {} \arrow[r, "1\times g"'] & {}
\end{tikzcd}
$$

works because the first 3-simplex is only missing its d_2 face, then the second 3-simplex is only missing its d_1 face, and then the last 3-simplex is only missing its d_3 face. □

We now have all of the pieces to prove Proposition 4.10.

Proof of Proposition 4.10 Let $I = J$ or K. We wish to show that the maps $(\partial I \odot I) \cup (I \odot \{\varepsilon\}) \hookrightarrow I \odot I$ for $\varepsilon = 0, 1$ can be realized as a sequence of pushouts of almost-I-augmented horn inclusions. The central claim of the proof is that all of the missing simplices of $\Delta[1] \times I \subseteq I \otimes I$ can be glued in via a sequence of pushouts of k-horns for varying $k \geq 2$, which we proved for $I = J$ in Lemma 4.20 and for $I = K$ in Lemma 4.21. Now, we apply Lemma 4.16 to upgrade it to a sequence of pushouts of almost-I-augmented horn inclusions, and note that all of the new edges are also almost-I-edges. The last step is to glue in a copy of I along the vertical edge that was missing when we started, ie the edge $\Delta[1] \times \{\varepsilon'\}$ where $\varepsilon' \neq \varepsilon$. We can witness this gluing as a sequence of pushouts of almost-I-augmented horn inclusions by Corollary 4.17. □

4.5 The resulting model structures

Having proved Propositions 4.9 and 4.10, and therefore Proposition 4.7, Cisinski's theory gives us our desired model structures.

Proposition 4.22 *For $I = J$ or K, there is a cofibrantly generated model structure on* sSet *whose cofibrations are the monomorphisms, and whose fibrant objects are the simplicial sets X such that $X \to *$ has the right lifting property with respect to the set of almost-I-augmented horn inclusions.*

Proof Let S denote the set containing the maps $\{\varepsilon\} \hookrightarrow I$ for $\varepsilon = 0, 1$ together with the almost-I-augmented horn inclusions. By Proposition 4.7, the set S generates an $(I \odot -)$-anodyne class, so Theorem 2.10 gives us a model structure whose fibrant objects are the simplicial sets X such that $S \boxslash (X \to *)$. Since $(* \hookrightarrow A) \boxslash (Y \to *)$ for all simplicial sets A and Y, a simplicial set is fibrant in this model structure if it has lifts of almost-I-augmented horn inclusions. □

Remark 4.23 Recall that Theorem 2.10 gives a description of the weak equivalences in this model structure as well, but that we do not get an explicit description of the fibrations.

To check that a simplicial set is fibrant in one of these model structures, it is actually not necessary to check all almost-I-augmented horn inclusions.

Proposition 4.24 *Given $I \neq \Delta[1]$ with two 0-simplices and an inclusion $\Delta[1] \hookrightarrow I$, if X is a simplicial set such that $X \to *$ has the right lifting property with respect to all I-augmented horn inclusions, then every almost-I-edge is an I-edge too.*

Proof The result follows if we can show that if an edge $\Delta[1] \to X$ factors through some $\mathcal{T}^I$ then it factors through a $\mathcal{U}^I$ of strictly smaller size. Since there is at least one 2-simplex of every I-augmented unordered triangulation where two of the edges are I-edges, it suffices to show that if two edges of a 2-simplex in X are I-edges, then so is the third edge.

Since $\Delta[1] \hookrightarrow I$ is bijective on 0-simplices, we know by Corollary 4.13 that the maps

$$g_\varepsilon \colon (\{\varepsilon\} \odot I) \cup (I \odot \Delta[1]) \hookrightarrow I \odot I$$

can be witnessed as sequences of pushouts of I-augmented horn inclusions for $\varepsilon = 0, 1$, so $X \to *$ has the right lifting property with respect to g_ε. Let $\mathcal{W}^I$ be $\Delta[2]$ with a copy of I glued along two edges. Then we can choose $\varepsilon \neq \varepsilon'$ and a surjective map $f \colon (\{\varepsilon\} \odot I) \cup (I \odot \Delta[1]) \to \mathcal{W}^I$ that sends the $\{\varepsilon'\} \times \Delta[1]$ edge to the edge e of $\mathcal{W}^I$ that is not an I-edge. Let $g' \colon \mathcal{W}^I \to P$ be the pushout of g along f, so that $X \to *$ also has the right lifting property with respect to g'. Note that the edge e becomes an I-edge in P, so we have shown that every $\mathcal{W}^I$-edge of X is also an I-edge because we can extend every $\mathcal{W}^I \to X$ along $g' \colon \mathcal{W}^I \hookrightarrow P$. $\square$

Corollary 4.25 *Let $I = J$ or K. A simplicial set X is fibrant in the model structure from Proposition 4.22 if and only if $X \to *$ has the right lifting property with respect to all I-augmented horn inclusions.*

Proof The forward implication is immediate because every I-augmented horn inclusion is also an almost-I-augmented horn inclusion. For the other implication, let us assume $X \to *$ has the right lifting property with respect to all I-augmented horn inclusions. Given an almost-I-augmented horn in X, by Proposition 4.24 the almost-I edge of the horn can be turned into an I-edge, so we get a lift. $\square$

Definition 4.26 We call the model structures from Proposition 4.22 the *minimal homotopically behaved model structure* and *K-minimal homotopically behaved model structure*.

The word "minimal" in the above definition is justified by the following remark.

Remark 4.27 By Corollary 3.14, a Cisinski model structure is homotopically behaved if and only if the J-augmented horn inclusions are weak equivalences in that model structure. The minimal homotopically behaved model structure therefore has the smallest class of weak equivalences possible for a homotopically behaved model structure. Similarly, if the map $K \to *$ is a weak equivalence in a Cisinski model structure, then that model structure is homotopically behaved if and only if all K-augmented horn inclusions are weak equivalences as well, so the K-minimal homotopically behaved model structure has smallest class of weak equivalences possible for a homotopically behaved model structure where $K \to *$ is a weak equivalence.

Corollary 4.28 *The Joyal model structure is a localization of the K-minimal homotopically behaved model structure.*

Proof Let X be a quasicategory. Given a K-augmented horn $\Lambda^j[n]^K_{i\to i+1} \to X$, there are two cases: either $0 < j < n$, in which case the horn is inner so there is a lift, or the horn is an outer horn with edge $0 \to 1$ or $n-1 \to n$ factoring through K. In the latter case, the edge factoring through K means it is sent to a categorical preisomorphism, making the horn a special outer horn, so there is a lift in this case as well.

We have shown that every fibrant object in the Joyal model structure is also fibrant in the K-minimal homotopically behaved model structure, which implies that the K-minimal homotopically behaved model structure is a localization of the Joyal model structure since their cofibrations are the same. □

Furthermore, these model structures are also minimal with respect to the exact cylinders $J \odot -$ and $\kappa \odot -$, as we now explain.

Remark 4.29 Observe that for all $n \geq 2$, the I-augmented horn inclusion $\Lambda^1[n]^I_{0\to 1} \hookrightarrow \Delta[n]^I_{0\to 1}$ is a retract of the inclusion

$$(\{0\} \odot \Delta[n-1]) \cup (I \odot \partial\Delta[n-1]) \hookrightarrow I \odot \Delta[n-1].$$

The same cannot be said of $\Lambda^j[n]^I_{0\to 1} \hookrightarrow \Delta[n]^I_{j-1\to j}$ for $1 < j \leq n$. However, it is instead true that it is a retract of a closely related inclusion. To illustrate, recall the notation from Lemma 4.12, along with the diagram

$$\begin{tikzcd}
 & Q_0 \arrow[dl, red] \\
P_0 & \\
 & Q_1^0 \arrow[ul] \arrow[dl, red] \\
P_1 & \\
\vdots & \vdots \arrow[ul] \arrow[dl, red] \\
P_{n-1} & \\
 & Q_n^{n-1} \arrow[ul] \arrow[dl, red] \\
P_n &
\end{tikzcd}$$

showing the simplices of $\Delta[1] \odot \Delta[n]$ that are not in $(\{0\} \odot \Delta[n]) \cup (I \odot \partial\Delta[n])$. Let us denote by A_j the union of $(\{0\} \odot \Delta[n]) \cup (I \odot \partial\Delta[n])$ with the simplices $P_{n-j}, P_{n-j+1}, \dots, P_n$. Then $\Lambda^j[n]^I_{0\to 1} \hookrightarrow \Delta[n]^I_{j-1\to j}$ is a retract of $A_j \hookrightarrow I \odot \Delta[n-1]$.

The above remark sets us up to prove the following proposition inductively.

Proposition 4.30 *Given a simplicial set I and $\Delta[1] \hookrightarrow I$ and $n \geq 2$, if the map*

$$(\{0\} \odot \Delta[n-1]) \cup (I \odot \partial\Delta[n-1]) \hookrightarrow I \odot \Delta[n-1]$$

is a weak equivalence, then so is every I-augmented horn inclusion of the form

$$\Lambda^j[n]^I_{0\to 1} \hookrightarrow \Delta[n]^I_{j-1\to j}.$$

Similarly, if the map

$$(\{1\} \odot \Delta[n-1]) \cup (I \odot \partial\Delta[n-1]) \hookrightarrow I \odot \Delta[n-1]$$

is a weak equivalence, then so is every I-augmented horn inclusion of the form

$$\Lambda^j[n]^I_{0\to 1} \hookrightarrow \Delta[n]^I_{j\to j+1}.$$

Proof We prove the first claim, as the second is similar. We proceed by induction on j. As observed in Remark 4.29, for the base case $j = 1$, the inclusion

$$\Lambda^j[n]^I_{0\to 1} \hookrightarrow \Delta[n]^I_{j-1\to j}$$

is a retract of

$$(\{0\} \odot \Delta[n-1]) \cup (I \odot \partial\Delta[n-1]) \hookrightarrow I \odot \Delta[n-1],$$

so is a weak equivalence. Now, assuming $1 < j \le n$ and each

$$\Lambda^\ell[n]^I_{0\to 1} \hookrightarrow \Delta[n]^I_{\ell-1\to \ell}$$

for $1 \le \ell < j$ is a weak equivalence, then, using the notation from Remark 4.29, the inclusion

$$(\{0\} \odot \Delta[n-1]) \cup (I \odot \partial\Delta[n-1]) \hookrightarrow A_j$$

is a weak equivalence, and so $A_j \hookrightarrow I \odot \Delta[n-1]$ is by the 2-out-of-3 property, and therefore

$$\Lambda^j[n]^I_{0\to 1} \hookrightarrow \Delta[n]^I_{j-1\to j}$$

is a weak equivalence since it is a retract of $A_j \hookrightarrow I \odot \Delta[n-1]$. □

Corollary 4.31 *Let S be a set of monomorphisms.*

(1) *If S generates an $(I \odot -)$-anodyne class for some I, then the corresponding Cisinski model structure $\mathcal{M}$ (whose fibrant objects are simplicial sets with the right lifting property with respect to S) is a localization of the minimal homotopically behaved model structure.*

(2) *If S generates a $(\kappa \odot -)$-anodyne class, then the corresponding Cisinski model structure $\mathcal{M}$ is a localization of the K-minimal homotopically behaved model structure.*

Proof We first prove (1). The maps $(\{\varepsilon\} \odot \Delta[n-1]) \cup (I \odot \partial\Delta[n-1]) \hookrightarrow I \odot \Delta[n-1]$ are necessarily in ${}^{\boxslash}(S^{\boxslash})$ and hence weak equivalences for $\varepsilon = 0, 1$ and $n \ge 2$ in $\mathcal{M}$, and so by Proposition 4.30 all of the I-augmented horn inclusions are also weak equivalences. Since the inclusion $\{0\} \hookrightarrow I$ is also in ${}^{\boxslash}(S^{\boxslash})$ and hence a weak equivalence, by the 2-out-of-3 property the map $I \to *$ is also a weak equivalence. We may therefore apply Corollary 3.14 to see that $\mathcal{M}$ is homotopically behaved, and so is a localization of the minimal homotopically behaved model structure by Remark 4.27.

To prove (2), we apply (1) with $I = K$ to see that $\mathcal{M}$ is homotopically behaved. In our proof of (1) we also see that $K \to *$ is a weak equivalence in $\mathcal{M}$, so by Remark 4.27 we can conclude that $\mathcal{M}$ is a localization of the K-minimal homotopically behaved model structure. $\square$

Remark 4.32 We note that K-augmented horns $\Lambda^i[n]^K_{i \to i+1}$ are themselves fibrant in the minimal homotopically behaved model structure, but not in the K-minimal homotopically behaved model structure, so these model structures are distinct. Furthermore, the map $K \to *$ is not a weak equivalence in the minimal homotopically behaved model structure because otherwise Proposition 3.10 would imply that all K-augmented horn inclusions would be as well and so the model structures would be the same.

We summarize the results of this section with the following theorem.

Theorem 4.33 *Let $\mathcal{M}_{\mathrm{mhb}}$ be the minimal homotopically behaved model structure on* sSet, *and let $\mathcal{M}_{K,\mathrm{mhb}}$ be the K-minimal homotopically behaved model structure on* sSet.

(1) (a) *Every homotopically behaved model structure is a localization of $\mathcal{M}_{\mathrm{mhb}}$.*
   (b) *Every Cisinski model structure corresponding to an $(I \odot -)$-anodyne class for some I is a localization of $\mathcal{M}_{\mathrm{mhb}}$.*
   (c) *The fibrant objects in $\mathcal{M}_{\mathrm{mhb}}$ are the simplicial sets with the right lifting property with respect to all J-augmented horn inclusions.*
   (d) *The Joyal model structure is a localization of $\mathcal{M}_{\mathrm{mhb}}$.*

(2) (a) *The model structure $\mathcal{M}_{K,\mathrm{mhb}}$ is the localization of $\mathcal{M}_{\mathrm{mhb}}$ with respect to $K \to *$.*
   (b) *Every Cisinski model structure corresponding to a $(K \odot -)$-anodyne class is a localization of $\mathcal{M}_{K,\mathrm{mhb}}$.*
   (c) *The fibrant objects in $\mathcal{M}_{K,\mathrm{mhb}}$ are the simplicial sets with the right lifting property with respect to all K-augmented horn inclusions.*
   (d) *The Joyal model structure is a localization of $\mathcal{M}_{K,\mathrm{mhb}}$.*

We conclude this section with one more useful corollary to this theorem.

Corollary 4.34 *Given a set of monomorphisms $S = \{A_i \hookrightarrow B_i\}$ of simplicial sets such that the inclusions*

$$(A_i \times \Delta[1]) \cup (B_i \times \partial\Delta[1]) \hookrightarrow B_i \times \Delta[1]$$

are in ${}^{\boxslash}(S^{\boxslash})$, there exists a homotopically behaved model structure on sSet *whose fibrant objects are those with lifts against S and all J-augmented horn inclusions.*

Proof We claim that the set S together with $\{0\} \hookrightarrow J$ and the set of almost-J-augmented horn inclusions generates a $(J \otimes -)$-anodyne class. Because we know that $\{0\} \hookrightarrow J$ together with the J-augmented horn inclusions generate such a class, it suffices to check that S satisfies axiom (An2′). However, the maps that we must show are in ${}^{\boxslash}(S^{\boxslash})$ are pushouts of the maps we have assumed are in ${}^{\boxslash}(S^{\boxslash})$, so we are done. $\square$

5 A model structure for special horn inclusions

Intuitively, considering I-edges to be "invertible" implies we want I-augmented horn inclusions to be weak equivalences. So far, the invertibility of I-edges has come from $I \to *$ being a weak equivalence. In particular, since every $\mathcal{T}^K \to *$ is a weak equivalence in the Joyal model structure, the almost-K-augmented horn inclusions are also Joyal weak equivalences. However, the Joyal model structure comes with its own notion of invertible edges, the categorical preisomorphisms. The following example demonstrates that, in an arbitrary simplicial set, a categorical preisomorphism need not be an almost-K-edge.

Example 5.1 Let T be the simplicial set depicted by

$$\begin{tikzcd}
z \arrow[d] & y \arrow[l] & y \arrow[l, dotted] \arrow[d] \\
x & x \arrow[l, dotted] \arrow[u, "e_T"] \arrow[ul] & w \arrow[l] \arrow[ul]
\end{tikzcd}$$

with $e_T : \Delta[1] \hookrightarrow T$ the vertical edge in the middle. We have a nondegenerate 2-simplex for each of the four triangles in the picture. The dotted arrows indicate degenerate edges. If T were Joyal equivalent to $\Delta[0]$, then every edge of T would be a categorical preisomorphism, but in fact e_T is the only nondegenerate categorical preisomorphism in T as the functor $T \to \mathrm{Set}$

$$\begin{tikzcd}
\{a, c\} \arrow[d] & \{a'\} \arrow[l, "a"'] & \{a'\} \arrow[l, dotted] \arrow[d] \\
\{a\} & \{a\} \arrow[l, dotted] \arrow[u] \arrow[ul, "a"] & \{a, c\} \arrow[l] \arrow[ul]
\end{tikzcd}$$

which sends every other nondegenerate edge of T to a nonisomorphism in Set, demonstrates.

Furthermore, there is no simplicial set T' that is Joyal equivalent to $\Delta[0]$ such that the inclusion $e_T \colon \Delta[1] \hookrightarrow T$ from the above example factors through $e_{T'} \colon \Delta[1] \hookrightarrow T'$, because the sequence of edges in T' that provide a left inverse to $e_{T'}$ are all categorical preisomorphisms, but there is no directed sequence of categorical preisomorphisms in T from y to x. Therefore, while all T-edges are necessarily categorical preisomorphisms, this example shows that they need not be almost-K-edges.

The goal of this section is to address this disparity. We identify a set of inclusions $\mathscr{E} = \{\Delta[1] \hookrightarrow T\}$ such that an edge in an arbitrary simplicial set is a categorical equivalence if and only if it is a T-edge for some $\Delta[1] \hookrightarrow T$ in $\mathscr{E}$. We then define the set of *special horn inclusions* to be the set of T-augmented horn inclusions for all $\Delta[1] \hookrightarrow T$ in $\mathscr{E}$. There turns out to be an intermediate model structure between the K-minimal homotopically behaved model structure and the Joyal model structure where the fibrant objects are precisely the simplicial sets with lifts of special horn inclusions.

We begin by establishing notation and terminology. This first definition is standard.

Definition 5.2 For $n \geq 1$, let $\mathrm{Sp}[n]$ denote the *spine* of $\Delta[n]$, the union of the edges $i \to (i+1)$ in $\Delta[n]$ ranging over $0 \leq i \leq n-1$.

$$\begin{tikzcd}[row sep=small]
\mathrm{Sp}[n] & C_i[n,n+1] & & \mathrm{Sp}[n+1] \\
n \arrow[r, dotted, no head] & n+1 \arrow[rr, dotted, no head] & & n+1 \\
\vdots \arrow[u] & \vdots \arrow[u] & & \vdots \arrow[u] \\
i \arrow[u] \arrow[r, dotted, no head] & i+1 \arrow[u] \arrow[rr, dotted, no head] & & i+1 \arrow[u] \\
& & i \arrow[ul] \arrow[r, dotted, no head] & i \arrow[u] \\
i-1 \arrow[uu] \arrow[r, dotted, no head] & i-1 \arrow[uu] \arrow[ur] \arrow[rr, dotted, no head] & & i-1 \arrow[u] \\
\vdots \arrow[u] & \vdots \arrow[u] & & \vdots \arrow[u] \\
0 \arrow[u] \arrow[r, dotted, no head] & 0 \arrow[u] \arrow[rr, dotted, no head] & & 0 \arrow[u]
\end{tikzcd}$$

Figure 8

For the purposes of this section, we introduce some new notions in the following definitions.

Definition 5.3 For $n \geq 1$ and $1 \leq i \leq n$, let $C_i[n, n+1]$ denote the union in $\Delta[n+1]$ of $\mathrm{Sp}[n+1]$ with the 2-simplex $(i-1) \to i \to (i+1)$. Call $C_i[n, n+1]$ a *composition tile*.

Remark 5.4 There are two inclusions of spines into the composition tile $C_i[n, n+1]$ that preserve the initial and final vertex, the inclusion $\mathrm{Sp}[n+1] \hookrightarrow C_i[n, n+1]$ that hits every vertex of $C_i[n, n+1]$ and the inclusion $\mathrm{Sp}[n] \hookrightarrow C_i[n, n+1]$ that avoids the i^{th} vertex. These inclusions are depicted in Figure 8.

Definition 5.5 Call a simplicial set a *composition tiling* if it is a colimit of a diagram of the form

$$\begin{tikzcd}[column sep=small]
C_i[n,n+1] & & C_{i'}[n',n'+1] & & \cdots & & C_{i^{(k)}}[n^{(k)},n^{(k)}+1] \\
& \mathrm{Sp}[m] \arrow[ul, hook'] \arrow[ur, hook] & & \mathrm{Sp}[m'] \arrow[ul, hook'] \arrow[ur, hook] & & \mathrm{Sp}[m^{(k-1)}] \arrow[ul, hook'] \arrow[ur, hook] &
\end{tikzcd}$$

built out of the inclusions from above. (For such a diagram to make sense, we must have $n^{(j)} = n^{(j+1)} \pm 1$ and $m^{(j)} = \max(n^{(j)}, n^{(j+1)})$ for all $0 \leq j \leq k-1$.)

A composition tiling C comes with two important inclusions of spines, coming from the unused inclusions of the composition tiles on the left and right in the diagram above. These spines must start at the same vertex and end at the same vertex, and their union is precisely the outer edges of the composition tiling. We view a composition tiling as linking these two spines. For our purposes, those two spines are the crucial data to keep track of in a composition tiling, so we use $C^{r,s}$ to denote a composition tiling linking a length r spine to a length s spine.

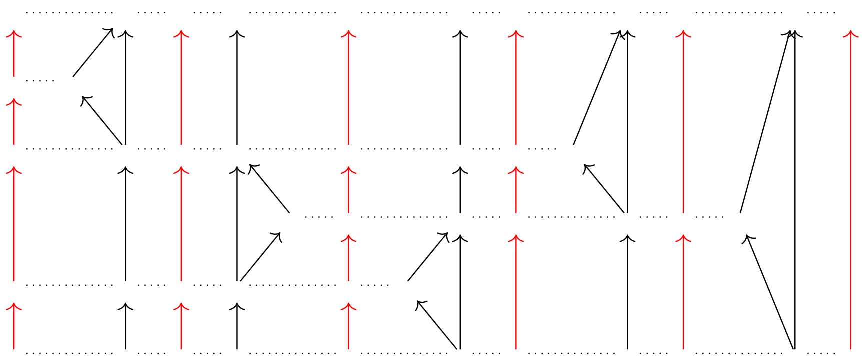
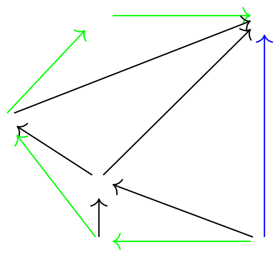


Figure 9

Example 5.6 We visualize the components of the diagram

$$\begin{tikzcd}
C_3[3,4] & & C_2[3,4] & & C_1[3,4] & & C_2[2,3] & & C_1[1,2] \\
& \mathrm{Sp}[3] \arrow[ul, hook] \arrow[ur, hook] & & \mathrm{Sp}[4] \arrow[ul, hook] \arrow[ur, hook] & & \mathrm{Sp}[3] \arrow[ul, hook] \arrow[ur, hook] & & \mathrm{Sp}[2] \arrow[ul, hook] \arrow[ur, hook] &
\end{tikzcd}$$

along with the leftmost and rightmost spines as shown in Figure 9, and then taking the colimit we get a composition tiling $C^{4,1}$



The green edges show the spine $\mathrm{Sp}[4] \hookrightarrow C^{4,1}$ and the blue edge shows $\mathrm{Sp}[1] \hookrightarrow C^{4,1}$.

Example 5.7 An unordered triangulation need not be a composition tiling. For example, in the unordered triangulation

$$\begin{tikzcd}
\bullet \arrow[r] \arrow[d] \arrow[dr] & \bullet \\
\bullet & \bullet \arrow[l] \arrow[u]
\end{tikzcd}$$

there is not a choice of precisely two spines whose union is the set of outer edges. However, every (ordered) triangulation of the $(n+1)$-gon is a composition tiling, linking the spine $0 \to 1 \to \cdots \to (n-1) \to n$ with the spine $0 \to n$. At the same time, not every composition tiling from $\mathrm{Sp}[n]$ to $\mathrm{Sp}[1]$ is a triangulation since in general composition tilings can have interior vertices.

Remark 5.8 Recall that $h\colon \mathrm{sSet} \to \mathrm{Cat}$ is the left adjoint of the nerve functor. Given a simplicial set X, we can construct hX explicitly by first letting the set of objects of hX equal the set of 0-simplices X_0. To define $\mathrm{Hom}_{hX}(x, y)$ for $x, y \in X_0$, we take the set of all maps $\mathrm{Sp}[n] \to X$ (for varying n) that start at

x and end at y and then quotient out by the equivalence relation where $f: \mathrm{Sp}[r] \to X$ is equivalent to $g: \mathrm{Sp}[s] \to X$ if there exists a composition tiling $C^{r,s}$ and a map $C^{r,s} \to X$ such that restricting along $\mathrm{Sp}[r] \hookrightarrow C^{r,s}$ is f and restricting along $\mathrm{Sp}[s] \hookrightarrow C^{r,s}$ is g. The composition functions are induced by concatenation of spines.

Although this construction of hX is nonstandard, one can check that is just another way of phrasing the more standard explicit construction given in [16].

Definition 5.9 For any $r \geq 1$ and composition tiling $C^{r,1}$, call the pushout

$$\begin{tikzcd} \mathrm{Sp}[1] = \Delta[1] \arrow[r] \arrow[d, hook] & \Delta[0] \arrow[d, hook] \\ C^{r,1} \arrow[r] & \widetilde{C}^r \end{tikzcd}$$

a *pinched tiling*. Call the inclusions $\Delta[1] \hookrightarrow \mathrm{Sp}[r] \hookrightarrow \widetilde{C}^r$ coming from the $0 \to 1$ and $(r-1) \to r$ edge inclusions $\Delta[1] \to \mathrm{Sp}[r]$ the *first edge inclusion* and *last edge inclusion*, respectively.

Example 5.10 In Example 5.6, we collapse the rightmost arrow of $C^{4,1}$ to a degeneracy to get a pinched tiling $\widetilde{C}^4$. The first edge inclusion is the bottom-most edge, and the last edge inclusion is the top-most edge in the picture.

Example 5.11 The standard 2-simplex $\Delta[2]$ is itself a composition tiling $C^{2,1}$. We collapse the $0 \to 2$ edge to get a pinched tiling $\widetilde{C}^2$, whose first edge inclusion is $0 \to 1$ and last edge inclusion is $1 \to 2$.

Since a map from a composition tiling $C^{r,1} \to X$ is capturing that the restriction to $\mathrm{Sp}[r] \to X$ and to $\Delta[1] \to X$ correspond to the same morphism in hX, we can see that a map from the respective pinched tiling $\widetilde{C}^r \to X$ is capturing that the restriction to $\mathrm{Sp}[r] \to X$ becomes the identity in hX. In particular, the first edge inclusion of a pinched tiling (the $0 \to 1$ edge of the spine $\mathrm{Sp}[r]$) has a left inverse (coming from the $1 \to 2 \to \cdots \to r$ edges), and the last edge inclusion (the $(r-1) \to r$ edge) has a right inverse $(0 \to 1 \to \cdots \to (r-1))$. In fact, by Remark 5.8, an edge $\Delta[1] \to X$ has a right or left inverse in hX if and only if it extends along a pinched tiling. We record this observation as a lemma.

Lemma 5.12 *Given an edge $e: \Delta[1] \to X$, the morphism $h(e)$ has left (right) inverse in hX if and only if there exists a pinched tiling $\widetilde{C}^r$ and a map $\widetilde{C}^r \to X$ that restricts to e along the first (last) edge inclusion.*

Since we can use pinched tilings to identify edges of a simplicial set X that have left or right inverses, we can use a pushout of pinched tilings to identify edges that have both inverses.

Definition 5.13 Given two pinched tilings $\widetilde{C}^r$ and $(\widetilde{C}')^s$, let T be the pushout

$$\begin{tikzcd} \Delta[1] \arrow[r, hook, "\text{first}"] \arrow[d, hook, "\text{last}"'] & (\widetilde{C}')^s \arrow[d, hook] \\ \widetilde{C}^r \arrow[r, hook] & T \end{tikzcd}$$

Call T an *inverting tiling*, and let e_T denote the diagonal composite map $\Delta[1] \hookrightarrow T$. Call e_T the *inverting inclusion* of T.

The following proposition shows that categorical preisomorphisms are characterized by maps out of inverting tilings.

Proposition 5.14 *An edge $e\colon \Delta[1] \to X$ is a categorical preisomorphism if and only if there exists an inverting tiling T such that e extends along the inclusion e_T:*

$$\begin{tikzcd} \Delta[1] \arrow[r, "e"] \arrow[d, "e_T"'] & X \\ T \arrow[ur, dotted] & \end{tikzcd}$$

Proof By Lemma 5.12, the morphism $h(e)$ has a left and a right inverse if and only if there exist two pinched tilings such that e extends along the first inclusion of one and the last inclusion of the other, which happens if and only if e extends along the pushout of those inclusions. □

Example 5.15 The simplicial set K from Example 4.5 is an inverting tiling built out of the pinched tiling $\widetilde{C}^2$.

This set of inverting tilings $\{T\}$ characterizes which edges of a simplicial set we want to think of as invertible, in the context of the Joyal model structure. So, in the spirit of Section 4, let us consider the set of T-augmented horn inclusions from Definition 3.3.

Definition 5.16 Let $\mathscr{E} = \{\Delta[1] \hookrightarrow T\}$ be the set of all inverting inclusions into inverting tilings. Given $\Delta[1] \hookrightarrow T$ in $\mathscr{E}$, we say that a T-augmented horn inclusion is a *special horn inclusion*. If the horn is outer, we say that it is a *special outer horn inclusion*. Let SpHorn be the set of all special horn inclusions and let SpOH be the set of all special outer horn inclusions.

Remark 5.17 The sets SpHorn and SpOH are countable by Remark 3.26.

Recall that the standard phrasing of the special outer horn lifting property of quasicategories is that there exist lifts of outer horns so long as they satisfy the additional property that a certain edge is sent to a categorical preisomorphism. We can now rephrase this condition directly as a lifting condition with respect to the set of special horn inclusions.

Proposition 5.18 *If Q is a quasicategory, then $Q \to *$ has the right lifting property with respect to* SpOH.

Proof All inner special horns are pushouts of ordinary inner horns, so it suffices just to check outer special horns. By symmetry, it suffices to consider an inverting inclusion $\Delta[1] \hookrightarrow T$ and a special horn map $\Lambda^0[n]^T_{0\to 1} \to Q$. Because the edge $0 \to 1$ factors through $T \to Q$, it is sent to a categorical preisomorphism in Q, and so by the special horn lifting property, we get an extension of the horn $\Lambda^0[n] \to Q$ to $\Delta[n] \to Q$, inducing a lift of the original special horn map $\Delta[n]^T_{0\to 1} \to Q$. □

We now turn to constructing the special horn model structure using Cisinski's theory.

Lemma 5.19 *The class generated by the set of special horn inclusions satisfies axiom* (An2′) *from Lemma 2.8.*

Proof Apply Lemma 4.14 where $I' = T$, the inverting tiling for a given special horn. □

Corollary 5.20 *The class generated by* $\mathrm{SpHorn} \cup \{\{\varepsilon\} \hookrightarrow K\}$ *together with the set of almost-K-augmented horn inclusions is* $(\kappa \odot -)$*-anodyne.*

Proof We already knew from Proposition 4.7 that axiom (An1′) is satisfied, as well as (An2′) for $\{\{\varepsilon\} \hookrightarrow K\}$ together with the set of almost-K-augmented horn inclusions. Lemma 5.19 tells us that (An2′) is satisfied for the remaining maps. □

Theorem 5.21 *There is a Cisinski model structure on* sSet *whose fibrant objects are the simplicial sets X such that $X \to *$ has the right lifting property with respect to the set of special horn inclusions.*

Proof By Theorem 2.10, we get a Cisinski model structure from Corollary 5.20 whose fibrant objects are those with lifts against $\mathrm{SpHorn} \cup \{\{\varepsilon\} \hookrightarrow K\}$ as well as the set of K-augmented horn inclusions. We claim that simply knowing $X \to *$ has lifts of special horn inclusions is enough to conclude that X is fibrant in this model structure. We first note that $X \to *$ has the right lifting property with respect to $\{\{\varepsilon\} \hookrightarrow K\}$ for all simplicial sets X. Now, note that K is itself an inverting tiling, so if $X \to *$ has lifts of special horn inclusions, it in particular has lifts of K-augmented horn inclusions, so by Corollary 4.25 we see that $X \to *$ has lifts of all almost-K-augmented horn inclusions. □

Definition 5.22 We call the model structure in Theorem 5.21 the *special horn model structure*. We say a simplicial set is *special horn fibrant* if it is fibrant in this model structure.

Remark 5.23 The special outer horn lifting property of quasicategories implies that the Joyal model structure is a localization of the special horn model structure. While these model structures have a close relationship in sharing a notion of "invertible edges", they are distinct because $\Lambda^1[2]$ is fibrant in the special horn model structure but not in the Joyal model structure.

The fact that K is itself an inverting tiling means that every special horn fibrant simplicial set has lifts of K-augmented horns. Therefore, the special horn model structure is a localization of the K-minimal homotopically behaved model structure. These model structures are also distinct; if T is as in Example 5.1, then the special horns $\Lambda^i[n]^T_{i\to i+1}$ are fibrant in the K-minimal homotopically behaved model structure but not in the special horn model structure.

The special horn model structure is therefore a curious intermediate between the K-minimal homotopically behaved model structure and the Joyal model structure. The fibrant objects are very similar to those of the K-minimal homotopically behaved model structure, making it tempting to claim it as a model structure with "the homotopical properties of quasicategories without the composition aspects". However, compositionality actually does play a subtle but key role in determining the notion of homotopy for the special horn model structure.

We conclude this section by conjecturing a partial characterization of the trivial cofibrations in the Joyal model structure.

Conjecture 5.24 *The class of trivial cofibrations in the Joyal model structure that are bijective-on-0-simplices is generated by the set of inner horn inclusions together with the set of special outer horn inclusions. That is, the bijective-on-0-simplices trivial cofibrations are precisely* ${}^{\boxslash}((\mathrm{IH} \cup \mathrm{SpOH})^{\boxslash})$.

Remark 5.25 The special outer horn inclusions are weak equivalences in the special horn model structure, and so are also Joyal weak equivalences. The uncertain aspect of the conjecture is whether the containment of ${}^{\boxslash}((\mathrm{IH} \cup \mathrm{SpOH})^{\boxslash})$ in the class of bijective-on-0-simplices trivial cofibrations in the Joyal model structure is strict.

Joyal [13] left open whether the inner horn inclusions alone generated the bijective-on-0-simplices trivial cofibrations in his model structure, but Campbell [1] recently provided a counterexample of a map that is bijective-on-0-simplices and a weak equivalence in the Joyal model structure but is not in ${}^{\boxslash}(\mathrm{IH}^{\boxslash})$. Since Campbell's map is in fact a pushout of a special 2-horn, it is not a counterexample to Conjecture 5.24.

An intuitive argument for Conjecture 5.24 is that this set of maps seems as close as possible to the set of ordinary horn inclusions (which generate the trivial cofibrations of the Kan–Quillen model structure on sSet) while still being weak equivalences in the Joyal model structure.

We also state a similar conjecture for the special horn model structure.

Conjecture 5.26 *The class of bijective-on-0-simplices trivial cofibrations in the special horn model structure is generated by the set of special horn inclusions. That is, the bijective-on-0-simplices trivial cofibrations are precisely* ${}^{\boxslash}(\mathrm{SpHorn}^{\boxslash})$.

6 Comparing model structures

In this section, we compare the fibrant objects in the minimal model structure to the fibrant objects of homotopically behaved model structures to get a better understanding of what it means to be homotopically behaved. We begin by explaining the horn-based characterization of the minimal model structure's fibrant objects from [6].

Definition 6.1 Fix $n \geq 1$ and $0 \leq i \leq n-1$. Let $[n]_i$ denote the category $[n]$ with the morphism $c_i \to c_{i+1}$ inverted,

$$c_0 \to c_1 \to \cdots \to c_{i-1} \to c_i \rightleftarrows c_{i+1} \to c_{i+2} \to \cdots \to c_{n-1} \to c_n,$$

and let $\overline{\nabla}_i[n]$ denote the nerve of $[n]_i$. Call $\overline{\nabla}_i[n]$ an *n-isoplex*, or simply an *isoplex*.

Let $d_j \overline{\nabla}_i[n]$ denote the nerve of the full subcategory of $[n]_i$ that includes all but the j^{th} vertex. Call $d_j \overline{\nabla}_i[n]$ the *j^{th} face* of $\overline{\nabla}_i[n]$.

For $j \neq i, i+1$, the d_j face of the isoplex $\overline{\nabla}_i[n]$ is an $(n-1)$-isoplex, while the d_i and d_{i+1} faces are standard $(n-1)$-simplices. We can therefore think of the n-isoplex as an "isomorphism of $(n-1)$-simplices" between its d_{i+1} face and its d_i face.

Having defined our "iso" analogue of simplices and faces, we can now define "iso-horns" to be the union of all but one face of an isoplex. However, just like we saw with augmented horns, we want to limit ourselves to horns that omit the d_k or d_{k+1} face, where $k \to k+1$ is a J-edge. Furthermore, due to the symmetry of isoplexes, it suffices just to consider horns where the d_k face is missing.

Definition 6.2 Let $n \geq 1$ and $0 \leq i \leq n-1$ be as in Definition 6.1. Let $\mathbb{V}_i[n]$ be the union of all but the i^{th} face of $\overline{\nabla}_i[n]$. We call $\mathbb{V}_i[n]$ an *iso-horn*, and call the inclusion $\mathbb{V}_i[n] \hookrightarrow \overline{\nabla}_i[n]$ an *iso-horn inclusion*.

We can now state the main result of [6].

Theorem 6.3 *A simplicial set X is fibrant in the minimal model structure if and only if it has lifts of iso-horn inclusions.*

The takeaway of this theorem is that, from a certain perspective, isoplexes are the fundamental building blocks of homotopies in the minimal model structure, and hence in any Cisinski model structure. They are inherently equipped with all the higher invertibility data we want from a good notion of homotopy.

To understand what is happening when we localize to a homotopically behaved model structure, consider the diagram

$$\begin{tikzcd} \Delta[n]^J_{i\to i+1} \arrow[r, hook] \arrow[dr] & \overline{\nabla}_i[n] \arrow[d, "\sim"] \\ & \Delta[n-1] \end{tikzcd}$$

Since the map on the right is a weak equivalence in any Cisinski model structure, by localizing with respect to the maps on the left to yield a homotopically behaved model structure, we are equivalently localizing with respect to the horizontal maps. These maps being weak equivalences is effectively saying that "n-simplices with a J-edge along $i \to (i+1)$ extend to full-fledged homotopies of $(n-1)$-simplices". In other words, we are justified in considering a n-simplex with a J-edge in its spine to be a "homotopy" since all of the higher invertibility data comes along for free.

$$\begin{tikzcd}
\mathrm{Min}(\varnothing) \arrow[r] \arrow[d] & \mathrm{Min}(\mathrm{HtpyBehaved}) \arrow[d] & & \\
\mathrm{Min}(K \to *) \arrow[r] & \mathrm{Min}(K, \mathrm{HtpyBehaved}) \arrow[r] & \mathrm{Special} \arrow[r] & \mathrm{Joyal}
\end{tikzcd}$$

Figure 10

We conclude by reviewing the broader picture of Cisinski model structures on sSet that localize to the Joyal model structure. Figure 10 shows the model structures discussed in this paper, with an arrow drawn to indicate that the target is a localization of the source. (The minimal model structure is denoted by $\mathrm{Min}(\varnothing)$, and its localization with respect to $K \to *$ by $\mathrm{Min}(K \to *)$. The minimal and K-minimal homotopically behaved model structures are depicted in the second column.) By Remark 5.23, the two localizations on the right are nontrivial. The vertical arrows indicate nontrivial localizations because $K \to *$ is not a weak equivalence in the two upper model structures by Remark 4.32. The upper horizontal arrow indicates a nontrivial localization because the simplicial set $\Delta[2]^*_{0\to1}$ is fibrant in the minimal model structure but not in the minimal homotopically behaved model structure. The remaining localization is nontrivial by the following lemma.

Lemma 6.4 *Given a simplicial set I with exactly two vertices a and b and with 1-simplices $a \to b$ and $b \to a$, the localization of the minimal model structure at the map $I \to *$ is not homotopically behaved.*

Proof Since the map $\Delta[2]^*_{0\to1} \to *$ is a weak equivalence but not a trivial fibration in a homotopically behaved model structure, it cannot be a fibration. So, it suffices to show that $\Delta[2]^*_{0\to1}$ is fibrant in the minimal model structure localized at $I \to *$. A similar argument as in [6, Section 3] shows that a simplicial set is fibrant in this model structure if and only if it has lifts with respect to all maps in the set

$$\mathscr{A}_I = \{(I \times \partial\Delta[n]) \cup (\{v\} \times \Delta[n]) \hookrightarrow I \times \Delta[n]\}_{v\in\{a,b\}, n\geq 0}.$$

We therefore see that $\Delta[2]^*_{0\to1}$ is fibrant by observing that any map

$$(I \times \partial\Delta[n]) \cup (\{v\} \times \Delta[n]) \to \Delta[2]^*_{0\to1}$$

either factors through the collapse map $(I \times \partial\Delta[n]) \cup (\{v\} \times \Delta[n]) \to \Delta[n]$ or through $\Delta[1] \hookrightarrow \Delta[2]^*_{0\to1}$ because the 1-simplices of I go in opposite directions. □

$$\begin{tikzcd}
\mathrm{Min}(\varnothing) \arrow[r] \arrow[d] & \mathrm{Min}(\mathrm{HtpyBehaved}) \arrow[d] & & \\
\mathrm{Min}(K' \to *) \arrow[r] \arrow[d] & \mathrm{Min}(K', \mathrm{HtpyBehaved}) \arrow[r] \arrow[d] & \text{2-Special} \arrow[r] \arrow[d] & \text{Quasi-2-Seg} \arrow[d] \\
\mathrm{Min}(K \to *) \arrow[r] & \mathrm{Min}(K, \mathrm{HtpyBehaved}) \arrow[r] & \mathrm{Special} \arrow[r] & \mathrm{Joyal}
\end{tikzcd}$$

Figure 11

In Figure 11, we indicate how we expect the 2-Segal analogue of these model structures to fit in. In particular, it appears likely that there is a separate simplicial set K' that plays the same role for the 2-Segal situation as K does for quasicategories. It also seems likely that there be a notion of 2-Segal preisomorphism that is distinct from categorical preisomorphisms, and hence a notion of 2-special horns that is distinct from that of special horns. These model structures are the topic of [7].

References

[1] **A Campbell**, *A counterexample in quasi-category theory*, Proc. Amer. Math. Soc. 148 (2020) 37–40 MR Zbl

[2] **A Campbell**, **E Lanari**, *On truncated quasi-categories*, Cah. Topol. Géom. Différ. Catég. 61 (2020) 154–207 MR Zbl

[3] **D-C Cisinski**, *Les préfaisceaux comme modèles des types d'homotopie*, Astérisque 308, Soc. Math. France, Paris (2006) MR Zbl

[4] **D-C Cisinski**, *Higher categories and homotopical algebra*, Cambridge Stud. Adv. Math. 180, Cambridge Univ. Press (2019) MR Zbl

[5] **T Dyckerhoff**, **M Kapranov**, *Higher Segal spaces*, Lecture Notes in Math. 2244, Springer (2019) MR Zbl

[6] **M Feller**, *A horn-like characterization of the fibrant objects in the minimal model structure on simplicial sets*, High. Struct. 7 (2023) 166–181 MR Zbl

[7] **M Feller**, *Quasi-2-Segal sets*, Tunis. J. Math. 5 (2023) 327–367 MR Zbl

[8] **I Gálvez-Carrillo**, **J Kock**, **A Tonks**, *Decomposition spaces, incidence algebras and Möbius inversion, I: Basic theory*, Adv. Math. 331 (2018) 952–1015 MR Zbl

[9] **P G Goerss**, **J F Jardine**, *Simplicial homotopy theory*, Progr. Math. 174, Birkhäuser, Basel (1999) MR Zbl

[10] **P S Hirschhorn**, *Model categories and their localizations*, Math. Surv. Monogr. 99, Amer. Math. Soc., Providence, RI (2003) MR Zbl

[11] **M Hovey**, *Model categories*, Math. Surv. Monogr. 63, Amer. Math. Soc., Providence, RI (1999) MR Zbl

[12] **A Joyal**, *Quasi-categories and Kan complexes*, J. Pure Appl. Algebra 175 (2002) 207–222 MR Zbl

[13] **A Joyal**, *Notes on quasi-categories*, preprint (2008) Available at `https://www.math.uchicago.edu/~may/IMA/Joyal.pdf`

[14] **A Joyal**, *The theory of quasi-categories and its applications*, *II*, lecture notes, CRM (2008) Available at `https://mat.uab.cat/~kock/crm/hocat/advanced-course/Quadern45-2.pdf`

[15] **D G Quillen**, *Homotopical algebra*, Lecture Notes in Math. 43, Springer (1967) MR Zbl

[16] **C Rezk**, *Introduction to quasicategories*, lecture notes, University of Illinois (2022) Available at `https://rezk.web.illinois.edu/quasicats.pdf`

Department of Mathematics, University of Virginia
Charlottesville, VA, United States

`feller@virginia.edu`

Received: 23 November 2021 Revised: 22 May 2023

Algebraic & Geometric Topology **25**:1 (2025) 399–488
DOI: 10.2140/agt.2025.25.399
Published: 24 March 2025

Quasiconvexity of virtual joins and separability of products in relatively hyperbolic groups

ASHOT MINASYAN
LAWK MINEH

A relatively hyperbolic group G is said to be QCERF if all finitely generated relatively quasiconvex subgroups are closed in the profinite topology on G.

Assume that G is a QCERF relatively hyperbolic group with double coset separable (eg virtually polycyclic) peripheral subgroups. Given any two finitely generated relatively quasiconvex subgroups $Q, R \leqslant G$ we prove the existence of finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ such that the join $\langle Q', R' \rangle$ is again relatively quasiconvex in G. We then show that, under the minimal necessary hypotheses on the peripheral subgroups, products of finitely generated relatively quasiconvex subgroups are closed in the profinite topology on G. From this we obtain the separability of products of finitely generated subgroups for several classes of groups, including limit groups, Kleinian groups and balanced fundamental groups of finite graphs of free groups with cyclic edge groups.

20E26, 20F65, 20F67; 20H10

1 Introduction

Any group can be equipped with the *profinite topology*, whose basic open sets are cosets of finite-index subgroups. A subset of a group is said to be *separable* if it is closed in the profinite topology. The trivial subgroup of a group G is separable if and only if the profinite topology is Hausdorff; in this case G is said to be *residually finite*. If every finitely generated subgroup of G is separable then G is called *LERF* (or *subgroup separable*), and if the product of any two finitely generated subgroups is separable, G is said to be *double coset separable*.

In this paper we will be interested in various separability properties of relatively hyperbolic groups. The notion of a relatively hyperbolic group was originally suggested by Gromov [25] as a generalisation of word hyperbolic groups. The concept was further developed by Farb [20], Bowditch [8], Druțu and Sapir [17], Osin [46], and Groves and Manning [26], whose various definitions were later shown to be equivalent by Hruska [30]. Relative hyperbolicity is a relative property of a group G in the sense that one must specify a collection of *peripheral subgroups* $\{H_\nu \mid \nu \in \mathcal{N}\}$ with respect to which G is relatively hyperbolic (see Definition 5.3). Typical examples of relatively hyperbolic groups include geometrically finite Kleinian groups, fundamental groups of finite-volume manifolds of pinched negative curvature, and small cancellation quotients of free products. Respectively, these groups are hyperbolic relative to their maximal parabolic subgroups, their cusp subgroups and the images of the free factors (see, for example, [46]).

1.1 Quasiconvexity of virtual joins

Since general finitely generated subgroups of word hyperbolic (relatively hyperbolic) groups can be quite wild and need not be separable, it is customary to restrict one's attention to quasiconvex (respectively, relatively quasiconvex) subgroups.

Quasiconvex subgroups play a central role in the study of word hyperbolic groups. They are precisely the finitely generated quasi-isometrically embedded subgroups, and, hence, they are hyperbolic themselves and are generally well behaved.

If Q and R are two quasiconvex subgroups of a hyperbolic group G then the intersection $S = Q \cap R$ is also quasiconvex (see, for example, Short [57]) but the join $\langle Q, R\rangle$ need not be. This can be remedied by considering a *virtual join* of Q and R, which is defined as $\langle Q', R'\rangle$, for some finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$. The existence of a quasiconvex virtual join $\langle Q', R'\rangle$ was proved by Gitik [23] under the assumption that $S = Q \cap R$ is separable in G. More precisely, Gitik's theorem states that there exist finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ such that $Q' \cap R' = S$ and the virtual join $\langle Q', R'\rangle$ is quasiconvex in G; moreover, $\langle Q', R'\rangle$ will be naturally isomorphic to the amalgamated free product $Q' *_S R'$. This theorem was an important ingredient in the proof that double cosets of quasiconvex subgroups are separable in LERF hyperbolic groups (see [22; 43]).

In the setting of relatively hyperbolic groups, the natural subobjects are the *relatively quasiconvex subgroups*, which are themselves relatively hyperbolic in a way that is compatible with the ambient group. Basic examples of relatively quasiconvex subgroups are *maximal parabolic subgroups* (that is, conjugates of the peripheral subgroups), *parabolic subgroups* (subgroups of maximal parabolics) and finitely generated undistorted (equivalently, quasi-isometrically embedded) subgroups (see [30]).

In [30], Hruska proved that the intersection of two relatively quasiconvex subgroups is again relatively quasiconvex. However, until now the existence of a relatively quasiconvex virtual join $\langle Q', R'\rangle$, for two relatively quasiconvex subgroups Q and R in a relatively hyperbolic group G, such that $S = Q \cap R$ is separable in G, was only known in special cases:

- Martínez-Pedroza [37] proved it in the case when $R \leqslant P$, for some maximal parabolic subgroup P of G, such that $Q \cap P \subseteq R$;
- Martínez-Pedroza and Sisto [38] proved it when Q and R have *compatible parabolics* (that is, for every maximal parabolic subgroup P of G either $Q \cap P \subseteq R \cap P$ or $R \cap P \subseteq Q \cap P$);
- Yang [60] (unpublished; see also McClellan's thesis [40]) proved it when R is a *full subgroup* of G (that is, for every maximal parabolic subgroup P in G, $R \cap P$ is either finite or has finite index in P).

Similarly to Gitik's theorem [23], in all three cases above the authors establish an isomorphism between the virtual join $\langle Q', R'\rangle$ and the amalgamated free product $Q' *_{S'} R'$, where $S' = Q' \cap R' \leqslant_f S$.

The extra assumptions on Q and R in each of the above results from [37; 38; 40; 60] imply that Q and R have *almost compatible parabolics* (see Definition 1.5 below). Unfortunately this is still a significant restriction and a more general result is desirable. Moreover, in the absence of almost compatibility one cannot expect a virtual join to split as an amalgamated free product of Q' and R'. Indeed, for example if both Q and R are subgroups of an abelian peripheral subgroup of G then any virtual join $\langle Q', R'\rangle$ would again be abelian.

One of the goals of the present paper is to establish quasiconvexity of virtual joins without making any compatibility assumptions on Q and R. However we need to impose stronger assumptions on the properties of the profinite topology on G than just separability of $S = Q \cap R$: we will require the finitely generated relatively quasiconvex subgroups to be separable and the peripheral subgroups to be double coset separable.

Definition 1.1 (QCERF) We will say that a relatively hyperbolic group G is *QCERF* if every finitely generated relatively quasiconvex subgroup in G is separable.

Theorem 1.2 *Let G be a finitely generated relatively hyperbolic group. Suppose that G is QCERF and the peripheral subgroups of G are double coset separable. If $Q, R \leqslant G$ are finitely generated relatively quasiconvex subgroups and $S = Q \cap R$ then there exist finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$, with $Q' \cap R' = S$, such that the virtual join $\langle Q', R' \rangle$ is relatively quasiconvex in G.*

More precisely, there exists $L \leqslant_f G$, with $S \subseteq L$, such that for any $L' \leqslant_f L$, satisfying $S \subseteq L'$, we can choose $Q' = Q \cap L' \leqslant_f Q$, and there exists $M \leqslant_f L'$, with $Q' \subseteq M$, such that for any $M' \leqslant_f M$, satisfying $Q' \subseteq M'$, we can choose $R' = R \cap M' \leqslant_f R$.

One can observe that the choice of $R' \leqslant_f R$ in the above theorem depends on the choice of $Q' \leqslant_f Q$. In the case when the peripheral subgroups are abelian the situation is easier:

Theorem 1.3 *Let G be a finitely generated group hyperbolic relative to a finite collection of abelian subgroups. Assume that G is QCERF. If $Q, R \leqslant G$ are relatively quasiconvex subgroups and $S = Q \cap R$ then there exists a finite-index subgroup $L \leqslant_f G$, with $S \subseteq L$, such that the virtual join $\langle Q', R' \rangle$ is relatively quasiconvex in G, for arbitrary subgroups $Q' \leqslant_f Q \cap L$ and $R' \leqslant_f R \cap L$, satisfying $Q' \cap R' = S$.*

In fact, one can slightly weaken the assumptions in Theorem 1.3 by requiring the peripheral subgroups of G to be virtually abelian instead of abelian; see Corollary 14.2.

Unlike the previous results from [38; 60], Theorem 1.2 does not require any (almost) compatibility of parabolics from the subgroups Q and R. To work in this general setting, we develop a novel approach which uses the profinite topology on G to carefully select the finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ satisfying certain metric properties (see Sections 3.1, 3.2 and 11). We also give a new and simple criterion for establishing separability of double cosets in amalgamated free products in Section 12.

Theorem 1.2 applies to a wide class of relatively hyperbolic groups, including all limit groups, all Kleinian groups and many groups acting on CAT(0) cube complexes. Regarding QCERF-ness, Manning and Martínez-Pedroza [36] proved that the following two statements are equivalent:

(a) every finitely generated group hyperbolic relative to a finite collection of LERF and slender subgroups is QCERF;

(b) all word hyperbolic groups are residually finite.

Recall that a group is called *slender* if every subgroup is finitely generated. The question of whether statement (b) is true is a well-known open problem. If the answer to it is positive then, for example, all finitely generated groups hyperbolic relative to virtually polycyclic subgroups will be QCERF.

Large classes of relatively hyperbolic groups have already been proved to be QCERF. One of the first results in this direction is due to Wilton [58], who established QCERF-ness of limit groups. The groundbreaking work of Haglund and Wise [28] and Agol [2] implies that any word hyperbolic group acting geometrically on a CAT(0) cube complex is QCERF. One of the consequences of this result is that all finitely generated Kleinian groups are QCERF. More recently, Einstein and Groves [18] and Groves and Manning [27] extended this theory to relatively hyperbolic groups acting (weakly) relatively geometrically on CAT(0) cube complexes. Einstein and Ng [19] used it to show that full relatively quasiconvex subgroups of $C'(1/6)$-small cancellation quotients of free products of residually finite groups are separable. In the case when the free factors are LERF and slender the latter result can be combined with a theorem of Manning and Martínez-Pedroza [36, Theorem 1.7] to conclude that such small cancellation free products are QCERF.

By a theorem of Lennox and Wilson [33] all virtually polycyclic groups are double coset separable; hence the assumption about peripheral subgroups in Theorem 1.2 is automatically true in many relevant cases. However whether this assumption is actually necessary is less obvious. It is required in our approach, but it would be interesting to see whether the theorem remains valid without it. As expected from the results in [38; 60], it is not needed if the relatively quasiconvex subgroups Q and R have almost compatible parabolics; see Theorem 14.5 below.

1.2 Separability of double cosets

In group theory, knowing that double cosets of certain subgroups are separable is often quite useful. For example, the separability of double cosets of hyperplane subgroups was used by Haglund and Wise in [28] to give a criterion for virtual specialness of a compact non-positively curved cube complex. Separability of double cosets of abelian subgroups in Kleinian groups was an important ingredient in the theorem of Hamilton, Wilton and Zalesskii [29] that fundamental groups of compact orientable 3-manifolds are conjugacy separable.

Double coset separability of free groups was first proved by Gitik and Rips [24]. Shortly after, Niblo [44] came up with a new criterion for separability of double cosets and applied it to show that finitely generated Fuchsian groups and fundamental groups of Seifert-fibred 3-manifolds are double coset separable. Separability of double cosets of quasiconvex subgroups in QCERF word hyperbolic groups was proved by the first author in [43]. Martínez-Pedroza and Sisto [38] generalised this to double cosets of relatively quasiconvex subgroups with compatible parabolics in QCERF relatively hyperbolic groups; Yang [60] and McClellan [40] treated the case when at least one of the factors is full. Our proof of Theorem 1.2 almost immediately yields the following.

Corollary 1.4 *Let G be a finitely generated group hyperbolic relative to a finite collection of subgroups $\{H_\nu \mid \nu \in \mathcal{N}\}$. Suppose that G is QCERF and H_ν is double coset separable, for every $\nu \in \mathcal{N}$. Then for all finitely generated relatively quasiconvex subgroups $Q, R \leqslant G$, the double coset QR is separable in G.*

Clearly the assumptions of Corollary 1.4 are the minimal possible. This result is powerful enough to prove a conjecture of Hsu and Wise from [31]; see Corollary 2.3.

In the case when the relatively hyperbolic group G admits a weakly relatively geometric action on a CAT(0) cube complex, Corollary 1.4 was proved by Groves and Manning [27]. Groves and Manning's argument uses Dehn fillings to approximate G by QCERF word hyperbolic groups; thus reducing the statement to separability of double cosets in hyperbolic groups from [43]. Our approach is completely different as we always work within G.

In the following definition we will use a preorder $\preccurlyeq$ on the sets of subsets of a group G, introduced by the first author in [42]:

given $U, V \subseteq G$ we will write $U \preccurlyeq V$ if there exists a finite subset $Y \subseteq G$ such that $U \subseteq VY$.

If d_X is the word metric on G, corresponding to a finite generating set X, and U and V are subsets of G then $U \preccurlyeq V$ if and only if U is contained in a finite d_X-neighbourhood of V. If U and V are subgroups of G then $U \preccurlyeq V$ is equivalent to $|U : (U \cap V)| < \infty$ (see [42, Lemma 2.1]).

Definition 1.5 (almost compatible parabolics) Let Q and R be subgroups of a relatively hyperbolic groups G. We will say that Q and R have *almost compatible parabolics* if for every maximal parabolic subgroup P of G either $Q \cap P \preccurlyeq R \cap P$ or $R \cap P \preccurlyeq Q \cap P$.

Clearly if G is a relatively hyperbolic group and Q and R are subgroups with compatible parabolics then they have almost compatible parabolics. The same is true if at least one of Q or R is a full subgroup of G.

In the case when the relatively quasiconvex subgroups Q and R have almost compatible parabolics, the assumption that the peripheral subgroups H_ν are double coset separable can be dropped from Corollary 1.4, allowing us to recover the double coset separability results from [38; 40; 60].

Corollary 1.6 *Suppose that G is a finitely generated QCERF relatively hyperbolic group. If Q and R are finitely generated relatively quasiconvex subgroups of G with almost compatible parabolics then the double coset QR is separable in G.*

1.3 Separability of products of quasiconvex subgroups

The third part of this paper is dedicated to proving separability for more general products $F_1 \cdots F_s$, where $s \in \mathbb{N}$ is arbitrary and $F_1, \dots, F_s$ are relatively quasiconvex subgroups in a relatively hyperbolic group.

Definition 1.7 (RZ_s and product separability) Let P be a group and let $s \in \mathbb{N}$. We say that P has property RZ_s if for arbitrary finitely generated subgroups $E_1, \dots, E_s \leqslant P$ the product $E_1 \cdots E_s$ is separable in P. If P has property RZ_s for all $s \in \mathbb{N}$, we say that P is *product separable*.

Thus RZ_1 means that the group is LERF and RZ_2 is equivalent to double coset separability. The definition of RZ_s is due to Coulbois [14]; he named it after Ribes and Zalesskii, who proved in [53] that free groups are product separable, confirming a conjecture of Pin and Reutenauer from [49]. Pin and Reutenauer showed that product separability of free groups implies Rhodes' type II conjecture from semigroup theory (see [48; 49] for the background).

In [43], generalising the result of [53], the first author proved that the product of finitely many quasiconvex subgroups is separable in a QCERF word hyperbolic group. Moreover, in [14] Coulbois showed that, for every $s \in \mathbb{N}$, free products of groups with property RZ_s also have property RZ_s. Taken together, these facts motivate the following theorem.

Theorem 1.8 *Let G be a finitely generated group hyperbolic relative to a finite collection of subgroups $\{H_\nu \mid \nu \in \mathcal{N}\}$, and let $s \in \mathbb{N}$. Suppose that G is QCERF and H_ν has property RZ_s, for each $\nu \in \mathcal{N}$. If $F_1, \dots, F_s \leqslant G$ are finitely generated relatively quasiconvex subgroups of G, then the product $F_1 \cdots F_s$ is separable in G.*

We note that separability of products of full relatively quasiconvex subgroups in a QCERF relatively hyperbolic group was proved by McClellan [40].

Finitely generated virtually abelian groups are product separable. Therefore, Theorem 1.8 applies to finitely generated QCERF relatively hyperbolic groups with virtually abelian peripheral subgroups. Examples of such groups include limit groups, geometrically finite Kleinian groups and $C'(1/6)$-small cancellations quotients of free products of finitely generated virtually abelian groups (see [51]). We discuss some applications of Theorem 1.8 in Section 2.2, and give a brief outline of the proof at the beginning of Part III.

Acknowledgements

The authors would like to thank Pavel Zalesskii, Sam Shepherd and Benjamin Steinberg for helpful discussions, and the referee for careful reading of the paper and for valuable corrections.

2 Applications

In this section we list some applications of the main results from the introduction.

2.1 Geometrically finite virtual joins

A *Kleinian group* is a discrete subgroup of the (orientation-preserving) isometries of the real hyperbolic 3-space, $\mathrm{Isom}(\mathbb{H}^3)$. Recall that a Kleinian group G has an induced action on the ideal boundary $\partial\mathbb{H}^3$ of hyperbolic space by homeomorphisms, under which the smallest G-invariant compact subset, ΛG, is called its *limit set*. A subgroup $P \leqslant G$ is called *parabolic* if it has a single fixed point p in $\partial\mathbb{H}^3$ and setwise fixes some horosphere centred at p. We say that G is *geometrically finite* if every point of ΛG is either a conical limit point or a bounded parabolic point (see [7] for definitions). Examples of geometrically finite Kleinian groups include the fundamental groups of finite-volume hyperbolic 3-manifolds.

As noted in the introduction, geometrically finite Kleinian groups are relatively hyperbolic with respect to conjugacy class representatives of their maximal parabolic subgroups (which are virtually abelian). Moreover, geometrically finite subgroups are exactly the relatively quasiconvex subgroups of geometrically finite Kleinian groups [30, Corollary 1.6].

Baker and Cooper [5] showed, using geometric methods, that if G is a finitely generated Kleinian group and Q and R are geometrically finite subgroups of G with almost compatible parabolics, then there are finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ such that the join $\langle Q', R' \rangle$ is geometrically finite. In [38] Martínez-Pedroza and Sisto recover this result for geometrically finite Kleinian groups as a special case of their work, using techniques closer to those in the present paper. Using Theorem 1.2, we are able to eliminate the hypothesis of compatible parabolic subgroups in these results:

Corollary 2.1 *Let G be a geometrically finite Kleinian group, and suppose that $Q, R \leqslant G$ are geometrically finite subgroups of G. Then there are finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ such that $\langle Q', R' \rangle$ is a geometrically finite subgroup of G.*

Proof The group G is geometrically finite, so it is finitely generated [50, Theorem 12.4.9] and hyperbolic relative to a finite collection of finitely generated virtually abelian subgroups [8; 30]. Agol proved that all finitely generated Kleinian groups are LERF [2, Corollary 9.4]; in particular, this means that they are QCERF. Therefore G is a QCERF relatively hyperbolic group with double coset separable peripheral subgroups. By Hruska's result [30, Corollary 1.6], a subgroup of G is geometrically finite if and only if it is relatively quasiconvex. We may now apply Theorem 1.2 to obtain the desired conclusion. □

2.2 Product separability

Recall that a group G is product separable if the product of finitely many finitely generated subgroups is closed in the profinite topology on G. Until now, few examples of groups were known to be product separable: free abelian groups, free groups [53], groups of the form $F \times \mathbb{Z}$, where F is free [61], and locally quasiconvex LERF hyperbolic groups [43] (eg, surface groups). Additionally, the class of product separable groups is closed under taking subgroups, finite-index supergroups and free products [14]. However, this class is not closed under direct products (eg, the direct product of two non-abelian free groups is not even LERF [4]). It also does not contain some polycyclic groups: in [33] Lennox and Wilson proved that the integral Heisenberg group $H_3(\mathbb{Z})$, which is polycyclic (in fact, finitely generated nilpotent of class 2), is not product separable as it does not have property RZ_3.

We use Theorem 1.8 to establish product separability for many more groups.

Theorem 2.2 *The following groups are product separable*:

(i) *limit groups*;

(ii) *finitely generated Kleinian groups*;

(iii) *fundamental groups of finite graphs of free groups with cyclic edge groups*, *as long as they are balanced.*

Recall that a group G is called a *limit group* if it is finitely generated and fully residually free (that is, for every finite subset $A \subset G$, there is a free group F and a homomorphism $\varphi\colon G \to F$ that is injective when restricted to A). Limit groups played an important role in the solutions of Tarski's problems about the first order theory of free groups by Sela [54] and Kharlampovich and Myasnikov [32].

Following Wise, we say that a group G is *balanced* if for every infinite order element $g \in G$ the conjugacy between g^m and g^n implies that $n = \pm m$. In [59], Wise proved that the fundamental group G of a finite graph of free groups with cyclic edge groups is LERF if and only if it is balanced if and only if G does not contain any *non-Euclidean* Baumslag–Solitar subgroups $BS(m,n) = \langle a, t \mid ta^m t^{-1} = a^n \rangle$, with $m, n \in \mathbb{Z} \setminus \{0\}$ and $n \neq \pm m$.

Part (iii) of Theorem 2.2 generalises a result of Coulbois [13, Theorem 5.18], who proved that the free amalgamated product of two free groups along a cyclic subgroup is product separable. Theorem 2.2(iii) confirms (in a strong way) a conjecture of Hsu and Wise [31, Conjecture 15.5], which states that a balanced group splitting as a finite graph of free groups with cyclic edge groups is double coset separable.

Corollary 2.3 *Suppose that G splits as a fundamental group of a finite graph of finitely generated free groups with cyclic edge groups. If G is balanced then it is virtually compact special*; *in other words*, G *has a finite-index subgroup which is isomorphic to the fundamental group of a compact non-positively curved special cube complex* (*in the sense of Haglund and Wise* [28]).

Proof Hsu and Wise [31, Theorem 10.4] proved that G admits a proper cocompact action on a CAT(0) cube complex $\mathscr{X}$. By Theorem 2.2, G is double coset separable; hence, by a result of Haglund and Wise [28, Theorem 9.19], G has a finite-index subgroup K such that $K \backslash \mathscr{X}$ is a special cube complex. □

After the completion of this paper the authors learned of a recent result of Shepherd and Woodhouse [56, Theorem 1.2], which gives an alternative proof of Corollary 2.3, using different methods.

One of the original motivations for considering product separability of groups came from semigroups and automata theory. Pin and Reutenauer [49] used this property to characterise the profinitely closed rational subsets of free groups.

Recall that for a monoid M, the *rational subsets* $\mathrm{Rat}(M) \subseteq 2^M$ form the smallest collection of subsets of M satisfying the following conditions:

(1) $\varnothing \in \mathrm{Rat}(M)$ and, for each $m \in M$, $\{m\} \in \mathrm{Rat}(M)$;

(2) if $A, B \in \mathrm{Rat}(M)$, then $AB \in \mathrm{Rat}(M)$ and $A \cup B \in \mathrm{Rat}(M)$;

(3) if $A \in \mathrm{Rat}(M)$, then $A^* \in \mathrm{Rat}(M)$, where A^* is the submonoid of M generated by A.

We refer the reader to [49] for an account of the basic theory of rational subsets.

In a group G it makes sense to consider the subgroup closure instead of the $*$-closure. Thus we define the set $\mathrm{Rat}^0(G) \subseteq 2^G$ as the smallest collection of subsets of G containing all finite subsets, closed under finite unions, products and subgroup closure. It is easy to see that $\mathrm{Rat}^0(G)$ consists of all subsets of the form $gF_1 \cdots F_s$, where $s \in \mathbb{N}_0$, $g \in G$ and $F_1, \dots, F_s$ are finitely generated subgroups of G [49, Proposition 2.2]. Evidently $\mathrm{Rat}^0(G) \subseteq \mathrm{Rat}(G)$; moreover, it is not difficult to show that $\mathrm{Rat}^0(G) = \mathrm{Rat}(G)$ if and only if G is torsion.

The following theorem was proved by Pin and Reutenauer [49, Corollary 2.5] in the case of free groups (see also [52, Section 12.3] for a slightly different argument); however the proof is readily seen to remain valid in all product separable groups.

Theorem 2.4 (Pin and Reutenauer) *If G is a product separable group then $\mathrm{Rat}^0(G)$ is precisely the class of all separable rational subsets of G.*

Corollary 2.5 *If G is a group from one of the classes* (i)–(iii), *described in Theorem 2.2, then the set of separable rational subsets of G coincides with $\mathrm{Rat}^0(G)$.*

3 Plan of the paper

3.1 The metric quasiconvexity theorem

Let G be a relatively hyperbolic group generated by a finite set X, and let Q and R be relatively quasiconvex subgroups of G. The technical heart of this paper is Theorem 3.5 below, which, given some relatively quasiconvex subgroups $Q' \leqslant Q$ and $R' \leqslant R$, provides sufficient metric conditions for the relative quasiconvexity of the join $\langle Q', R' \rangle$.

Definition 3.1 ($\min_X$) Let G be a group with finite generating set X, and let $Y \subseteq G$. Then we denote the number $\min\{|g|_X \mid g \in Y\}$ by $\min_X(Y)$, with the usual convention that minimum over the empty set is $+\infty$.

Let $S = Q \cap R$ and $A \geq 0$ be some constant. We will be interested in finding subgroups $Q' \leqslant Q$ and $R' \leqslant R$ satisfying the following properties:

(P1) if Q' and R' are relatively quasiconvex in G then so is the subgroup $\langle Q', R' \rangle$;

(P2) $\min_X(\langle Q', R' \rangle \setminus S) \geq A$;

(P3) $\min_X(Q\langle Q', R' \rangle R \setminus QR) \geq A$.

Remark 3.2

- Quasiconvexity of Q' and R' is only required in property (P1).
- Property (P2) says that all "short" elements of $\langle Q', R' \rangle$ belong to S.
- Property (P3) is the key ingredient for proving that the double coset QR is separable in G in Corollary 1.4.

Let us now describe the metric conditions used to establish the above properties. Given a finite collection $\mathcal{P}$ of maximal parabolic subgroups of G, constants $B, C \geq 0$ and subgroups $Q' \leqslant Q$ and $R' \leqslant R$, we will consider the following conditions:

(C1) $Q' \cap R' = S$;

(C2) $\min_X(Q\langle Q', R' \rangle Q \setminus Q) \geq B$ and $\min_X(R\langle Q', R' \rangle R \setminus R) \geq B$;

(C3) $\min_X((PQ' \cup PR') \setminus PS) \geq C$, for each $P \in \mathcal{P}$.

Moreover, if not all of the subgroups in $\mathcal{P}$ are abelian then we will need two more conditions (here for subgroups $H, P \leqslant G$, we use H_P to denote the intersection $H \cap P \leqslant P$):

(C4) $Q_P \cap \langle Q'_P, R'_P \rangle = Q'_P$ and $R_P \cap \langle Q'_P, R'_P \rangle = R'_P$, for every $P \in \mathcal{P}$;

(C5) $\min_X(q\langle Q'_P, R'_P \rangle R_P \setminus qQ'_P R_P) \geq C$, for each $P \in \mathcal{P}$ and all $q \in Q_P$.

Remark 3.3 If the peripheral subgroups of G are abelian then condition (C4) follows from (C1) and condition (C5) is trivially true.

Indeed, if P is abelian, then, in the notation of (C4), $\langle Q'_P, R'_P \rangle = Q'_P R'_P$; hence

$$Q'_P \subseteq Q_P \cap \langle Q'_P, R'_P \rangle = Q_P \cap Q'_P R'_P = Q'_P(Q_P \cap R'_P) \subseteq Q'_P S_P = Q'_P,$$

where the last equality used that $S_P = S \cap P \subseteq Q'_P$ by (C1). The second equality of (C4) can be proved in the same fashion.

Similarly, if $q \in Q_P$ then $q\langle Q'_P, R'_P \rangle R_P = qQ'_P R'_P R_P = qQ'_P R_P$, so

$$\min_X(q\langle Q'_P, R'_P \rangle R_P \setminus qQ'_P R_P) = \min_X(\varnothing) = +\infty;$$

thus (C5) holds.

Remark 3.4 In this paper we will be primarily interested in the existence of finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ satisfying the above conditions. This may be easier to interpret through the lens of the profinite topology on G (see Section 11):

- Conditions (C1) and (C4) can be ensured by choosing any finite-index subgroup $M \leqslant_f G$ with $S \subseteq M$, and setting $Q' = Q \cap M$, $R' = R \cap M$.
- The existence of finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ satisfying condition (C2) can be deduced from separability of Q and R in G.
- The existence of finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ satisfying condition (C3) can be deduced from separability of the double coset PS in G.

- If $Q'_P \leqslant_f Q_P$ is already chosen then $R'_P \leqslant_f R_P$, satisfying (C5), can be constructed with the help of separability of the double coset $Q'_P R_P$ in P. Indeed, if $Q_P = \bigcup_{j=1}^n a_j Q'_P$, then the inequality in (C5) can be rewritten as $\min_X(a_j \langle Q'_P, R'_P \rangle Q'_p R_P \setminus a_j Q'_P R_P) \geq C$, for every $j = 1, \dots, n$. Thus our approach to establishing (C5) will be to choose $R' \leqslant_f R$ after $Q' \leqslant_f Q$ has already been constructed (in other words, R' will depend on Q').

Theorem 3.5 (metric quasiconvexity theorem) *Let G be relatively hyperbolic group generated by a finite set X. Suppose that $Q, R \leqslant G$ are relatively quasiconvex subgroups and denote $S = Q \cap R$. There exists a finite collection $\mathcal{P}$ of maximal parabolic subgroups of G such that for any $A \geq 0$ there are constants $B, C \geq 0$ satisfying the following.*

Suppose that $Q' \leqslant Q$ and $R' \leqslant R$ are subgroups of G satisfying conditions (C1)–(C5). *Then these subgroups enjoy properties* (P1)–(P3) *above.*

Rough sketches of the proofs of Theorems 3.5 and 1.2 are given in the beginning of Part II of the paper.

3.2 The separability assumptions

As the reader may notice, our main results in the introduction assume that the underlying relatively hyperbolic group G is QCERF and the peripheral subgroups of G are double coset separable. Indeed, the essence of our method is in finding (sufficiently many) finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ satisfying conditions (C1)–(C5) by using properties of the profinite topology. However, a careful analysis of the arguments reveals that instead of the full QCERF assumption it is possible to require the separability only of certain finitely generated relatively quasiconvex subgroups related to Q and R. For example, the proof of Theorem 1.2 relies on the separability conditions (S1)–(S3) from Theorem 11.3, which are established in Section 13 using the separability of relatively quasiconvex subgroups Q, R, K, $\langle K, T \rangle$ and $\langle K, V \rangle$, where $K \leqslant_f P \in \mathcal{P}$, $T \leqslant_f Q$, $V \leqslant_f R$ and $\mathcal{P} = \mathcal{P}_1$ is a finite collection of maximal parabolic subgroups of G that depends on Q and R (see Notation 10.2). The exact requirements for double coset separability of the peripheral subgroups are easier to trace: it suffices to look at condition (S4) of Theorem 11.3.

3.3 Section outline

This paper is structured as follows. There are three parts: Part I contains background material and useful preliminary results (Sections 4–5), Part II is dedicated to the proof of the metric quasiconvexity theorem and the double coset separability results that follow from them (Sections 6–15), and Part III is essentially dedicated to the proof and applications of Theorem 1.8 (Sections 16–21).

Section 4 covers generalities and Section 5 covers definitions and results specific to relatively hyperbolic groups. In Section 6 we introduce the terminology of *path representatives*, their associated *types*, and make some observations about path representatives that have minimal type. Sections 7 and 8 are devoted to controlling certain instances of backtracking in minimal type path representatives. In Section 9 we describe the "shortcutting" of a broken line, and establish its quasigeodesicity under some technical assumptions. Section 10 contains the proof of Theorem 3.5. In Sections 11 and 13 we show how finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ satisfying conditions (C1)–(C5) can be obtained using separability, with the help of a new criterion for separability of double cosets in amalgamated products from Section 12. Section 14 contains proofs of Theorems 1.2 and 1.3, while Section 15 contains the proof of Corollary 1.6.

In Section 16 we generalise the content of Section 6 to the setting of products of subgroups, as well as introducing new metric conditions (C2-m) and (C5-m). Sections 17 and 18 are product analogues to Section 8; similarly, Section 19 generalises Section 11. Finally, Section 20 contains the proof of Theorem 1.8, and Section 21 establishes new examples of product separable groups, proving Theorem 2.2.

Part I Background

In this part we will present the definitions and basic results that will be necessary for the rest of the paper.

4 Preliminaries

4.1 Notation

We write $\mathbb{N}$ for the set of natural numbers $\{1, 2, 3, \dots\}$, and $\mathbb{N}_0$ for $\mathbb{N} \cup \{0\}$.

Let G be a group. If H is a finite-index (respectively, finite-index normal) subgroup of G, then we write $H \leqslant_f G$ (respectively, $H \lhd_f G$). For a subgroup $T \leqslant G$ and elements $a, b \in G$ we will write $T^a = aTa^{-1} \leqslant G$ and $b^a = aba^{-1} \in G$.

By a generating set $\mathcal{A}$ of G we will mean a set $\mathcal{A}$ together with a map $\mathcal{A} \to G$ such that the image of $\mathcal{A}$ under this map generates G.

If $\mathcal{A}$ is a generating set for G, then we denote by $\Gamma(G, \mathcal{A})$ the (left) Cayley graph of G with respect to $\mathcal{A}$. The standard edge path length metric on $\Gamma(G, \mathcal{A})$ will be denoted by $d_{\mathcal{A}}(\cdot, \cdot)$. After identifying G with the vertex set of $\Gamma(G, \mathcal{A})$, this metric induces the *word metric* associated to $\mathcal{A}$: $d_{\mathcal{A}}(g, h) = |g^{-1}h|_{\mathcal{A}}$ for all $g, h \in G$, where $|g|_{\mathcal{A}}$ denotes the length of the shortest word in $\mathcal{A}^{\pm 1}$ representing g in G.

Abusing the notation, we will identify the combinatorial Cayley graph $\Gamma(G, \mathcal{A})$ with its geometric realisation. The latter is a geodesic metric space and, given two points x and y in this space, we will use $[x, y]$ to denote a geodesic path from x to y in $\Gamma(G, \mathcal{A})$. In general $\Gamma(G, \mathcal{A})$ need not be uniquely geodesic, so there will usually be a choice for $[x, y]$, which will either be specified or will be clear from the context (eg, if x and y already belong to some geodesic path under discussion, then $[x, y]$ will be chosen as the subpath of that path).

If $Y \subseteq G$ is a subset of G and $K \geq 0$, we denote by

$$N_{\mathcal{A}}(Y, K) = \{g \in G \mid d_{\mathcal{A}}(g, Y) \leq K\}$$

the K-neighbourhood of Y with respect to $d_{\mathcal{A}}$. Note that when $\mathcal{A}$ is a finite generating set, the metric $d_{\mathcal{A}}$ is proper. However, in this paper we will also be working with infinite generating sets; see Section 5 below, where generating sets of the form $\mathcal{A} = X \cup \mathcal{H}$ are considered.

The following general fact will be used quite often.

Lemma 4.1 *Let G be a group generated by a finite set $\mathcal{A}$. If $A, B \leqslant G$ are subgroups of G then for every $K \geq 0$ there is a constant $K' = K'(A, B, K) \geq 0$ such that for any $x \in G$ we have*

$$N_{\mathcal{A}}(xA, K) \cap N_{\mathcal{A}}(xB, K) \subseteq N_{\mathcal{A}}(x(A \cap B), K').$$

Proof After applying the left translation by x^{-1}, which preserves the metric $d_{\mathcal{A}}$, we can assume that $x = 1$. Now the statement follows, for example, from [30, Proposition 9.4]. □

Suppose that γ is a combinatorial path (edge path) in $\Gamma(G, \mathcal{A})$. We will denote the initial and terminal endpoints of γ by γ_- and γ_+ respectively. We will write $\ell(\gamma)$ for the length (that is, the number of edges) of γ. We will also use γ^{-1} to denote the inverse of γ, which is the path starting at γ_+, ending at γ_- and traversing γ in the reverse direction. If $\gamma_1, \dots, \gamma_n$ are combinatorial paths with $(\gamma_i)_+ = (\gamma_{i+1})_-$, for each $i \in \{1, \dots, n-1\}$, we will denote their concatenation by $\gamma_1 \cdots \gamma_n$.

Since $\Gamma(G, \mathcal{A})$ is a labelled graph, every combinatorial path γ comes with a label $\mathrm{Lab}(\gamma)$, which is a word over the alphabet $\mathcal{A}^{\pm 1}$. We denote by $\tilde{\gamma} \in G$ the element represented by $\mathrm{Lab}(\gamma)$ in G. Finally, we write $|\gamma|_{\mathcal{A}} = |\tilde{\gamma}|_{\mathcal{A}} = d_{\mathcal{A}}(\gamma_-, \gamma_+)$. Note that $\mathrm{Lab}(\gamma^{-1})$ is the formal inverse of $\mathrm{Lab}(\gamma)$, so that and $|\gamma^{-1}|_{\mathcal{A}} = |\gamma|_{\mathcal{A}}$ and $\widetilde{\gamma^{-1}} = \tilde{\gamma}^{-1}$.

4.2 Quasigeodesic paths

In this section we assume that Γ is a graph equipped with the standard path length metric $d(\cdot,\cdot)$.

Definition 4.2 (quasigeodesic) Let $\lambda \geq 1$ and $c \geq 0$ be some numbers and let p be an edge path in Γ. Recall that p is said to be (λ, c)-*quasigeodesic* if for every combinatorial subpath q of p we have

$$\ell(q) \leq \lambda d(q_-, q_+) + c.$$

Lemma 4.3 *Suppose that $s = rpt$ is a concatenation of three combinatorial paths r, p and t in Γ such that $\ell(r) \leq D$ and $\ell(t) \leq D$, for some $D \geq 0$, and p is (λ, c)-quasigeodesic, for some $\lambda \geq 1$ and $c \geq 0$. Then the path s is (λ, c')-quasigeodesic, where $c' = c + 2(\lambda + 1)D$.*

Proof Consider an arbitrary combinatorial subpath q of s. We need to show that

$$\ell(q) \leq \lambda d(q_-, q_+) + c + 2(\lambda + 1)D. \tag{4-1}$$

If q is contained in r or in t then the desired inequality follows from the assumptions that $\ell(r) \leq D$ and $\ell(t) \leq D$. Therefore we can further suppose that q_- is a vertex of rp and q_+ is a vertex of pt. The bounds on the lengths of r and t imply that there is a combinatorial subpath a of p such that there are at most D edges of s between q_- and a_- and between a_+ and q_+. Thus $d(q_-,a_-) \le D$, $d(q_+,a_+) \le D$ and $\ell(q) \le \ell(a) + 2D$

The assumption that p is (λ, c)-quasigeodesic implies that

$$\ell(q) \le \ell(a) + 2D \le \lambda d(a_-,a_+) + c + 2D. \tag{4-2}$$

The triangle inequality gives $d(a_-,a_+) \le d(q_-,q_+) + 2D$, which, combined with (4-2), shows that (4-1) holds, as required. □

Lemma 4.4 *Let $\lambda \ge 1$, $c \ge 0$ and $K \in \mathbb{N}$. Suppose that p is a combinatorial path in Γ and let p' be a path obtained by replacing some edges of p with combinatorial paths of length at most K. If p is (λ, c)-quasigeodesic then p' is $(K\lambda, 2K^2\lambda + Kc + 2K)$-quasigeodesic.*

Proof Let q be any combinatorial subpath of p' and write $q_- = x$ and $q_+ = y$. We need to show that

$$\ell(q) \le K\lambda d(x,y) + 2K^2\lambda + Kc + 2K. \tag{4-3}$$

If q does not contain any vertices of p then $\ell(q) \le K$ and (4-3) holds. Otherwise, let z and w be the first and the last vertices of q that lie on p respectively, and let r be the subpath of p starting at z and ending at w. The assumptions imply that $d(x,z) \le K$, $d(y,w) \le K$ and

$$\ell(q) \le K\ell(r) + 2K. \tag{4-4}$$

Using the quasigeodesicity of p and the triangle inequality, we obtain

$$\ell(r) \le \lambda d(z,w) + c \le \lambda d(x,y) + 2K\lambda + c,$$

which, combined with (4-4), gives (4-3). □

4.3 Hyperbolic metric spaces

In this subsection let (Γ, d) be a geodesic metric space.

Definition 4.5 (Gromov product) Let $x, y, z \in \Gamma$ be points. The *Gromov product* of x and y with respect to z is

$$\langle x, y\rangle_z = \tfrac{1}{2}(d(x,z) + d(y,z) - d(x,y)).$$

It is easy to see that the Gromov products satisfy

$$d(x,y) = \langle y,z\rangle_x + \langle x,z\rangle_y, \quad d(y,z) = \langle x,z\rangle_y + \langle x,y\rangle_z, \quad d(z,x) = \langle x,y\rangle_z + \langle y,z\rangle_x.$$

The following elementary property of Gromov products is an immediate consequence of the triangle inequality.

Remark 4.6 Suppose that x, y and z are points in Γ, u is a point on any geodesic segment $[x,z]$, from x to z, and v is a point on any geodesic segment $[z,y]$, from z to y. Then

$$\langle u, v\rangle_z \le \langle x, y\rangle_z.$$

Definition 4.7 (δ-thin triangle) Let Δ be a geodesic triangle in Γ with vertices x, y and z, and let $\delta \geq 0$. Denote by T_Δ the (possibly degenerate) tripod with edges of length $\langle x, y\rangle_z$, $\langle y, z\rangle_x$ and $\langle z, x\rangle_y$ respectively. There is an map from $\{x, y, z\}$ to the extremal vertices of T_Δ, which extends uniquely to a map $\phi\colon \Delta \to T_\Delta$, whose restriction to each side of Δ is an isometry. If the diameter in Γ of $\phi^{-1}(\{t\})$ is at most δ, for all $t \in T_\Delta$, then Δ is said to be *δ-thin*.

Definition 4.8 (hyperbolic space) The space Γ is said to be a *hyperbolic metric space* if there is a constant $\delta \geq 0$ such that every geodesic triangle in Γ is δ-thin.

The above definition of δ-hyperbolicity is not the most commonly used in the literature, though it is well-known to be equivalent to other definitions after possibly increasing δ; see, for example, [9, III.H.1.17]. For technical reasons we will always assume that δ is chosen to be sufficiently large so that all the definitions in this reference are satisfied.

In the remainder of this subsection we assume that Γ is a δ-hyperbolic graph, for some $\delta \geq 0$, and $d(\cdot,\cdot)$ is the standard path length metric on Γ.

Definition 4.9 (broken line) A *broken line* in Γ is a path p which comes with a fixed decomposition as a concatenation of combinatorial geodesic paths $p_1, \dots, p_n$ in Γ, so $p = p_1 p_2 \cdots p_n$. The paths $p_1, \dots, p_n$ will be called the *segments* of the broken line p, and the vertices

$$p_- = (p_1)_-, (p_1)_+ = (p_2)_-, \dots, (p_{n-1})_+ = (p_n)_-, (p_{n+1})_+ = p_+$$

will be called the *nodes* of p.

The following statement is a special case of [41, Lemma 4.2], applied to the situation when each p_i is geodesic (so, in the notation of that lemma, we can take $\bar\lambda = 1$, $\bar c = 0$ and $\nu = \delta$). Note that due to a slightly different definition of quasigeodesicity used in [41], a (λ, c)-quasigeodesic in the sense of [41] is $(1/\lambda, c/\lambda)$-quasigeodesic in the sense of Definition 4.2 above, and vice versa.

Lemma 4.10 *Let c_0, c_1 and c_2 be constants such that $c_0 \geq 14\delta$, $c_1 = 12(c_0+\delta)+1$ and $c_2 = 10(\delta+c_1)$. Suppose that $p = p_1 \cdots p_n$ is a broken line in Γ, where p_i is a geodesic with $(p_i)_- = x_{i-1}$ and $(p_i)_+ = x_i$ for $i = 1, \dots, n$. If $d(x_{i-1}, x_i) \geq c_1$ for $i = 1, \dots, n$, and $\langle x_{i-1}, x_{i+1}\rangle_{x_i} \leq c_0$ for each $i = 1, \dots, n-1$, then the path p is $(4, c_2)$-quasigeodesic.*

We will need an extension of the above lemma which allows the first and the last geodesic segments p_1 and p_n to be short.

Lemma 4.11 *For any constant c_0 satisfying $c_0 \geq 14\delta$, let*

$$c_1 = c_1(c_0) = 12(c_0+\delta)+1 \quad \textit{and} \quad c_3 = c_3(c_0) = 10(\delta + 2c_1).$$

Suppose that $p = p_1 \cdots p_n$ is a broken line in Γ, where p_i is a geodesic with $(p_i)_- = x_{i-1}$ and $(p_i)_+ = x_i$ for $i = 1, \dots, n$. If $d(x_{i-1}, x_i) \geq c_1$ for $i = 2, \dots, n-1$, and $\langle x_{i-1}, x_{i+1}\rangle_{x_i} \leq c_0$ for each $i = 1, \dots, n-1$, then the path p is $(4, c_3)$-quasigeodesic.

Proof This follows easily by combining Lemma 4.10 with Lemma 4.3. Indeed, there are four possibilities depending on whether or not $d(x_0, x_1) \geq c_1$ and $d(x_{n-1}, x_n) \geq c_1$. Since all of these cases are similar, let us concentrate on the situation when $d(x_0, x_1) < c_1$ and $d(x_{n-1}, x_n) \geq c_1$. Then the path $q = p_2 p_3 \cdots p_n$ is $(4, c_2)$-quasigeodesic by Lemma 4.10, where $c_2 = 10(\delta + c_1)$. Since $\ell(p_1) = d(x_0, x_1) < c_1$, we can apply Lemma 4.3 to deduce that the path $p = p_1 \cdots p_n = p_1 q$ is $(4, c_3)$-quasigeodesic, where $c_3 = c_2 + 10c_1 = 10(\delta + 2c_1)$, as required. □

4.4 Profinite topology and separable subsets

Let G be a group. The *profinite topology* on G is the topology $\mathcal{PT}(G)$ whose basis consists of left cosets to finite-index subgroups of G.

A subset $Z \subseteq G$ is called *separable* (in G) if it is closed in $\mathcal{PT}(G)$. Evidently finite unions and arbitrary intersections of separable subsets are separable. It is easy to see that a subset $Z \subseteq G$ is separable if and only if for every $g \in G \setminus Z$, there is a finite group Q and a homomorphism $\varphi\colon G \to Q$ such that $\varphi(g) \notin \varphi(Z)$ in Q. A subgroup $H \leq G$ is separable if and only if it is the intersection of the finite-index subgroups of G containing it.

The following observation stems from the fact that the group operations of taking an inverse and multiplying by a fixed element are homeomorphisms with respect to the profinite topology.

Remark 4.12 Let Z be a separable subset of a group G. Then for every $g \in G$ the subsets Z^{-1}, gZ and Zg are also separable.

Lemma 4.13 *Suppose that A is a subgroup of a group G.*

(a) *Every subset of A which is closed in $\mathcal{PT}(G)$ is also closed in $\mathcal{PT}(A)$.*

(b) *If every finite-index subgroup of A is separable in G then every closed subset of $\mathcal{PT}(A)$ is closed in $\mathcal{PT}(G)$.*

Proof Claim (a) immediately follows from the observation that the intersection of A with any basic closed subset from $\mathcal{PT}(G)$ is either empty or is a basic closed subset of $\mathcal{PT}(A)$.

If each finite-index subgroup of A is separable in G then, in view of Remark 4.12, every basic closed set in $\mathcal{PT}(A)$ is closed in the profinite topology of G. Claim (b) of the lemma now follows from the fact that any closed subset of A is the intersection of basic closed sets. □

Lemma 4.14 *Let G be a group with subgroups A and B. Suppose that $A' \leqslant_f A$, $B' \leqslant_f B$ and $A'B'$ is separable in G. Then AB is separable in G.*

Proof Let $A = \bigsqcup_{i=1}^m a_i A'$ and $B = \bigsqcup_{j=1}^n B' b_j$. Then

$$AB = \bigcup_{i=1}^m \bigcup_{j=1}^n a_i A' B' b_j,$$

which is separable in G by Remark 4.12. □

The next two lemmas use the notation introduced in Sections 1.2 and 3.1.

Lemma 4.15 *Let A and B be subgroups of a group G such that $A \preccurlyeq B$. If B is separable in G then so are the double cosets AB and BA.*

Proof By [42, Lemma 2.1] $A \cap B$ has finite index in A, so $A = \bigsqcup_{i=1}^{m} a_i(A \cap B)$, for some $a_1, \dots, a_m \in A$. It follows that $AB = \bigcup_{i=1}^{m} a_i B$, so it is separable by Remark 4.12. The same remark also implies that $BA = (AB)^{-1}$ is separable in G. □

The main use of the profinite topology in this paper stems from the following elementary facts.

Lemma 4.16 *Let G be a group generated by a finite set X, and let $P \leqslant G$ be a subgroup. Suppose that Z is a separable subset of P.*

(a) *If a finite subset $U \subseteq P$ is disjoint from Z then there is a normal finite-index subgroup $N \lhd_f P$ such that $U \cap ZN = \varnothing$. Thus the image of U in the quotient P/N will be disjoint from the image of Z.*

(b) *For every constant $C \geq 0$ there is a finite-index normal subgroup $N \lhd_f P$ such that*
$$\min_X(ZN \setminus Z) \geq C.$$

(c) *For any finite subset $A \subseteq P$ and any $C \geq 0$ there exists $N \lhd_f P$ such that*
$$\min_X(aZN \setminus aZ) \geq C \quad \text{for all } a \in A.$$

Proof For (a), let $U = \{u_1, \dots, u_m\} \subseteq P$. Since $u_i \notin Z$ and Z is separable in P, there exists $N_i \lhd_f P$ such that $u_i N_i \cap Z = \varnothing$, for each $i = 1, \dots, m$. We set $N = \bigcap_{i=1}^{m} N_i \lhd_f P$, so that $u_i N \cap Z = \varnothing$. That is, $u_i \notin ZN$ for all $i = 1, \dots, m$. Therefore $U \cap ZN = \varnothing$ and (a) has been proved.

Claim (b) follows by applying claim (a) to the finite subset $U = \{g \in P \setminus Z \mid |g|_X < C\}$ of P.

To prove (c), suppose that $A = \{a_1, \dots, a_k\} \subseteq P$. By Remark 4.12, $a_j Z$ is separable in P, for every $j = 1, \dots, k$, so, according to part (b), there exists $N_j \lhd_f P$ such that
$$\min_X(a_j Z N_j \setminus a_j Z) \geq C \quad \text{for each } j = 1, \dots, k.$$
It is easy to see that the normal subgroup $N = \bigcap_{j=1}^{k} N_j \lhd_f P$ enjoys the required property. □

The following statement is well known; we include a proof for completeness.

Lemma 4.17 *Let G be a group with subgroups $K \leqslant_f H \leqslant G$. If K is separable in G, then there is $L \leqslant_f G$ such that $L \cap H = K$*

Proof Since K is of finite index in H, we can write
$$H = K \cup Kh_1 \cup \dots \cup Kh_m$$

for some $h_1, \dots, h_m \in H \setminus K$. The subgroup K is separable in G, meaning that it is closed in $\mathcal{PT}(G)$. Following Remark 4.12, the union $Kh_1 \cup \cdots \cup Kh_m$ is also closed in $\mathcal{PT}(G)$. Thus the subset

$$(G \setminus H) \cup K = G \setminus (Kh_1 \cup \cdots \cup Kh_m)$$

is open in $\mathcal{PT}(G)$ and contains the identity. It follows from the definition of the profinite topology that there is a finite-index normal subgroup $N \lhd_f G$ with $N \subseteq (G \setminus H) \cup K$. Observe that $Kh_i \cap N = \varnothing$, for every $i = 1, \dots, m$, so $N \cap H \leqslant K$. Now set $L = KN \leqslant_f G$. Then $L \cap H = KN \cap H = K(N \cap H) = K$, as required. □

5 Relatively hyperbolic groups

In this section we define relatively hyperbolic groups and collect various properties that will be used throughout the paper.

5.1 Definition

We will define relatively hyperbolic groups following the approach of Osin (for full details, see [46]).

Definition 5.1 (relative generating set, relative presentation) Let G be a group, $X \subseteq G$ a subset and $\{H_\nu \mid \nu \in \mathcal{N}\}$ a collection of subgroups of G. The group G is said to be *generated by X relative to* $\{H_\nu \mid \nu \in \mathcal{N}\}$ if it is generated by $X \sqcup \mathcal{H}$, where $\mathcal{H} = \bigsqcup_{\nu \in \mathcal{N}} (H_\nu \setminus \{1\})$ (with the obvious map $X \sqcup \mathcal{H} \to G$). If this is the case, then there is a surjection

$$F = F(X) * (*_{\nu \in \mathcal{N}} H_\nu) \to G,$$

where $F(X)$ denotes the free group on X. Suppose that the kernel of this map is the normal closure of a subset $\mathcal{R} \subseteq F$. Then G can equipped with the *relative presentation*

$$\langle X, H_\nu, \nu \in \mathcal{N} \mid \mathcal{R} \rangle. \tag{5-1}$$

If X is a finite set, then G is said to be *finitely generated relative to* $\{H_\nu \mid \nu \in \mathcal{N}\}$. If $\mathcal{R}$ is also finite, G is said to be *finitely presented relative to* $\{H_\nu \mid \nu \in \mathcal{N}\}$ and the presentation above is a *finite relative presentation*.

With the above notation, we call the Cayley graph $\Gamma(G, X \cup \mathcal{H})$ the *relative Cayley graph* of G with respect to X and $\{H_\nu \mid \nu \in \mathcal{N}\}$. Note that when X is itself a generating set of G, $d_{X \cup \mathcal{H}}(g, h) \le d_X(g, h)$, for all $g, h \in G$.

Definition 5.2 (relative Dehn function) Suppose that G has a finite relative presentation (5-1) with respect to a collection of subgroups $\{H_\nu \mid \nu \in \mathcal{N}\}$. If w is a word in the free group $F(X \sqcup \mathcal{H})$, representing the identity in G, then it is equal in F to a product of conjugates

$$w \overset{F}{=} \prod_{i=1}^{n} a_i r_i a_i^{-1},$$

where $a_i \in F$ and $r_i \in \mathcal{R}$, for each i. The *relative area* of the word w with respect to the relative presentation, $\mathrm{Area}^{\mathrm{rel}}(w)$, is the least number n among products of conjugates as above that are equal to w in F.

A *relative isoperimetric function* of the above presentation is a function $f : \mathbb{N} \to \mathbb{N}$ such that $\mathrm{Area}^{\mathrm{rel}}(w)$ is at most $f(|w|)$, for every freely reduced word w in $F(X \sqcup \mathcal{H})$ representing the identity in G. If an isoperimetric function exists for the presentation, the smallest such function is called the *relative Dehn function* of the presentation.

Definition 5.3 (relatively hyperbolic group) Let G be a group and let $\{H_\nu \mid \nu \in \mathcal{N}\}$ be a collection of subgroups of G. If G admits a finite relative presentation with respect to this collection of subgroups which has a well-defined linear relative Dehn function, it is called *hyperbolic relative to* $\{H_\nu \mid \nu \in \mathcal{N}\}$. When it is clear what the relevant collection of subgroups is, we refer to G simply as a *relatively hyperbolic group*. The groups $\{H_\nu \mid \nu \in \mathcal{N}\}$ are called the *peripheral subgroups* of the relatively hyperbolic group G, and their conjugates in G are called *maximal parabolic subgroups*. Any subgroup of a maximal parabolic subgroup is said to be *parabolic*.

Lemma 5.4 [46, Corollary 2.54] *Suppose that G is a group generated by a finite set X and hyperbolic relative to a collection of subgroups $\{H_\nu \mid \nu \in \mathcal{N}\}$, and let $\mathcal{H} = \bigsqcup_{\nu \in \mathcal{N}} (H_\nu \setminus \{1\})$. Then the Cayley graph $\Gamma(G, X \cup \mathcal{H})$ is δ-hyperbolic, for some $\delta \geq 0$.*

In the remainder of this section (namely, in Sections 5.2–5.4, we will assume that G is a group generated by a finite subset X and hyperbolic relative to a finite collection of subgroups $\{H_\nu \mid \nu \in \mathcal{N}\}$. As usual, we will let $\mathcal{H} = \bigsqcup_{\nu \in \mathcal{N}} (H_\nu \setminus \{1\})$.

5.2 Geodesics and quasigeodesics in relatively hyperbolic groups

Definition 5.5 (path components) Let p be a combinatorial path in $\Gamma(G, X \cup \mathcal{H})$. A non-trivial combinatorial subpath of p whose label consists entirely of elements of $H_\nu \setminus \{1\}$, for some $\nu \in \mathcal{N}$, is called an *H_ν-subpath* of p.

An H_ν-subpath is called an *H_ν-component* if it is not contained in any strictly longer H_ν-subpath. We will call a subpath of p an $\mathcal{H}$-subpath (respectively, an *$\mathcal{H}$-component*) if it is an H_ν-subpath (respectively, an H_ν-component), for some $\nu \in \mathcal{N}$.

Definition 5.6 (connected and isolated components) Let p and q be edge paths in $\Gamma(G, X \cup \mathcal{H})$ and suppose that s and t are H_ν-subpaths of p and q respectively, for some $\nu \in \mathcal{N}$. We say that s and t are *connected* if s_- and t_- belong to the same left coset of H_ν in G. The latter means that for all vertices u of s and v of t either $u = v$ or there is an edge e in $\Gamma(G, X \cup \mathcal{H})$ with $\mathrm{Lab}(e) \in H_\nu \setminus \{1\}$ and $e_- = u, e_+ = v$.

If s is an H_ν-component of a path p and s is not connected to any other H_ν-component of p then we say that s is *isolated* in p.

Definition 5.7 (phase vertex) A vertex v of a combinatorial path p in $\Gamma(G, X \cup \mathcal{H})$ is called *non-phase* if it is an interior vertex of an $\mathcal{H}$-component of p (that is, if it lies in an $\mathcal{H}$-component for which it is not an endpoint). Otherwise v is called *phase*.

Definition 5.8 (backtracking) If all $\mathcal{H}$-components of a combinatorial path p are isolated, then p is said to be *without backtracking*. Otherwise we say that p *has backtracking*.

Remark 5.9 If p is a geodesic edge path in $\Gamma(G, X \cup \mathcal{H})$ then every $\mathcal{H}$-component of p will consist of a single edge, labelled by an element from $\mathcal{H}$. Therefore every vertex of p will be phase. Moreover, it is easy to see that p will be without backtracking.

The following is a basic observation about the lengths of paths in the relative Cayley graph whose $\mathcal{H}$-components are uniformly short.

Lemma 5.10 *Let p be a path in $\Gamma(G, X \cup \mathcal{H})$ and suppose there is a constant $\Theta \geq 1$ such that for any $\mathcal{H}$-component h of p, we have $|h|_X \leq \Theta$. Then $|p|_X \leq \Theta\ell(p)$.*

Proof We can write p as a concatenation $p = a_0 h_1 a_1 \cdots a_{n-1} h_n a_n$, where $h_1, \dots, h_n$ are the $\mathcal{H}$-components of p and $a_0, \dots, a_n$ are subpaths of p all whose edges are labelled by elements of $X^{\pm 1}$.

It follows from the triangle inequality that

$$|p|_X = d_X(p_-, p_+) \leq \sum_{i=0}^{n} d_X((a_i)_-, (a_i)_+) + \sum_{i=1}^{n} d_X((h_i)_-, (h_i)_+).$$

Since each edge of a_i is labelled by an element of $X^{\pm 1}$, we have that $d_X((a_i)_-, (a_i)_+) \leq \ell(a_i)$, for all $i = 0, \dots, n$. Moreover, $d_X((h_i)_-, (h_i)_+) = |h_i|_X \leq \Theta\ell(h_i)$, for each $i = 1, \dots, n$, by the hypothesis of the lemma, as $\ell(h_i) \geq 1$.

Combining the above three inequalities with the fact that $\Theta \geq 1$, we obtain

$$|p|_X \leq \sum_{i=0}^{n} \ell(a_i) + \sum_{i=1}^{n} \Theta\ell(h_i) \leq \Theta\left(\sum_{i=0}^{n} \ell(a_i) + \sum_{i=1}^{n} \ell(h_i)\right) = \Theta\ell(p). \qquad \square$$

Lemma 5.11 [46, Lemma 3.1] *There is a constant $M \geq 1$ such that if $h_1, \dots, h_n$ are isolated $\mathcal{H}$-components of a cycle q in $\Gamma(G, X \cup \mathcal{H})$, then*

$$\sum_{i=1}^{n} |h_i|_X \leq M\ell(q).$$

Lemma 5.12 *For any $\lambda \geq 1$, $c \geq 0$ and $A \geq 0$ there is a constant $\eta = \eta(\lambda, c, A) \geq 0$ such that the following is true.*

Suppose that p is a (λ, c)-quasigeodesic path in $\Gamma(G, X \cup \mathcal{H})$ possessing an isolated $\mathcal{H}$-component h such that $|h|_X \geq \eta$. Then $|p|_X \geq A$.

Proof Let $M \geq 1$ be the constant from Lemma 5.11, and set

$$\eta = M(1+\lambda)A + Mc. \tag{5-2}$$

Let q be a path in $\Gamma(G, X \cup \mathcal{H})$, labelled by a word over $X^{\pm 1}$, with endpoints $q_- = p_-$ and $q_+ = p_+$, such that $\ell(q) = |p|_X$.

Consider the cycle $r = pq^{-1}$ in $\Gamma(G, X \cup \mathcal{H})$, formed by concatenating p and the inverse of q. By the quasigeodesicity of p, $\ell(p) \leq \lambda |p|_{X \cup \mathcal{H}} + c \leq \lambda |p|_X + c$. Now $\ell(r) = \ell(p) + \ell(q)$; therefore

$$\ell(r) \leq (1+\lambda)|p|_X + c. \tag{5-3}$$

Since h is isolated in p it must also be an isolated $\mathcal{H}$-component of the cycle r (because all edges of q are labelled by letters from $X^{\pm 1}$). Hence $|h|_X \leq M\ell(r)$ by Lemma 5.11, so (5-3) implies that

$$|p|_X \geq \frac{1}{1+\lambda}(\ell(r) - c) \geq \frac{1}{M(1+\lambda)}(|h|_X - Mc). \tag{5-4}$$

Combining the above inequality with (5-2) and the assumption that $|h|_X \geq \eta$, we obtain the desired bound $|p|_X \geq A$. □

Proposition 5.13 [47, Proposition 3.2] *There is a constant $L \geq 0$ such that if Δ is a geodesic triangle in $\Gamma(G, X \cup \mathcal{H})$ and some side p is an isolated $\mathcal{H}$-component of Δ then $|p|_X \leq L$.*

Lemma 5.14 *There is a constant $L \geq 0$ such that if p_1 and p_2 are geodesic paths in $\Gamma(G, X \cup \mathcal{H})$ with $(p_1)_+ = (p_2)_-$, and s and t are connected H_ν-components of p_1 and p_2 respectively, for some $\nu \in \mathcal{N}$, then $d_X(s_+, t_-) \leq L$.*

Proof Let $L \geq 0$ be the constant provided by Proposition 5.13.

Since the component s of p_1 is connected to the component t of p_2, we know that $h = (s_+)^{-1} t_- \in H_\nu$. If $h = 1$ then $s_+ = t_-$ and there is nothing to prove, otherwise s_+ and t_- are endpoints of an edge e labelled by h in $\Gamma(G, X \cup \mathcal{H})$.

Consider the geodesic triangle Δ with vertices s_+, $(p_1)_+$ and t_-, where the sides $[s_+, (p_1)_+]$ and $[(p_1)_+, t_-]$ are chosen to be subpaths of p_1 and p_2 respectively, and the side $[s_+, t_-]$ is the edge e.

If $v \in [s_+, (p_1)_+]$ is a vertex belonging to the left coset $s_+ H_\nu$ then $d_{X \cup \mathcal{H}}(s_-, v) = 1$ and $s_+ \in [s_-, v]$ in p_1. Since $d_{X \cup \mathcal{H}}(s_-, s_+) = 1$ and p_1 is geodesic, we can conclude that $v = s_+$. Similarly, the only vertex of $[(p_1)_+, t_-]$ which belongs to the left coset $t_- H_\nu = s_+ H_\nu$ is t_-. It follows that the edge e is an isolated H_ν-component of Δ. Hence $d_X(s_+, t_-) \leq L$ by Proposition 5.13. □

Proposition 5.15 [46, Theorem 3.26] *Let Δ be a combinatorial geodesic triangle in $\Gamma(G, X \cup \mathcal{H})$ with sides p, q and r. There is a constant $\sigma = \sigma(G, \mathcal{H}, X) \in \mathbb{N}_0$ such that for any vertex $u \in p$, there is a vertex $v \in q \cup r$ with $d_X(u, v) \leq \sigma$.*

Definition 5.16 (k-similar paths) Let p and q be paths in $\Gamma(G, X \cup \mathcal{H})$, and let $k \geq 0$. The paths p and q are said to be *k-similar* if $d_X(p_-, q_-) \leq k$ and $d_X(p_+, q_+) \leq k$.

Proposition 5.17 [46, Proposition 3.15, Lemma 3.21 and Theorem 3.23] *For any $\lambda \geq 1$ and $c, k \geq 0$ there is a constant $\kappa = \kappa(\lambda, c, k) \geq 0$ such that if p and q are k-similar (λ, c)-quasigeodesics in $\Gamma(G, X \cup \mathcal{H})$ and p is without backtracking, then*

(1) *for every phase vertex u of p, there is a phase vertex v of q with $d_X(u, v) \leq \kappa$;*

(2) *every $\mathcal{H}$-component s of p, with $|s|_X \geq \kappa$, is connected to an $\mathcal{H}$-component of q.*

Moreover, if q is also without backtracking then

(3) *if s and t are connected $\mathcal{H}$-components of p and q respectively, then*

$$\max\{d_X(s_-, t_-), d_X(s_+, t_+)\} \leq \kappa.$$

5.3 Quasigeodesicity of paths with long components

One of the tools for proving Theorem 3.5 will be the next result of Martínez-Pedroza from [37].

Proposition 5.18 [37, Proposition 3.1] *There are constants $\zeta_0 \geq 0$ and $\lambda_0 \geq 1$ such that the following holds. If $q = r_0 s_1 \cdots r_n s_{n+1}$ is a concatenation of geodesic paths $r_0, s_1, \dots, r_n, s_{n+1}$ in $\Gamma(G, X \cup \mathcal{H})$ such that*

(1) *s_i is an $\mathcal{H}$-component of q, for each $i = 1, \dots, n+1$,*

(2) *$|s_i|_X \geq \zeta_0$, for every $i = 1, \dots, n+1$,*

(3) *s_i is not connected to s_{i+1}, for every $i = 1, \dots, n$,*

then q is $(\lambda_0, 0)$-quasigeodesic in $\Gamma(G, X \cup \mathcal{H})$ without backtracking.

We will actually need a slightly more general version of Proposition 5.18, as follows.

Proposition 5.19 *There exist constants $\lambda \geq 1$ and $c \geq 0$ such that for every $\rho \geq 0$ there is $\zeta_1 > 0$ such that the following holds. Suppose that $p = a_0 b_1 a_1 \cdots b_n a_n$ is a concatenation of geodesic paths $a_0, b_1, \dots, b_n, a_n$ in $\Gamma(G, X \cup \mathcal{H})$ such that*

(1) *b_i is an $\mathcal{H}$-subpath of p, for each $i = 1, \dots, n$,*

(2) *$|b_i|_X \geq \zeta_1$, for each $i = 1, \dots, n$;*

(3) *b_i is not connected to b_{i+1}, for every $i = 1, \dots, n-1$;*

(4) *if b_i is connected to a component h of a_i or a_{i-1} then $|h|_X \leq \rho$, $i = 1, \dots, n$.*

Then p is a (λ, c)-quasigeodesic without backtracking.

Proof The argument below employs the following trick: for each $i = 1, \dots, n$, we replace the $\mathcal{H}$-component of p containing b_i by a single edge s_i, and then embed the resulting path p' into a larger path q to which Proposition 5.18 can be applied. Since a subpath of a (λ, c)-quasigeodesic path without backtracking is again (λ, c)-quasigeodesic and without backtracking, this will complete the proof. In order to construct the path q we add an extra infinite peripheral subgroup Z by embedding G into a larger relatively hyperbolic group G_1.

Let us consider the free product $G_1 = G * Z$, where $Z = \langle z \rangle$ is an infinite cyclic group. Since G is hyperbolic relative to the family $\{H_\nu \mid \nu \in \mathcal{N}\}$, the group G_1 is hyperbolic relative to the union $\{H_\nu \mid \nu \in \mathcal{N}\} \cup \{Z\}$ (this can be fairly easily deduced from the definition or from many existing combination theorems for relatively hyperbolic groups, eg [45, Corollary 1.5]).

Note that G embeds in G_1 and G_1 is generated by the finite set $X' = X \sqcup \{z\}$. Let $\mathcal{H}' = \mathcal{H} \sqcup Z \setminus \{1\}$, so that the Cayley graph $\Gamma(G, X \cup \mathcal{H})$ is naturally a subgraph of the Cayley graph $\Gamma(G_1, X' \cup \mathcal{H}')$. Therefore we can think of p as a path in $\Gamma(G_1, X' \cup \mathcal{H}')$.

The normal form theorem for free products [35, Theorem IV.1.2] implies that the embedding of G into G_1 is isometric with respect to both proper and relative metrics; more precisely,

$$d_X(g,h) = d_{X'}(g,h) \quad \text{and} \quad d_{X\cup\mathcal{H}}(g,h) = d_{X'\cup\mathcal{H}'}(g,h) \quad \text{for all } g,h \in G. \tag{5-5}$$

An alternative way to see this is to use the retraction $r\colon G_1 \to G$, such that $r(x) = x$ for all $x \in X$ and $r(z) = 1$. Then $r(X') = X \cup \{1\}$, $r(H_\nu) = H_\nu$, for all $\nu \in \mathcal{N}$, and $r(Z) = \{1\}$.

Let $\zeta_0 \geq 0$ and $\lambda_0 \geq 1$ be the constants provided by Proposition 5.18 applied to the group G_1, its finite generating set X' and its Cayley graph $\Gamma(G_1, X' \cup \mathcal{H}')$. Set $\zeta_1 = \zeta_0 + 2\rho + 1 > 0$.

For each $i = 1, \dots, n$, let t_i denote the H_{ν_i}-component of p containing the edge b_i, $\nu_i \in \mathcal{N}$. Note that $t_1, \dots, t_n$ are pairwise distinct by condition (3), in particular no two of them share a common edge. In view of Remark 5.9, for every $i = 1, \dots, n$ we can represent t_i as a concatenation $t_i = h_{i-1} b_i f_i$, where

- h_{i-1} is either the last edge and an H_{ν_i}-component of a_{i-1} if a_{i-1} ends with an H_{ν_i}-component, or h_{i-1} is the trivial path, consisting of the vertex $(a_{i-1})_+$, if a_{i-1} does not end with an H_{ν_i}-component;
- f_i is the first edge and an H_{ν_i}-component of a_i if a_i starts with an H_{ν_i}-component, or f_i is the trivial path, consisting of the vertex $(a_i)_-$, if a_i does not start with an H_{ν_i}-component.

Note that for each $i = 1, \dots, n$ we have $|h_{i-1}|_X \leq \rho$ and $|f_i|_X \leq \rho$, by condition (4). By (2) and the triangle inequality we get

$$|t_i|_X \geq |b_i|_X - 2\rho \geq \zeta_0 + 1 \quad \text{for } i = 1, \dots, n. \tag{5-6}$$

Therefore p decomposes as a concatenation

$$p = r_0 t_1 r_1 \cdots t_n r_n,$$

where r_i is a subpath of a_i, $i = 0, \dots, n$, such that $a_0 = r_0 h_0, a_1 = f_1 r_1 h_1, \dots, a_n = f_n r_n$.

By (5-6) the endpoints of the H_{ν_i}-component t_i of p must be distinct; hence there is an edge s_i joining them in $\Gamma(G, X \cup \mathcal{H})$, with $\mathrm{Lab}(s_i) \in H_{\nu_i} \setminus \{1\}$, $i = 1, \dots, n$. Now, (5-6) and (5-5) imply that

$$|s_i|_{X'} = |t_i|_{X'} = |t_i|_X \geq \zeta_0 \quad \text{for } i = 1, \dots, n.$$

Choose $k \in \mathbb{N}$ so that $|z^k|_{X'} \geq \zeta_0$ and let s_{n+1} be the edge in $\Gamma(G_1, X' \cup \mathcal{H}')$, starting at $p_+ = (r_n)_+$ and labelled by z^k. Observe that $|s_{n+1}|_{X'} = |z^k|_{X'} \geq \zeta_0$.

Consider the path q in $\Gamma(G_1, X' \cup \mathcal{H}')$, defined as the concatenation $q = r_0 s_1 \cdots r_n s_{n+1}$. By (5-5) the paths $r_0, \ldots, r_n$ are still geodesic in $\Gamma(G_1, X' \cup \mathcal{H}')$, and $s_1, \ldots, s_{n+1}$ are $\mathcal{H}'$-components of q, by construction. Finally, s_i is not connected to s_{i+1}, for $i = 1, \ldots, n-1$, because elements of G that belong to different H_ν-cosets continue to do so in G_1, and s_n is not connected to s_{n+1} because H_{ν_n} and Z are distinct peripheral subgroups of G_1. Therefore all of the assumptions of Proposition 5.18 are satisfied, which allows us to conclude that the path q is $(\lambda_0, 0)$-quasigeodesic without backtracking in $\Gamma(G_1, X' \cup \mathcal{H}')$.

Consequently, the path $p' = r_0 s_1 r_1 \cdots s_n r_n$ is $(\lambda_0, 0)$-quasigeodesic without backtracking in $\Gamma(G_1, X' \cup \mathcal{H}')$, as a subpath of q. Since p' only contains vertices and edges from $\Gamma(G, X \cup \mathcal{H})$, we see that p' is also $(\lambda_0, 0)$-quasigeodesic without backtracking in $\Gamma(G, X \cup \mathcal{H})$.

Now, the original path p can be obtained by replacing the edges $s_1, \ldots, s_n$ of p' by paths $t_1, \ldots, t_n$, each of which has length at most 3. Hence, by Lemma 4.4, p is $(3\lambda_0, 18\lambda_0 + 6)$-quasigeodesic. Since p' is without backtracking and every $\mathcal{H}$-component of p is connected to an $\mathcal{H}$-component of p' (and vice versa), by construction, the path p must also be without backtracking.

Thus we have shown that the path p is (λ, c)-quasigeodesic without backtracking in $\Gamma(G, X \cup \mathcal{H})$, where $\lambda = 3\lambda_0$ and $c = 18\lambda_0 + 6$. □

5.4 Quasiconvex subsets in relatively hyperbolic groups

In this paper we shall use the definition of a relatively quasiconvex subgroup given by Osin in [46]. For convenience we state it in the case of arbitrary subsets rather than just subgroups.

Definition 5.20 (relatively quasiconvex subset) A subset $Q \subseteq G$ is said to be *relatively quasiconvex* (with respect to $\{H_\nu \mid \nu \in \mathcal{N}\}$) if there exists $\varepsilon \geq 0$ such that for every geodesic path q in $\Gamma(G, X \cup \mathcal{H})$, with $q_-, q_+ \in Q$, and every vertex v of q we have $d_X(v, Q) \leq \varepsilon$.

Any number $\varepsilon \geq 0$ as above will be called a *quasiconvexity constant* of Q.

Osin proved that relative quasiconvexity of a subset is independent of the choice of a finite generating set X of G; see [46, Proposition 4.10] — the proof there is stated for relatively quasiconvex subgroups but actually works more generally for relatively quasiconvex subsets.

We outline some basic properties of quasiconvex subsets and subgroups of G in the next two lemmas.

Lemma 5.21 *Let Q be a relatively quasiconvex subset of G. Then*

(a) *the subset gQ is relatively quasiconvex, for every $g \in G$;*

(b) *if $T \subseteq G$ lies at a finite d_X-Hausdorff distance from Q then T is relatively quasiconvex.*

Proof Claim (a) follows immediately from the fact that left multiplication by g induces an isometry of G with respect to both the proper metric d_X and the relative metric $d_{X\cup\mathcal{H}}$.

To prove claim (b), suppose that $\varepsilon \geq 0$ is a quasiconvexity constant of Q and the d_X-Hausdorff distance between Q and T is less than $k \in \mathbb{N}$. Consider any geodesic path t in $\Gamma(G, X \cup \mathcal{H})$ with $t_-, t_+ \in T$, and take any vertex v of t. Then there are $x, y \in Q$ such that $d_X(x, t_-) \leq k$ and $d_X(y, t_+) \leq k$. Let q be any geodesic connecting x with y. Then q is k-similar to t, hence there is a vertex u of q such that $d_X(v, u) \leq \kappa$, where $\kappa = \kappa(1, 0, k) \geq 0$ is the global constant given by Proposition 5.17 applied to k-similar geodesics. By the relative quasiconvexity of Q, there exists $w \in Q$ such that $d_X(u, w) \leq \varepsilon$. Moreover, $d_X(w, T) \leq k$ by assumption. Therefore $d_X(v, T) \leq \kappa + \varepsilon + k$, thus T is relatively quasiconvex in G. □

Lemma 5.22 *Suppose that $Q \leqslant G$ is a relatively quasiconvex subgroup. Then for all $g \in G$ and $Q' \leqslant_f Q$ the subgroups gQg^{-1} and Q' are relatively quasiconvex in G.*

Proof By claim (a) of Lemma 5.21, the coset gQ is relatively quasiconvex and the d_X-Hausdorff distance between this coset and gQg^{-1} is at most $|g|_X$; hence gQg^{-1} is relatively quasiconvex in G by claim (b) of the same lemma.

Suppose that $Q = \bigcup_{i=1}^{m} Q'h_i$, where $h_i \in Q$, $i = 1, \ldots, m$. Then the d_X-Hausdorff distance between Q and Q' is bounded above by $\max\{|h_i|_X \mid 1 \leq i \leq m\}$, so Q' is relatively quasiconvex by Lemma 5.21(b). □

Corollary 5.23 *Any parabolic subgroup of G is relatively quasiconvex.*

Proof Let $H = gQg^{-1}$ be a parabolic subgroup, where $g \in G$ and $Q \leqslant H_\nu$, for some $\nu \in \mathcal{N}$. The subgroup Q is relatively quasiconvex in G (with quasiconvexity constant 0), because any geodesic connecting two elements of Q consists of a single edge in $\Gamma(G, X \cup \mathcal{H})$. Therefore H is relatively quasiconvex by Lemma 5.22. □

Lemma 5.24 *Let P be a maximal parabolic subgroup of G and let Q be a finitely generated relatively quasiconvex subgroup of G. Then the subgroups P and $Q \cap P$ are finitely generated.*

Proof The fact that each H_ν is finitely generated, provided G is finitely generated, was proved by Osin in [46, Theorem 1.1].

Now, Hruska [30, Theorem 9.1] proved that every quasiconvex subgroup Q of G is itself relatively hyperbolic and maximal parabolic subgroups of Q are precisely the infinite intersections of Q with maximal parabolic subgroups of G. In other words, if $P \leqslant G$ is maximal parabolic, then $Q \cap P$ is either finite or a maximal parabolic subgroup of Q. Combined with Osin's result [46, Theorem 1.1] mentioned above we can conclude that if Q is finitely generated then so is $Q \cap P$, as required. □

The following property of quasiconvex subgroups will be useful.

Lemma 5.25 *Let $Q, R \leqslant G$ be relatively quasiconvex subgroups of G. For every $\zeta \geq 0$ there exists a constant $\mu = \mu(\zeta) \geq 0$ such that the following holds.*

Suppose $x \in G$, $a \in Q$ and $b \in R$ are some elements, $[x, xa]$ and $[x, xb]$ are geodesic paths in $\Gamma(G, X \cup \mathcal{H})$, and $u \in [x, xa]$ and $v \in [x, xb]$ are vertices such that $d_X(u, v) \leq \zeta$. Then there is an element $z \in x(Q \cap R)$ such that $d_X(u, z) \leq \mu$ and $d_X(v, z) \leq \mu$.

Proof Denote by $\varepsilon \geq 0$ a quasiconvexity constant of the subgroups Q and R. After applying the left translation by x^{-1}, which is an isometry with respect to both metrics d_X and $d_{X \cup \mathcal{H}}$, we can assume that $x = 1$. Let $K' = K'(Q, R, \varepsilon + \zeta)$ be the constant given by Lemma 4.1.

Since $x = 1 \in Q \cap R$, $xa = a \in Q$ and $xb = b \in R$, by the relative quasiconvexity of Q and R we know that $u \in N_X(Q, \varepsilon)$ and $v \in N_X(R, \varepsilon)$. By the assumptions $d_X(u, v) \leq \zeta$, it follows that $u \in N_X(Q, \varepsilon + \zeta) \cap N_X(R, \varepsilon + \zeta)$; hence $u \in N_X(Q \cap R, K')$ by Lemma 4.1.

Thus there exists $z \in Q \cap R$ such that $d_X(u, z) \leq K'$, and, hence, $d_X(v, z) \leq K' + \zeta$ by the triangle inequality. Therefore the statement of the lemma holds for $\mu = K' + \zeta$. □

The next combination theorem was proved by Martínez-Pedroza.

Theorem 5.26 [37, Theorem 1.1] *Let G be a relatively hyperbolic group generated by a finite set X. Suppose that Q is a relatively quasiconvex subgroup of G, P is a maximal parabolic subgroup of G and $D = Q \cap P$. There is a constant $C \geq 0$ such that the following holds. If $H \leqslant P$ is any subgroup satisfying*

(1) *$H \cap Q = D$, and*

(2) *$\min_X(H \setminus D) \geq C$,*

then the subgroup $A = \langle H, Q \rangle$ is relatively quasiconvex in G and is naturally isomorphic to the amalgamated free product $H *_D Q$.

Moreover, for every maximal parabolic subgroup T of G, there exists $u \in A$ such that either

$$A \cap T \subseteq uQu^{-1} \quad \text{or} \quad A \cap T \subseteq uHu^{-1}.$$

Part II Quasiconvexity of virtual joins

This part of the paper is mostly devoted to the proofs of Theorems 3.5 and 1.2. Let us start by giving brief outlines of the arguments.

Suppose G is a group generated by finite set X and hyperbolic relative to a collection of subgroups $\{H_\nu \mid \nu \in \mathcal{N}\}$. Denote $\mathcal{H} = \bigsqcup_{\nu \in \mathcal{N}} H_\nu \setminus \{1\}$ and take any $A \geq 0$. Consider two finitely generated relatively quasiconvex subgroups $Q, R \leqslant G$. Set $S = Q \cap R$ and suppose that $Q' \leqslant Q$ and $R' \leqslant R$ are subgroups satisfying conditions (C1)–(C5) from Section 3.1, with some finite collection of maximal parabolic subgroups $\mathcal{P}$ of G (which is independent of A) and parameters B and C that are sufficiently large with respect to A.

Every element $g \in \langle Q', R' \rangle$ can be written as a product of elements of Q' and R', which gives rise to a broken geodesic line in $\Gamma(G, X \cup \mathcal{H})$ (not necessarily uniquely), whose label represents g in G. We choose a path p from the collection of such broken lines, representing g, that is minimal in a certain sense. The path p may fail to be uniformly quasigeodesic, as it may travel through H_ν-cosets for an arbitrarily long time. We do, however, have some metric control over such instances of backtracking, using the fact that Q' and R' satisfy conditions (C1)–(C5) and the minimality of p.

We construct a new path from p, which we call the *shortcutting* of p, that turns out to be uniformly quasigeodesic. Informally speaking, the shortcutting of p is obtained by replacing each maximal instance of backtracking in consecutive geodesic segments of p with a single edge, then connecting these edges in sequence by geodesics. The resulting path can be seen to satisfy the hypotheses of Proposition 5.19. It follows that the shortcutting of p is uniformly quasigeodesic, and hence $\langle Q', R' \rangle$ is relatively quasiconvex. Properties (P2) and (P3) also follow from this quasigeodesicity, giving us Theorem 3.5.

Now suppose that G is QCERF and its peripheral subgroups are double coset separable. In Theorem 11.3 we use the separability assumptions on G and $\{H_\nu \mid \nu \in \mathcal{N}\}$ to deduce the existence of a finite-index subgroup $M \leqslant_f G$ such that $Q' = Q \cap M \leqslant_f Q$, $R' = R \cap M \leqslant_f R$ satisfy conditions (C1)–(C5) with constants B and C large enough to apply Theorem 3.5 (as suggested in Remark 3.4). Conditions (C1) and (C4) are essentially automatic. Conditions (C2), (C3) and (C5) can be assured to hold for the subgroups Q' and R' using Lemma 4.16 by the QCERF condition on G, separability of double cosets PS (where P is one of finitely many maximal parabolic subgroups) and double coset separability of the peripheral subgroups, respectively.

The remaining technical difficulty is in showing that the double cosets of the form PS as above are separable in G. To this end, we prove a general result about lifting separability of certain double cosets in amalgamated free products. This is then combined with a result of Martínez-Pedroza (Theorem 5.26), allowing us to deduce Theorem 1.2 from Theorem 3.5.

6 Path representatives

Let us set the notation that will be used in the next few sections.

Convention 6.1 We fix a group G, generated by a finite set X, which is hyperbolic relative to a finite family of subgroups $\{H_\nu \mid \nu \in \mathcal{N}\}$. We let $\mathcal{H} = \bigsqcup_{\nu \in \mathcal{N}} (H_\nu \setminus \{1\})$. It follows that the Cayley graph $\Gamma(G, X \cup \mathcal{H})$ is δ-hyperbolic, for some $\delta \in \mathbb{N}$ (see Lemma 5.4).

Furthermore, we assume that $Q, R \leqslant G$ are fixed relatively quasiconvex subgroups of G, with a quasiconvexity constant $\varepsilon \geq 0$, and denote $S = Q \cap R$.

In this section Q' and R' will denote some subgroups of Q and R respectively. We will introduce path representatives of elements in $\langle Q', R' \rangle$ and will order such representatives by their types. This will be crucial in our proof of Theorem 3.5.

Definition 6.2 (path representative, I) Consider an arbitrary element $g \in \langle Q', R' \rangle$. Let $p = p_1 \cdots p_n$ be a broken line in $\Gamma(G, X \cup \mathcal{H})$ with geodesic segments $p_1, \dots, p_n$, such that $\tilde{p} = g$ and $\tilde{p}_i \in Q' \cup R'$ for each $i \in \{1, \dots, n\}$. We will call p a *path representative* of g.

To choose an optimal path representative we define their types.

Definition 6.3 (type of a path representative, I) Suppose that $p = p_1 \cdots p_n$ is a broken line in $\Gamma(G, X \cup \mathcal{H})$. For each $i = 1, \dots, n$, let T_i denote the set of all $\mathcal{H}$-components of p_i, and let $T = \bigcup_{i=1}^n T_i$. We define the *type* $\tau(p)$ of p to be the triple

$$\tau(p) = \left(n, \ell(p), \sum_{t \in T} |t|_X \right) \in {\mathbb{N}_0}^3,$$

where $\ell(p) = \sum_{i=1}^n \ell(p_i)$ is the length of p.

Definition 6.4 (minimal type) Given $g \in \langle Q', R' \rangle$, the set $\mathcal{S}$ of all path representatives of g is non-empty. Therefore the subset $\tau(\mathcal{S}) = \{\tau(p) \mid p \in \mathcal{S}\} \subseteq {\mathbb{N}_0}^3$, where ${\mathbb{N}_0}^3$ is equipped with the lexicographic order, will have a unique minimal element.

We will say that $p = p_1 \cdots p_n$ is a *path representative of g of minimal type* if $\tau(p)$ is the minimal element of $\tau(\mathcal{S})$.

Remark 6.5 If p_1 and p_2 are paths with $(p_1)_+ = (p_2)_-$ whose labels both represent elements of Q' (or, respectively, both of R'), then the label of any geodesic $[(p_1)_-, (p_2)_+]$ also represents an element of Q' (respectively, R'). Hence in a path representative of $g \in \langle Q', R' \rangle$ of minimal type, the labels of the consecutive segments necessarily alternate between representing elements of $Q' \setminus (Q' \cap R')$ and $R' \setminus (Q' \cap R')$, whenever g is not itself an element of $Q' \cap R'$.

The minimality of the type of a path representative is thus a numerical condition on the total lengths of the paths p_i and the total lengths of their components. In the next few sections we will study local properties induced by this global condition. The first such property is stated in the next lemma.

Notation 6.6 Let $x, y, z \in G$. We will write $\langle x, y \rangle_z^{\mathrm{rel}} = \frac{1}{2}(d_{X \cup \mathcal{H}}(x, z) + d_{X \cup \mathcal{H}}(y, z) - d_{X \cup \mathcal{H}}(x, y))$ to denote the Gromov product of x and y with respect to z in the relative metric $d_{X \cup \mathcal{H}}$.

Lemma 6.7 (Gromov products are bounded) *There is a constant $C_0 \geq 0$ such that the following holds.*

Let $Q' \leqslant Q$ and $R' \leqslant R$ be subgroups satisfying condition (C1). If $p = p_1 \cdots p_n$ is a minimal type path representative of an element $g \in \langle Q', R' \rangle$ and $f_0, \dots, f_n \in G$ are the nodes of p (that is, $f_{i-1} = (p_i)_-$, for $i = 1, \dots, n$, and $f_n = (p_n)_+$) then $\langle f_{i-1}, f_{i+1} \rangle_{f_i}^{\mathrm{rel}} \leq C_0$ for each $i = 1, \dots, n-1$.

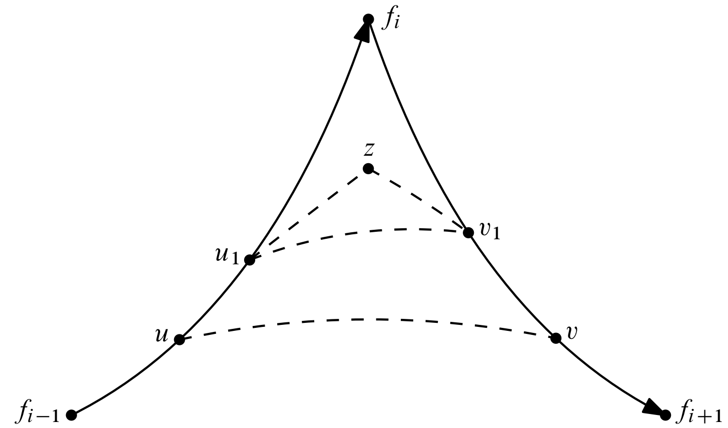


Figure 1: We obtain a different path representative for g by replacing p_i and p_{i+1} with geodesics from f_{i-1} to z to f_{i+1}.

Proof Let $\sigma \in \mathbb{N}_0$ be the constant from Proposition 5.15 and let $\mu = \mu(\sigma) \geq 0$ be given by Lemma 5.25. Set $C_0 = \mu + \delta + 2\sigma + 2$, and assume that $p = p_1 \cdots p_n$ is a path representative of $g \in \langle Q', R' \rangle$ of minimal type.

Take any $i \in \{1, \dots, n-1\}$. Choose vertices $u \in p_i$ and $v \in p_{i+1}$ so that

$$d_{X \cup \mathcal{H}}(f_i, u) = d_{X \cup \mathcal{H}}(f_i, v) = \lfloor \langle f_{i-1}, f_{i+1} \rangle_{f_i}^{\mathrm{rel}} \rfloor.$$

As $\Gamma(G, X \cup \mathcal{H})$ is δ-hyperbolic, we must have $d_{X \cup \mathcal{H}}(u, v) \leq \delta$.

If $\langle f_{i-1}, f_{i+1} \rangle_{f_i}^{\mathrm{rel}} < C_0$ then we are done, so suppose otherwise. Then $d_{X \cup \mathcal{H}}(u, f_i) \geq \delta + \sigma + 1 \in \mathbb{N}$, so there is a vertex u_1 on the subpath $[u, f_i]$ of p_i such that

$$d_{X \cup \mathcal{H}}(u_1, u) = \delta + \sigma + 1.$$

Applying Proposition 5.15 to the geodesic triangle Δ with sides $[u, f_i]$, $[f_i, v]$ and $[u, v]$ (here we choose $[f_i, v]$ to be a subpath of p_{i+1}), we can find some vertex $v_1 \in [u, v] \cup [f_i, v]$ with $d_X(v_1, u_1) \leq \sigma$. If $v_1 \in [u, v]$, then, by the triangle inequality,

$$d_{X \cup \mathcal{H}}(u_1, u) \leq d_{X \cup \mathcal{H}}(u_1, v_1) + d_{X \cup \mathcal{H}}(u, v) \leq \sigma + \delta,$$

which would contradict the choice of u_1. Therefore it must be that $v_1 \in [f_i, v]$ (see Figure 1).

Since the path representative p has minimal type, in view of Remark 6.5 we must have either $\tilde{p}_i \in Q'$ and $\tilde{p}_{i+1} \in R'$ or $\tilde{p}_i \in R'$ and $\tilde{p}_{i+1} \in Q'$. Without loss of generality let us assume the former. We can apply Lemma 5.25 to find $z \in f_i(Q \cap R)$ with $d_X(u_1, z) \leq \mu$ and $d_X(v_1, z) \leq \mu$. Let p_i' be a geodesic path in $\Gamma(G, X \cup \mathcal{H})$ joining $f_{i-1} = (p_i)_-$ with z and let p_{i+1}' be a geodesic path joining z with $f_{i+1} = (p_{i+1})_+$. Observe that $f_{i-1} \in f_i Q'$ and $Q \cap R \subseteq Q'$ by (C1), whence

$$\tilde{p}_i' = f_{i-1}^{-1} z \in Q' f_i^{-1} f_i (Q \cap R) = Q'.$$

Similarly, $\tilde{p}'_{i+1} \in R'$. It follows that the path $p' = p_1 \cdots p_{i-1} p'_i p'_{i+1} p_{i+2} \cdots p_n$ is also a path representative of the same element $g \in \langle Q', R' \rangle$.

Since p has minimal type, by the assumption, it must be that $\ell(p_i) + \ell(p_{i+1}) \le \ell(p'_i) + \ell(p'_{i+1})$, which can be rewritten as

$$d_{X \cup \mathcal{H}}(f_{i-1}, f_i) + d_{X \cup \mathcal{H}}(f_i, f_{i+1}) \le d_{X \cup \mathcal{H}}(f_{i-1}, z) + d_{X \cup \mathcal{H}}(z, f_{i+1}). \tag{6-1}$$

Since $u_1 \in p_i$, we have $d_{X \cup \mathcal{H}}(f_{i-1}, f_i) = d_{X \cup \mathcal{H}}(f_{i-1}, u_1) + d_{X \cup \mathcal{H}}(u_1, f_i)$. On the other hand,

$$d_{X \cup \mathcal{H}}(f_{i-1}, z) \le d_{X \cup \mathcal{H}}(f_{i-1}, u_1) + d_{X \cup \mathcal{H}}(u_1, z) \le d_{X \cup \mathcal{H}}(f_{i-1}, u_1) + \mu,$$

by the triangle inequality. Similarly,

$$d_{X \cup \mathcal{H}}(f_i, f_{i+1}) = d_{X \cup \mathcal{H}}(f_i, v_1) + d_{X \cup \mathcal{H}}(v_1, f_{i+1}) \quad \text{and} \quad d_{X \cup \mathcal{H}}(z, f_{i+1}) \le d_{X \cup \mathcal{H}}(v_1, f_{i+1}) + \mu.$$

Combining the above inequalities with (6-1), we obtain

$$d_{X \cup \mathcal{H}}(u_1, f_i) + d_{X \cup \mathcal{H}}(f_i, v_1) \le 2\mu. \tag{6-2}$$

Now, by construction, we have

$$d_{X \cup \mathcal{H}}(u_1, f_i) = d_{X \cup \mathcal{H}}(u, f_i) - d_{X \cup \mathcal{H}}(u_1, u) = \lfloor \langle f_{i-1}, f_{i+1} \rangle^{\mathrm{rel}}_{f_i} \rfloor - (\delta + \sigma + 1). \tag{6-3}$$

On the other hand, since $d_{X \cup \mathcal{H}}(v_1, u_1) \le \sigma$, we achieve

$$d_{X \cup \mathcal{H}}(f_i, v_1) \ge d_{X \cup \mathcal{H}}(u_1, f_i) - d_{X \cup \mathcal{H}}(v_1, u_1) \ge \lfloor \langle f_{i-1}, f_{i+1} \rangle^{\mathrm{rel}}_{f_i} \rfloor - (\delta + 2\sigma + 1). \tag{6-4}$$

After combining (6-3), (6-4) and (6-2), we obtain

$$2\lfloor \langle f_{i-1}, f_{i+1} \rangle^{\mathrm{rel}}_{f_i} \rfloor - (2\delta + 3\sigma + 2) \le 2\mu.$$

Therefore, we can conclude that $\langle f_{i-1}, f_{i+1} \rangle^{\mathrm{rel}}_{f_i} \le \mu + \delta + 2\sigma + 2 = C_0$, as required. □

7 Adjacent backtracking in path representatives of minimal type

In this section we continue working under Convention 6.1. Our goal here is to study the possible backtracking within two adjacent segments in a minimal type path representative.

Lemma 7.1 *For all non-negative numbers ζ and ξ there exists $\tau = \tau(\zeta, \xi) \ge 0$ such that the following holds.*

Suppose that $Q' \leqslant Q$ and $R' \leqslant R$ are subgroups satisfying (C1)*, $g \in \langle Q', R' \rangle$ and $p = p_1 \cdots p_n$ is a path representative of g of minimal type. If for some $i \in \{1, \dots, n-1\}$, s and t are connected $\mathcal{H}$-components of p_i and p_{i+1} respectively, such that $d_X(s_-, t_+) \le \zeta$ and $d_X(s_+, (p_i)_+) \le \xi$, then $|s|_X \le \tau$ and $|t|_X \le \tau$.*

Proof Let $\mu = \mu(\zeta) \ge 0$ be the constant from Lemma 5.25. Since $|X| < \infty$ and $|\mathcal{N}| < \infty$ we can define the constant $k \ge 0$ as

$$k = \max\{K'(Q \cap R, cH_\nu c^{-1}, \xi + \mu) \mid \nu \in \mathcal{N},\ c \in G,\ |c|_X \le \xi\}, \tag{7-1}$$

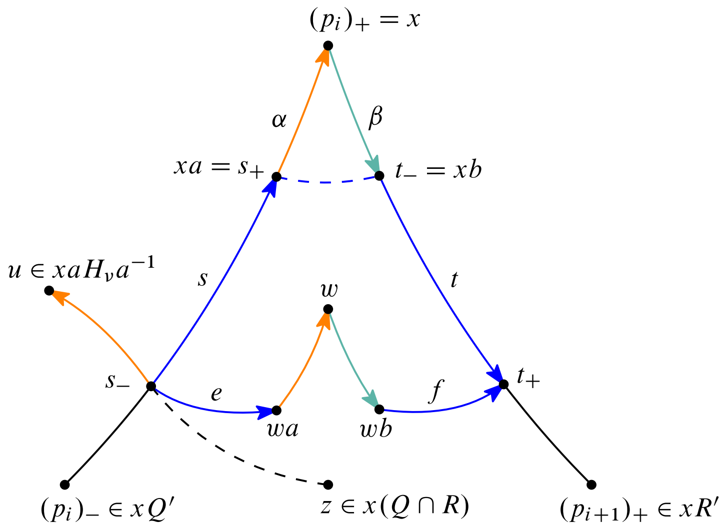


Figure 2: Illustration of Lemma 7.1.

where for each $c \in G$ and $\nu \in \mathcal{N}$ the constant $K'(Q \cap R, cH_\nu c^{-1}, \xi + \mu)$ is given by Lemma 4.1. Let $L \geq 0$ be the constant from Lemma 5.14 and set $\tau = 2k + 2\xi + \zeta + L \geq 0$.

Let $p = p_1 \cdots p_n$ be a path representative of some $g \in \langle Q', R' \rangle$ of minimal type. Suppose that s and t are connected H_ν-components of p_i and p_{i+1} respectively, for some $i \in \{1, \dots, n-1\}$ and $\nu \in \mathcal{N}$, such that $d_X(s_-, t_+) \leq \zeta$ and $d_X(s_+, (p_i)_+) \leq \xi$.

Note that, by Lemma 5.14,

$$d_X(s_+, t_-) \leq L. \tag{7-2}$$

Denote $x = (p_i)_+ = (p_{i+1})_- \in G$, $a = x^{-1}s_+ \in G$ and $b = x^{-1}t_- \in G$; see Figure 2.

Note that

$$aH_\nu = bH_\nu, \quad \text{hence } aH_\nu a^{-1} = bH_\nu b^{-1}, \tag{7-3}$$

because the H_ν-components s and t are connected. Using the lemma hypotheses and (7-2) we also have

$$|a|_X = d_X(x, s_+) \leq \xi \quad \text{and} \quad |b|_X \leq d_X(x, s_+) + d_X(s_+, t_-) \leq \xi + L. \tag{7-4}$$

In view of Remark 6.5, without loss of generality we can assume that $\mathrm{Lab}(p_i)$ represents an element of Q' and $\mathrm{Lab}(p_{i+1})$ represents an element of R' in G (the other case can be treated similarly). Applying Lemma 5.25, we can find $z \in x(Q \cap R)$ such that $d_X(s_-, z) \leq \mu$.

Consider the element $u = s_- a^{-1} = xa\tilde{s}^{-1}a^{-1} \in xaH_\nu a^{-1}$, and observe that $d_X(s_-, u) = |a^{-1}|_X \leq \xi$. On the other hand, $d_X(s_-, x(Q \cap R)) \leq d_X(s_-, z) \leq \mu$, whence

$$s_- \in N_X(x(Q \cap R), \xi + \mu) \cap N_X(xaH_\nu a^{-1}, \xi + \mu).$$

Therefore, according to Lemma 4.1, there exists $w \in x(Q \cap R \cap aH_\nu a^{-1})$ such that

$$d_X(s_-, w) \le k, \tag{7-5}$$

where $k \ge 0$ is the constant defined in (7-1).

Let α be the subpath of p_i from $s_+ = xa$ to $(p_i)_+ = x$. Choose the geodesic path $[wa, w]$ as the translate $wx^{-1}\alpha$. Observe that $s_- \in xaH_\nu$ and $wa \in xaH_\nu a^{-1}a = xaH_\nu$ lie in the same H_ν-coset. Thus $d_{X\cup\mathcal{H}}(s_-, wa) \le 1$; if $s_- = wa$ we let e be the trivial path in $\Gamma(G, X \cup \mathcal{H})$ consisting of the single vertex s_-, and otherwise we let e be the edge of $\Gamma(G, X \cup \mathcal{H})$ labelled by an element of $H_\nu \setminus \{1\}$ that joins s_- to wa. Define the path q in $\Gamma(G, X \cup \mathcal{H})$ as the concatenation

$$q = [(p_i)_-, s_-]e[wa, w], \tag{7-6}$$

where $[(p_i)_-, s_-]$ is chosen as the initial segment of p_i.

Since $\ell(e) \le 1 = d_{X\cup\mathcal{H}}(s_-, s_+)$, we can bound the length of the path q from above as follows:

$$\begin{aligned} \ell(q) &= d_{X\cup\mathcal{H}}((p_i)_-, s_-) + \ell(e) + d_{X\cup\mathcal{H}}(wa, w) \\ &\le d_{X\cup\mathcal{H}}((p_i)_-, s_-) + d_{X\cup\mathcal{H}}(s_-, s_+) + d_{X\cup\mathcal{H}}(xa, x) = \ell(p_i). \end{aligned} \tag{7-7}$$

Now we construct a similar path from w to $(p_{i+1})_+$. Let β be the subpath of p_{i+1} from $(p_{i+1})_- = x$ to $t_- = xb$. Choose the geodesic path $[w, wb]$ as the translate $wx^{-1}\beta$. Recall that $t_+ \in xbH_\nu$ and note that the inclusion $w \in xaH_\nu a^{-1}$, together with (7-3), imply that $wb \in xbH_\nu$ also. If $t_+ = wb$ then let f be the trivial path in $\Gamma(G, X \cup \mathcal{H})$ consisting of the single vertex t_+, otherwise let f be the edge in $\Gamma(G, X \cup \mathcal{H})$ joining the vertices wb and t_+ with $\mathrm{Lab}(f) \in H_\nu \setminus \{1\}$. We now define the path r in $\Gamma(G, X \cup \mathcal{H})$ as the concatenation

$$r = [w, wb]f[t_+, (p_{i+1})_+], \tag{7-8}$$

where $[t_+, (p_{i+1})_+]$ is chosen as the ending segment of p_{i+1}. Similarly to the case of q we can estimate that

$$\ell(r) \le \ell(p_{i+1}). \tag{7-9}$$

Note that since $q_- = (p_i)_- = x\tilde{p}_i^{-1} \in xQ'$, $q_+ = w \in x(Q \cap R)$ and $Q \cap R \subseteq Q'$, we have $\tilde{q} \in Q'$. Similarly, $\tilde{r} \in R'$.

Let p_i' be a geodesic path from $q_- = (p_i)_-$ to $q_+ = w$, and let p_{i+1}' be a geodesic path from $w = r_-$ to $(p_{i+1})_+ = r_+$. Since $\tilde{p}_i' = \tilde{q} \in Q'$ and $\tilde{p}_{i+1}' = \tilde{r} \in R'$, the broken line $p' = p_1 \cdots p_{i-1} p_i' p_{i+1}' p_{i+2} \cdots p_n$ is a path representative of the same element $g \in G$.

If at least one of the paths q, r is not geodesic in $\Gamma(G, X \cup \mathcal{H})$, then, in view of (7-7) and (7-9) we have

$$\ell(p_i') + \ell(p_{i+1}') < \ell(q) + \ell(r) \le \ell(p_i) + \ell(p_{i+1});$$

hence $\ell(p) = \sum_{i=1}^n \ell(p_i) > \ell(p')$, contradicting the minimality of the type of p.

Hence both q and r must be geodesic in $\Gamma(G, X \cup \mathcal{H})$, so we can further assume that $p'_i = q$ and $p'_{i+1} = r$. Moreover, the inequality $\ell(p) \le \ell(p')$ must hold by the minimality of the type of p. Therefore $\ell(p_i) + \ell(p_{i+1}) \le \ell(q) + \ell(r)$, which, in view of (7-7) and (7-9), implies that $\ell(q) = \ell(p_i)$, $\ell(r) = \ell(p_{i+1})$ and $\ell(p) = \ell(p')$. In particular, e and f are actual edges of $\Gamma(G, X \cup \mathcal{H})$ (and not trivial paths).

The definition (7-6) of q implies that $\mathrm{Lab}(q)$ can differ from $\mathrm{Lab}(p_i)$ in at most one letter, which is the label of the H_ν-component e in $\mathrm{Lab}(q)$ and the label of the H_ν-component s in $\mathrm{Lab}(p_i)$. Indeed,

$$\mathrm{Lab}(p_i) = \mathrm{Lab}([(p_i)_-, s_-])\,\mathrm{Lab}(s)\,\mathrm{Lab}(\alpha) \quad \text{and} \quad \mathrm{Lab}(q) = \mathrm{Lab}([(p_i)_-, s_-])\,\mathrm{Lab}(e)\,\mathrm{Lab}(\alpha),$$

where we used the fact that $[wa, w]$ is the left translate of α, by definition, and hence it has the same label as α.

Similarly, (7-8) implies $\mathrm{Lab}(r)$ can differ from $\mathrm{Lab}(p_i)$ in at most one letter which is the label of f in r and the label of t in p_{i+1}. The minimality of the type of p therefore implies that

$$|s|_X + |t|_X \le |e|_X + |f|_X. \tag{7-10}$$

Now, using the triangle inequality, (7-5) and (7-4) we obtain

$$|e|_X = d_X(s_-, wa) \le d_X(s_-, w) + d_X(w, wa) \le k + |a|_X \le k + \xi. \tag{7-11}$$

To estimate $|f|_X$ we also use the inequality $d_X(s_-, t_+) \le \zeta$:

$$\begin{aligned} |f|_X = d_X(t_+, wb) &\le d_X(t_+, w) + |b|_X \\ &\le d_X(t_+, s_-) + d_X(s_-, w) + \xi + L \le \zeta + k + \xi + L. \end{aligned} \tag{7-12}$$

Combining (7-10)–(7-12) together, we achieve

$$\max\{|s|_X, |t|_X\} \le |e|_X + |f|_X \le 2k + 2\xi + \zeta + L = \tau.$$

This inequality completes the proof of the lemma. □

The following auxiliary definition will only be used in the remainder of this section.

Definition 7.2 Let $C_0 \ge 0$ be the constant provided by Lemma 6.7, let $L \ge 0$ be the constant given by Lemma 5.14 and let $\kappa = \kappa(1, 0, L) \ge 0$ be the constant from Proposition 5.17.

Define the sequences $(\zeta_j)_{j\in\mathbb{N}}$, $(\xi_j)_{j\in\mathbb{N}}$ and $(\tau_j)_{j\in\mathbb{N}}$ of non-negative real numbers as follows.

Set $\zeta_1 = \kappa$, $\xi_1 = C_0 + 1$ and $\tau_1 = \max\{\kappa, \tau(\zeta_1, \xi_1)\}$, where $\tau(\zeta_1, \xi_1)$ is given by Lemma 7.1.

Now suppose that $j > 1$ and the first $j-1$ members of the three sequences have already been defined. Then we set

$$\zeta_j = \kappa, \quad \xi_j = C_0 + 1 + \sum_{k=1}^{j-1} \tau_k, \quad \tau_j = \max\{\kappa, \tau(\zeta_j, \xi_j)\},$$

where $\tau(\zeta_j, \xi_j)$ is given by Lemma 7.1.

Lemma 7.3 *There exists a constant $C_1 \geq 0$ such that the following is true.*

Let $Q' \leqslant Q$ and $R' \leqslant R$ be subgroups satisfying (C1) *and let $p = p_1 \cdots p_n$ be a minimal type path representative for an element $g \in \langle Q', R' \rangle$. Suppose that, for some $i \in \{1, \dots, n-1\}$, q and r are connected $\mathcal{H}$-components of p_i and p_{i+1} respectively. Then $d_X(q_+, (p_i)_+) \leq C_1$ and $d_X((p_i)_+, r_-) \leq C_1$.*

Proof Denote $x = (p_i)_+ = (p_{i+1})_- \in G$. First, let us show that

$$d_{X \cup \mathcal{H}}(q_+, x) \leq C_0 + 1, \tag{7-13}$$

where $C_0 \geq 0$ is the global constant provided by Lemma 6.7. Indeed, the latter lemma states that $\langle (p_i)_-, (p_{i+1})_+ \rangle_x^{\mathrm{rel}} \leq C_0$. Since q_+ and r_- are points on the geodesics p_i and p_{i+1}, Remark 4.6 implies that

$$\langle q_+, r_- \rangle_x^{\mathrm{rel}} \leq \langle (p_i)_-, (p_{i+1})_+ \rangle_x^{\mathrm{rel}} \leq C_0.$$

Consequently,

$$\begin{aligned} C_0 \geq \langle q_+, r_- \rangle_x^{\mathrm{rel}} &= \tfrac{1}{2}\big(d_{X \cup \mathcal{H}}(x, q_+) + d_{X \cup \mathcal{H}}(x, r_-) - d_{X \cup \mathcal{H}}(q_+, r_-)\big) \\ &\geq \tfrac{1}{2}\big(2 d_{X \cup \mathcal{H}}(x, q_+) - 2 d_{X \cup \mathcal{H}}(q_+, r_-)\big) \geq d_{X \cup \mathcal{H}}(x, q_+) - 1, \end{aligned}$$

where the last inequality used the fact that $d_{X \cup \mathcal{H}}(q_+, r_-) \leq 1$, which is true because q and r are connected $\mathcal{H}$-components. This establishes the inequality (7-13).

Let α denote the subpath of p_i starting at q_+ and ending at x, and let β denote the subpath of p_{i+1} starting at x and ending at r_-. Let $s_1, \dots, s_l$, $l \in \mathbb{N}_0$, be the set of all $\mathcal{H}$-components of α listed in the reverse order of their occurrence. That is, s_1 is the last $\mathcal{H}$-component of α (closest to $\alpha_+ = x$) and s_l is the first $\mathcal{H}$-component of α (closest to $\alpha_- = q_+$). Note that, by (7-13),

$$l \leq \ell(\alpha) = d_{X \cup \mathcal{H}}(x, q_+) \leq C_0 + 1. \tag{7-14}$$

Let $L \geq 0$ be the constant given by Lemma 5.14, then

$$d_X(\alpha_-, \beta_+) = d_X(q_+, r_-) \leq L. \tag{7-15}$$

It follows that the geodesic paths α and β^{-1} are L-similar in $\Gamma(G, X \cup \mathcal{H})$. Let $\kappa = \kappa(1, 0, L) \geq 0$ be the constant provided by Proposition 5.17.

We will now prove the following.

Claim 7.4 For each $j = 1, \dots, l$ we have

$$|s_j|_X \leq \tau_j, \tag{7-16}$$

where $\tau_j \geq 0$ is given by Definition 7.2.

We will establish the claim by induction on j. For the base of induction, $j = 1$, note that if $|s_1|_X < \kappa$ then the inequality $|s_1|_X \leq \tau_1$ will be true by definition of τ_1. Thus we can suppose that $|s_1|_X \geq \kappa$. In this case, by Proposition 5.17, s_1 must be connected to some $\mathcal{H}$-component of β^{-1}. Claim (3) of the same proposition implies that there is an $\mathcal{H}$-component t_1 of β, such that s_1 is connected to t_1 and $d_X((s_1)_-, (t_1)_+) \le \kappa$. Note that, by construction, s_1 and t_1 are also connected $\mathcal{H}$-components of p_i and p_{i+1} respectively.

Observe that the subpath of α from $(s_1)_+$ to x is labelled by letters from $X^{\pm 1}$ because it has no $\mathcal{H}$-components. Therefore $d_X((s_1)_+, x) \le \ell(\alpha) \le C_0 + 1$. Consequently, we can apply Lemma 7.1 to deduce that $|s_1|_X \le \tau(\zeta_1, \xi_1)$, where $\zeta_1 = \kappa$ and $\xi_1 = C_0 + 1$.

Thus we have shown that $|s_1|_X \le \tau_1$, where $\tau_1 = \max\{\kappa, \tau(\zeta_1, \xi_1)\}$, and the base of induction has been established.

Now, suppose that $j > 1$ and inequality (7-16) has been proved for all strictly smaller values of j. If $|s_j|_X < \kappa$ then are done, because $\tau_j \ge \kappa$ by definition. So we can assume that $|s_j|_X \ge \kappa$. As before, we can use Proposition 5.17, to find an $\mathcal{H}$-component t_j of β such that s_j is connected to t_j and $d_X((s_j)_-, (t_j)_+) \le \kappa$.

By construction, $s_1, \dots, s_{j-1}$ is the list of all $\mathcal{H}$-components of the subpath $[(s_j)_+, x]$ of α; hence

$$d_X((s_j)_+, x) \le \ell(\alpha) + \sum_{k=1}^{j-1} |s_k|_X \le C_0 + 1 + \sum_{k=1}^{j-1} \tau_k,$$

where the second inequality used (7-14) and the induction hypothesis. This allows us to apply Lemma 7.1 again, and conclude that $|s_j|_X \le \tau(\zeta_j, \xi_j)$, where $\zeta_j = \kappa$ and $\xi_j = C_0 + 1 + \sum_{k=1}^{j-1} \tau_k$.

Thus, $|s_j|_X \le \max\{\kappa, \tau(\zeta_j, \xi_j)\} = \tau_j$, as required. Hence the claim has been proved by induction on j.

We are finally ready to prove the main statement of the lemma. Since $s_1, \dots, s_l$ is the list of all $\mathcal{H}$-components of α, we can combine the inequalities (7-14) and (7-16) to achieve

$$d_X(q_+, (p_i)_+) = |\alpha|_X \le \ell(\alpha) + \sum_{j=1}^{l} |s_j|_X \le C_0 + 1 + \sum_{j=1}^{l} \tau_j \le C_0 + 1 + \sum_{j=1}^{\lfloor C_0+1 \rfloor} \tau_j.$$

On the other hand, by the triangle inequality and (7-15), we have

$$d_X((p_i)_+, r_-) \le L + d_X(q_+, (p_i)_+) \le L + C_0 + 1 + \sum_{j=1}^{\lfloor C_0+1 \rfloor} \tau_j.$$

We have shown that the constant $C_1 = L + C_0 + 1 + \sum_{j=1}^{\lfloor C_0+1 \rfloor} \tau_j > 0$ is an upper bound for $d_X(q_+, (p_i)_+)$ and $d_X((p_i)_+, r_-)$; thus the lemma is proved. □

Definition 7.5 (consecutive, adjacent and multiple backtracking) Let $p = p_1 \cdots p_n$ be a broken line in $\Gamma(G, X \cup \mathcal{H})$. Suppose that for some i and j, with $1 \le i < j \le n$, and $\nu \in \mathcal{N}$ there exist pairwise connected H_ν-components $h_i, h_{i+1}, \dots, h_j$ of the paths $p_i, p_{i+1}, \dots, p_j$, respectively. Then we will say that p has *consecutive backtracking* along the components $h_i, \dots, h_j$ of $p_i, \dots, p_j$. Moreover, if

$j = i + 1$, we will call it an instance of *adjacent backtracking*, while if $j > i + 1$ will use the term *multiple backtracking*.

The next lemma shows that, among path representatives of minimal type, instances of adjacent backtracking where at least one of the components is sufficiently long with respect to the proper metric d_X must have initial and terminal vertices far apart in d_X.

Lemma 7.6 (adjacent backtracking is long) *For any $\zeta \geq 0$ there is $\Theta_0 = \Theta_0(\zeta) \in \mathbb{N}$ such that the following holds.*

Let $Q' \leqslant Q$ and $R' \leqslant R$ be subgroups satisfying (C1) *and let $p = p_1 \cdots p_n$ be a minimal type path representative for an element $g \in \langle Q', R' \rangle$. Suppose that for some $i \in \{1, \dots, n-1\}$ the paths p_i and p_{i+1} have connected $\mathcal{H}$-components q and r respectively, satisfying*

$$\max\{|q|_X, |r|_X\} \geq \Theta_0.$$

Then $d_X(q_-, r_+) \geq \zeta$.

Proof For any $\zeta \geq 0$ we can define $\Theta_0 = \lfloor \tau(\zeta, C_1) \rfloor + 1$, where C_1 is the constant from Lemma 7.3 and $\tau(\zeta, C_1)$ is provided by Lemma 7.1.

It follows that if $d_X(q_-, r_+) < \zeta$ then $|q|_X < \Theta_0$ and $|r|_X < \Theta_0$, which is the contrapositive of the required statement. □

8 Multiple backtracking in path representatives of minimal type

As before, we keep working under Convention 6.1. In this section we deal with multiple backtracking in path representatives of elements from $\langle Q', R' \rangle$. Proposition 8.4 below uses condition (C3) to show that any instance of multiple backtracking essentially takes place inside a parabolic subgroup. In order to achieve this we first prove two auxiliary statements.

Notation 8.1 Throughout this section $C_1 \geq 0$ will be the constant given by Lemma 7.3 and $\mathcal{P}_1$ will denote the finite collection of parabolic subgroups of G defined by

$$\mathcal{P}_1 = \{tH_\nu t^{-1} \mid \nu \in \mathcal{N}, |t|_X \leq C_1\}.$$

Consider the subset $O = \{o \in PS \mid P \in \mathcal{P}_1, \ |o|_X \leq 2C_1\}$ of G. Since $|O| < \infty$, we can choose and fix a finite subset $\Omega \subseteq S$ such that every element $o \in O$ can be written as $o = fh$, where $f \in P$, for some $P \in \mathcal{P}_1$, and $h \in \Omega$. We define a constant E by

$$E = \max\{|h|_X \mid h \in \Omega\} \geq 0. \tag{8-1}$$

Lemma 8.2 *There exists a constant $D \geq 0$ such that the following holds.*

Let $\nu \in \mathcal{N}$ and $b \in G$ be an element with $|b|_X \leq C_1$, so that $P = bH_\nu b^{-1} \in \mathcal{P}_1$, and let p be a geodesic path in $\Gamma(G, X \cup \mathcal{H})$ with $\tilde{p} \in Q \cup R$.

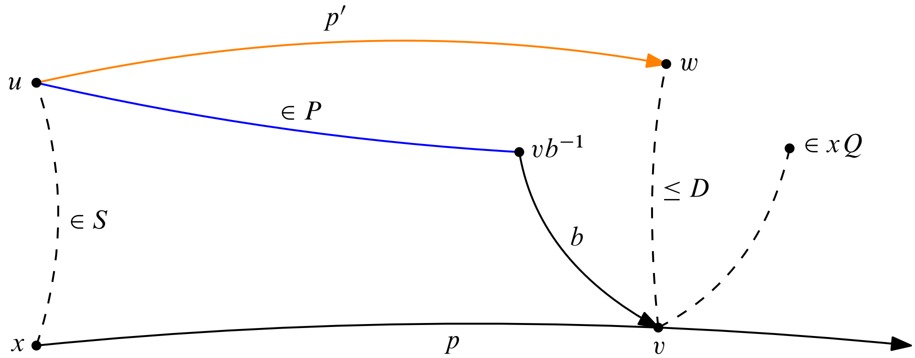


Figure 3: Illustration of Lemma 8.2.

Suppose there is a vertex v of p and an element $u \in P$ such that $v \in Pb = bH_\nu$ and $u^{-1}p_- \in S = Q \cap R$. Then there exists a geodesic path p' in $\Gamma(G, X \cup \mathcal{H})$ such that

- *$p'_- = u$ and $d_X(p'_+, v) \le D$;*
- *if $\tilde{p} \in Q$ then $\tilde{p}' \in Q \cap P$, otherwise $\tilde{p}' \in R \cap P$.*

Proof Let $K = \max\{C_1, \varepsilon\} \ge 0$, where ε is the quasiconvexity constant of Q and R, and let

$$D = \max\{K'(Q, P, K), K'(R, P, K) \mid P \in \mathcal{P}_1\}, \tag{8-2}$$

where $K'(Q, P, K)$ and $K'(R, P, K)$ are obtained from Lemma 4.1.

Denote $x = p_- \in G$ and assume, without loss of generality, that $\tilde{p} \in Q$ (the case $\tilde{p} \in R$ can be treated similarly). By the quasiconvexity of Q, we have that $d_X(v, xQ) \le \varepsilon$. Moreover, $xQ = uQ$ as $u^{-1}x \in S \subseteq Q$.

By the assumptions, $vb^{-1} \in P$; hence $d_X(v, P) \le |b|_X \le C_1$. Since $uP = P$ we see that

$$v \in N_X(uQ, \varepsilon) \cap N_X(uP, C_1).$$

Applying Lemma 4.1, we find $w \in u(Q \cap P)$ such that $d_X(v, w) \le D$ (see Figure 3).

Let p' be any geodesic in $\Gamma(G, X \cup \mathcal{H})$ starting at u and ending at w. It is easy to see that p' satisfies all of the required properties, so the lemma is proved. □

The next lemma describes how condition (C3) is used in this paper.

Lemma 8.3 *Assume that subgroups $Q' \leqslant Q$ and $R' \leqslant R$ satisfy conditions* (C1) *and* (C3) *with constant C and family $\mathcal{P}$ such that $C \ge 2C_1 + 1$ and $\mathcal{P}_1 \subseteq \mathcal{P}$. Let $P = bH_\nu b^{-1} \in \mathcal{P}_1$, for some $\nu \in \mathcal{N}$ and $b \in G$, with $|b|_X \le C_1$, and let p be a path in $\Gamma(G, X \cup \mathcal{H})$ with $\tilde{p} \in Q' \cup R'$.*

Suppose that there is a vertex v of p and an element $u \in P$ satisfying $u^{-1}p_- \in S$, $v \in Pb$, and $d_X(v, p_+) \le C_1$. Then there exists a geodesic path p' such that $(p')_- = u$, $\tilde{p}' \in P$, $(p')_+^{-1}p_+ \in S$ and

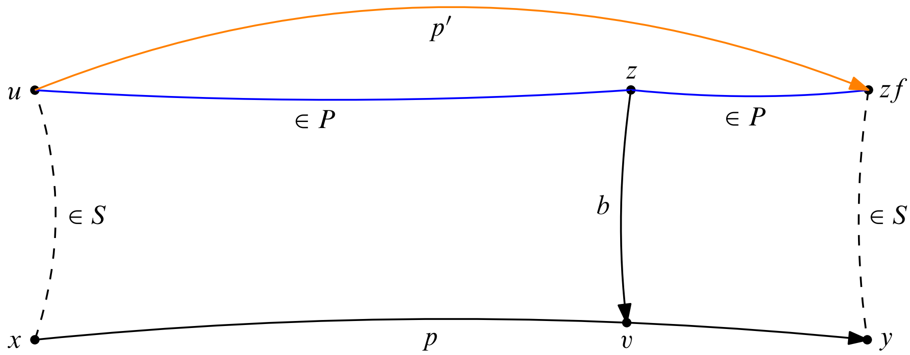


Figure 4: Illustration of Lemma 8.3.

$d_X((p')_+, p_+) \le E$, where E is the constant from (8-1). *In particular, if $\tilde{p} \in Q'$ (respectively, $\tilde{p} \in R'$) then $\tilde{p}' \in Q' \cap P$ (respectively, $\tilde{p}' \in R' \cap P$).*

Proof Denote $x = p_-$, $y = p_+$ and $z = vb^{-1} \in P$ (see Figure 4). Then $u^{-1}z \in P$ and $x^{-1}y = \tilde{p} \in Q' \cup R'$. Since $u^{-1}x \in S = Q' \cap R'$, we obtain

$$u^{-1}y = (u^{-1}x)(x^{-1}y) \in Q' \cup R',$$

whence $z^{-1}y = (z^{-1}u)(u^{-1}y) \in P(Q' \cup R')$. Now, observe that

$$|z^{-1}y|_X = d_X(z, y) \le d_X(z, v) + d_X(v, y) \le |b|_X + C_1 \le 2C_1 < C.$$

Condition (C3) now implies that $z^{-1}y \in PS$. That is, $z^{-1}y = fh$ for some $f \in P$ and $h \in \Omega$, where Ω is the finite subset of S defined above the statement of the lemma. Let p' be a geodesic path starting at u and ending at $zf \in P$. Then $\tilde{p}' = u^{-1}zf \in P$,

$$(p')_+^{-1} p_+ = f^{-1}z^{-1}y = h \in S \quad \text{and} \quad d_X((p')_+, p_+) = |h|_X \le E.$$

The last statement of the lemma follows from (C1) and the observation that

$$\tilde{p}' = u^{-1}(p')_+ = u^{-1}p_-\tilde{p}(p_+)^{-1}(p')_+ \in S\tilde{p}S. \qquad \square$$

Proposition 8.4 *Let $D \ge 0$ is the constant provided by Lemma 8.2, and let E be given by* (8-1). *Suppose that $Q' \leqslant Q$ and $R' \leqslant R$ are subgroups satisfying* (C1) *and* (C3), *with constant $C \ge 2C_1 + 1$ and family $\mathcal{P} \supseteq \mathcal{P}_1$.*

Let $p = p_1 \cdots p_n$ be a path representative for an element $g \in \langle Q', R' \rangle$ with minimal type. If p has consecutive backtracking along $\mathcal{H}$-components $h_i, \dots, h_j$ of the subpaths $p_i, \dots, p_j$ respectively, then there is a subgroup $P \in \mathcal{P}_1$ and a path $p' = p'_i \cdots p'_j$ satisfying the following properties:

(i) *p'_k is geodesic with $\tilde{p}'_k \in P$ for all $k = i, \dots, j$;*

(ii) *$(p'_i)_+ = (p_i)_+$, $(p'_k)_+^{-1}(p_k)_+ \in S$ and $d_X((p'_k)_+, (p_k)_+) \le E$, for all $k = i+1, \dots, j-1$;*

(iii) *$d_X(p'_-, (h_i)_-) \le D$ and $d_X(p'_+, (h_j)_+) \le D$;*

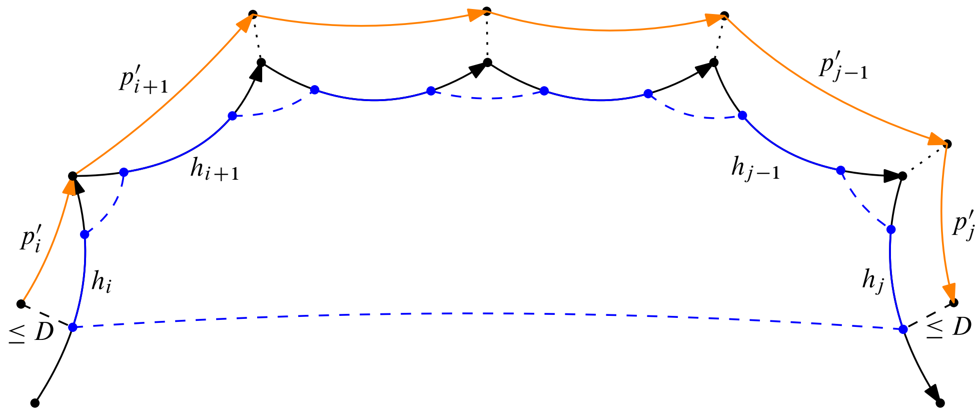


Figure 5: The new path p' constructed in Proposition 8.4. The dotted lines between p and p' are paths whose labels represent elements of S.

(iv) *$\tilde{p}'_i \in Q \cap P$ if $\tilde{p}_i \in Q'$, and $\tilde{p}'_i \in R \cap P$ if $\tilde{p}_i \in R'$; similarly, $\tilde{p}'_j \in Q \cap P$ if $\tilde{p}_j \in Q'$, and $\tilde{p}'_j \in R \cap P$ if $\tilde{p}_j \in R'$;*

(v) *for each $k \in \{i+1, \dots, j-1\}$, $\mathrm{Lab}(p'_k)$ either represents an element of $Q' \cap P$ or an element of $R' \cap P$.*

Proof Figure 5 below is a sketch of the path p' above the subpath $p_i p_{i+1} \cdots p_{j-1} p_j$ of p.

Note that claim (v) follows from claim (ii) and condition (C1), so we only need to establish claims (i)–(iv).

By the assumptions, there is $\nu \in \mathcal{N}$ such that for each $k \in \{i, \dots, j\}$, the path p_k is a concatenation $p_k = a_k h_k b_k$, where h_k is an H_ν-component of p_k and a_k, b_k are subpaths of p_k.

According to Lemma 7.3, we have

$$|b_k|_X \leq C_1 \quad \text{for } k = i, \dots, j-1. \tag{8-3}$$

After translating everything by $(p_i)_+^{-1}$ we can assume that $(p_i)_+ = 1$. From here on, we let $b = \tilde{b}_i^{-1} \in G$ and $P = bH_\nu b^{-1}$. As noted in (8-3), $|b|_X = |b_i|_X \leq C_1$, so $P \in \mathcal{P}_1$.

Since the components h_i and h_k are connected, for every $k = i+1, \dots, j$, the elements $(h_i)_+ = (b_i)_- = b$ and $(h_k)_+$ all belong to the same left coset $bH_\nu = Pb$; thus

$$(h_k)_+ \in Pb \quad \text{for all } k = i+1, \dots, j. \tag{8-4}$$

The rest of the argument will be divided into three steps.

Step 1 Construction of the path p'_i.

Set $u_i = (p_i)_+ = 1$ and $v_i = (h_i)_-$. Then $v_i = \tilde{b}_i^{-1} \tilde{h}_i^{-1} \in bH_\nu = Pb$, so the path p_i^{-1}, its vertex v_i and the element $u_i = 1 \in P$ satisfy the assumptions of Lemma 8.2. Therefore there exists a path q with $q_- = u_i$, $d_X(q_+, v) \leq D$ and such that $\tilde{q} \in Q \cap P$ if $\tilde{p}_i \in Q$ and $\tilde{q} \in R \cap P$ if $\tilde{p}_i \in R$.

It is easy to check that the path $p'_i = q^{-1}$ enjoys the required properties.

Step 2 Construction of the paths p'_k, for $k = i+1, \dots, j-1$.

We will define the paths p'_k by induction on k. For $k = i+1$ we consider the path p_{i+1}, its vertex $v_{i+1} = (h_{i+1})_+$ and the element $u_i = 1 = (p_{i+1})_-$. Since $v_{i+1} \in Pb$ by (8-4) and

$$d_X(v_{i+1}, (p_{i+1})_+) = |b_{i+1}|_X \le C_1$$

by (8-3), we can apply Lemma 8.3 to find a geodesic path p'_{i+1} starting at u_i and satisfying the required conditions.

Now suppose the required paths $p'_{i+1}, \dots, p'_m$ have already been constructed for some $m \in \{i+1, \dots, j-2\}$. To construct the path p'_{m+1}, let v_{m+1} be the vertex $(h_{m+1})_+$ of p_{m+1} and set $u_m = (p'_m)_+$. Then $u_m \in P$ and $u_m^{-1}(p_{m+1})_- = (p'_m)_+^{-1}(p_m)_+ \in S$ by the induction hypothesis. In view of (8-4) and (8-3), $v_{m+1} \in Pb$ and $d_X(v_{m+1}, (p_{m+1})_+) \le C_1$; therefore we can find a geodesic path p'_{m+1} with the desired properties by using Lemma 8.3.

Thus we have described an inductive procedure for constructing the paths p'_k, for $k = i+1, \dots, j-1$.

Step 3 Construction of the path p'_j.

This step is similar to Step 1: the path p'_j will start at $u_{j-1} = (p'_{j-1})_+ \in P$ and can be constructed by applying Lemma 8.2 to the path p_j and the elements $v_j = (h_j)_+ \in Pb$, $u_{j-1} \in P$.

We have thus constructed a sequence of geodesic paths $p'_i, \dots, p'_j$ whose concatenation p' satisfies all the properties from the proposition. □

We will now prove the main result of this section, which states that the initial and terminal vertices of an instance of multiple backtracking in a minimal type path representative must lie far apart in the proper metric d_X, provided $Q' \leqslant Q$ and $R' \leqslant R$ satisfy (C1)–(C5) with sufficiently large constants.

Proposition 8.5 (multiple backtracking is long) *For any $\zeta \ge 0$ there is a constant $C_2 = C_2(\zeta) \ge 0$ such that if $Q' \leqslant Q$ and $R' \leqslant R$ are subgroups satisfying conditions* (C1)–(C5) *with constants $B \ge C_2$ and $C \ge C_2$ and a family $\mathcal{P} \supseteq \mathcal{P}_1$, then the following is true.*

Let $p = p_1 \cdots p_n$ be a minimal type path representative for an element $g \in \langle Q', R' \rangle$. If p has multiple backtracking along $\mathcal{H}$-components $h_i, \dots, h_j$ of $p_i, \dots, p_j$ then $d_X((h_i)_-, (h_j)_+) \ge \zeta$.

Proof Let $\zeta \ge 0$ and define $C_2(\zeta) = \max\{2C_1, \zeta + 2D\} + 1$, where $D \ge 0$ is the constant obtained from Lemma 8.2.

In view of the assumptions we can apply Proposition 8.4 to find a path $p' = p'_i \cdots p'_j$ and $P \in \mathcal{P}_1$ satisfying properties (i)–(v) from its statement. Let α be a geodesic with $\alpha_- = (p'_j)_-$ and $\alpha_+ = (p_j)_-$, and let $\beta = p'_{i+1} \cdots p'_{j-1}$. We will denote $x_k = \tilde{p}_k$ and $x'_k = \tilde{p}'_k$, for each $k \in \{i, \dots, j\}$, and $z = \tilde{\alpha}$.

Condition (C1), together with claim (ii) of Proposition 8.4, tell us that $z \in S = Q' \cap R'$, and claim (v) yields that

$$\tilde{\beta} = x'_{i+1} \cdots x'_{j-1} \in \langle Q'_P, R'_P \rangle \tag{8-5}$$

(as before, for a subgroup $H \leqslant G$ we denote by $H_P \leqslant G$ the intersection $H \cap P$).

Now suppose, for a contradiction, that $d_X((h_i)_-, (h_j)_+) < \zeta$. Then

$$|p'|_X = d_X(p'_-, p'_+) < \zeta + 2D < C_2 \le \min\{B, C\}, \tag{8-6}$$

by claim (iii) of Proposition 8.4. There are four cases to consider depending on whether $\tilde{p}_i$ and $\tilde{p}_j$ are elements of Q' or R'.

Case 1 $x_i = \tilde{p}_i \in Q'$ and $x_j = \tilde{p}_j \in Q'$.

Then, by claim (iv) of Proposition 8.4, both x'_i and x'_j are elements of Q_P. It follows that

$$\tilde{p}' \in Q_P \langle Q'_P, R'_P \rangle Q_P \subseteq Q \langle Q', R' \rangle Q.$$

By (8-6) and (C2), there is $q \in Q$ such that $\tilde{p}' = q$. Therefore

$$\tilde{\beta} = x_i'^{-1} \tilde{p}' x_j'^{-1} = x_i'^{-1} q x_j'^{-1} \in Q. \tag{8-7}$$

Combining (8-7) with (8-5) and using condition (C4), we get

$$\tilde{\beta} \in Q \cap \langle Q'_P, R'_P \rangle = Q_P \cap \langle Q'_P, R'_P \rangle = Q'_P.$$

Let γ be any geodesic path in $\Gamma(G, X \cup \mathcal{H})$ starting at $(p_i)_-$ and ending at $(p_j)_+$. Then γ shares the same endpoints with the path $p_i \beta \alpha p_j$; therefore their labels represent the same element of G,

$$\tilde{\gamma} = x_i \tilde{\beta} z x_j \in Q' Q'_P S Q' = Q'.$$

Thus we can use γ to obtain another path representative for g through $p_1 \cdots p_{i-1} \gamma p_{j+1} \cdots p_n$, which consists of strictly fewer geodesic subpaths than $p = p_1 \cdots p_n$. This contradicts the minimality of the type of p, so Case 1 has been considered.

Case 2 Both $\tilde{p}_i$ and $\tilde{p}_j$ are elements of R'.

This case can be dealt with identically to Case 1.

Case 3 $x_i = \tilde{p}_i \in Q'$ and $x_j = \tilde{p}_j \in R'$.

Then $x'_i \in Q_P$ and $x'_j \in R_P$ by claim (iv) of Proposition 8.4. Hence $\mathrm{Lab}(p')$ represents an element of $x'_i \langle Q'_P, R'_P \rangle R_P$ with $x'_i \in Q_P$. In view of (8-6), we can use condition (C5) to deduce that $\tilde{p}' \in x'_i Q'_P R_P$. It follows that

$$\tilde{\beta} = (x'_i)^{-1} \tilde{p}' (x'_j)^{-1} \in Q'_P R_P,$$

so there exist $q \in Q'_P$ and $r \in R_P$ such that $\tilde{\beta} = qr$. Combining this with (8-5) we conclude that $r = q^{-1}\tilde{\beta} \in R_P \cap \langle Q'_P, R'_P \rangle$, so $r \in R'_P$ by condition (C4), whence

$$\tilde{\beta} = qr \in Q'_P R'_P. \tag{8-8}$$

Observe that the paths $\gamma = p_i \cdots p_j$ and $p_i \beta \alpha p_j$ have the same endpoints; hence their labels represent the same element of G,

$$\tilde{\gamma} = x_i \tilde{\beta} z x_j \in Q' Q'_P R'_P S R' \subseteq Q' R'.$$

Therefore there are elements $q_1 \in Q'$ and $r_1 \in R'$ such that $\tilde{\gamma} = q_1 r_1$.

Let γ_1 be a geodesic path in $\Gamma(G, X \cup \mathcal{H})$ starting at $\gamma_- = (p_i)_-$ and ending at $\gamma_- q_1$ and let γ_2 be a geodesic path starting at $(\gamma_1)_+$ and ending at $(\gamma_1)_+ r_1 = \gamma_+ = (p_j)_+$. Since $\tilde{\gamma}_1 = q_1 \in Q'$ and $\tilde{\gamma}_2 = r_1 \in R'$ the path $p_1 \cdots p_{i-1} \gamma_1 \gamma_2 p_{j+1} \cdots p_n$ is a path representative of g. Moreover, it consists of fewer than n geodesic segments because $j > i + 1$ (by the definition of multiple backtracking), contradicting the minimality of the type of p. This contradiction shows that Case 3 is impossible.

Case 4 $x_i = \tilde{p}_i \in R'$ and $x_j = \tilde{p}_j \in Q'$.

Then $x'_i \in R_P$ while $x'_j \in Q_P$, which implies that $\tilde{p}' \in R_P \langle Q'_P, R'_P \rangle x'_j$, hence $\tilde{p}'^{-1} \in (x'_j)^{-1} \langle Q'_P, R'_P \rangle R_P$.

By (8-6), we can use (C5) to conclude that $\tilde{p}'^{-1} \in (x'_j)^{-1} Q'_P R_P$, thus $\tilde{p}' \in R_P Q'_P x'_j$. The rest of the argument proceeds similarly to the previous case, leading to a contradiction with the minimality of the type of p. Hence Case 4 is also impossible.

We have arrived at a contradiction in each of the four cases, so $d_X((h_i)_-, (h_j)_+) \geq \zeta$, as required. □

9 Constructing quasigeodesics from broken lines

In this section we detail a procedure that takes as input a broken line and a natural number, and outputs another broken line together with some additional vertex data. We show that if a broken line satisfies certain metric conditions, then the new path constructed through this procedure is uniformly quasigeodesic.

We assume that G is a group generated by a finite set X and hyperbolic relative to a finite family of subgroups $\{H_\nu \mid \nu \in \mathcal{N}\}$. As usual we set $\mathcal{H} = \bigsqcup_{\nu \in \mathcal{N}} (H_\nu \setminus \{1\})$, and by Lemma 5.4 we know that the Cayley graph $\Gamma(G, X \cup \mathcal{H})$ is δ-hyperbolic, for some $\delta \geq 0$.

The outline of the construction is as follows. We begin with a broken line $p = p_1 \cdots p_n$ in $\Gamma(G, X \cup \mathcal{H})$. Starting from the initial vertex p_-, we note in sequence (along the vertices of p) the vertices marking the start and end of maximal instances of consecutive backtracking in p involving sufficiently long $\mathcal{H}$-components. Once we have done this, we construct the new path by connecting (in the same sequence) the marked vertices with geodesics.

Procedure 9.1 (Θ-shortcutting) Fix a natural number $\Theta \in \mathbb{N}$ and let $p = p_1 \cdots p_n$ be a broken line in $\Gamma(G, X \cup \mathcal{H})$. Let $v_0, \dots, v_d$ be the enumeration of all vertices of p in the order they occur along the path (possibly with repetition), so that $v_0 = p_-$, $v_d = p_+$ and $d = \ell(p)$.

We construct a broken line $\Sigma(p, \Theta)$, called the Θ-*shortcutting* of p, which comes with a finite set $V(p, \Theta) \subset \{0, \dots, d\} \times \{0, \dots, d\}$ corresponding to indices of vertices of p that we shortcut along.

In the algorithm below we will refer to numbers $s, t, N \in \{0, \dots, d\}$ and a subset $V \subseteq \{0, \dots, d\} \times \{0, \dots, d\}$. To avoid excessive indexing these will change value throughout the procedure. The parameters s and t will indicate the starting and terminal vertices of subpaths of p in which all $\mathcal{H}$-components have lengths less than Θ. The parameter N will keep track of how far along the path p we have proceeded. The set V will collect all pairs of indices (s, t) obtained during the procedure. We initially take $s = 0$, $N = 0$ and $V = \varnothing$.

Step 1 If there are no edges of p between v_N and v_d that are labelled by elements of $\mathcal{H}$, then add the pair (s, d) to the set V and skip ahead to Step 4. Otherwise, continue to Step 2.

Step 2 Let $t \in \{0, \dots, d\}$ be the least natural number with $t \geq N$ for which the edge of p with endpoints v_t and v_{t+1} is an $\mathcal{H}$-component h_i of a geodesic segment p_i of p, for some $i \in \{1, \dots, n\}$.

If $i = n$ or if h_i is not connected to a component of p_{i+1} then set $j = i$. Otherwise, let $j \in \{i+1, \dots, n\}$ be the maximal integer such that p has consecutive backtracking along $\mathcal{H}$-components $h_i, \dots, h_j$ of segments $p_i, \dots, p_j$. Proceed to Step 3.

Step 3 If

$$\max\{|h_k|_X \mid k = i, \dots, j\} \geq \Theta,$$

then add the pair (s, t) to the set V and redefine $s = N$ in $\{1, \dots, d\}$ to be the index of the vertex $(h_j)_+$ in the above enumeration $v_0, \dots, v_d$ of the vertices of p. Otherwise let N be the index of $(h_i)_+$, and leave s and V unchanged.

Return to Step 1 with the new values of s, N and V.

Step 4 Set $V(p, \Theta) = V$. The above constructions gives a natural ordering of $V(p, \Theta)$,

$$V(p, \Theta) = \{(s_0, t_0), \dots, (s_m, t_m)\},$$

where $s_k \leq t_k < s_{k+1}$, for all $k = 0, \dots, m-1$. Note that $s_0 = 0$ and $t_m = d$. Proceed to Step 5.

Step 5 For each $k = 0, \dots, m$, let f_k be a geodesic segment (possibly trivial) connecting v_{s_k} with v_{t_k}. Note that when $k < m$, v_{t_k} and $v_{s_{k+1}}$ are in the same left coset of H_ν, for some $\nu \in \mathcal{N}$. If $v_{t_k} = v_{s_{k+1}}$ then let e_{k+1} be the trivial path at v_{t_k}, otherwise let e_{k+1} be an edge of $\Gamma(G, X \cup \mathcal{H})$ starting at v_{t_k}, ending at $v_{s_{k+1}}$ and labelled by an element of $H_\nu \setminus \{1\}$.

We define the broken line $\Sigma(p, \Theta)$ to be the concatenation $f_0 e_1 f_1 e_2 \cdots f_{m-1} e_m f_m$.

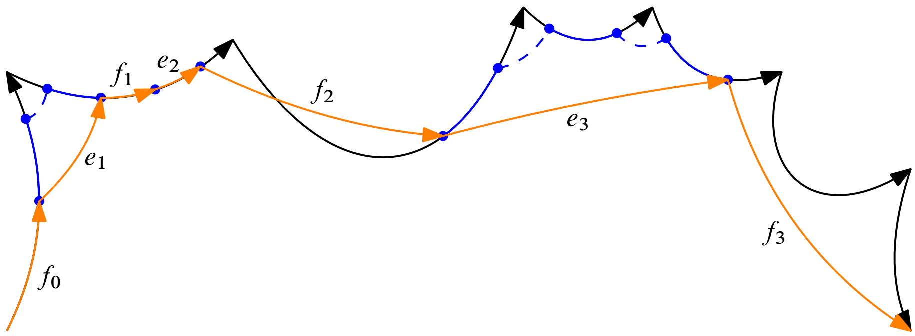


Figure 6: An example of a shortcutting of a path p in $\Gamma(G, X \cup \mathcal{H})$. The path p contains long $\mathcal{H}$-components, some of which are involved in instances of consecutive backtracking, as indicated by the dashed lines. The path $\Sigma(p, \Theta) = f_0 e_1 f_1 e_2 f_2 e_3 f_3$ is drawn on top of p.

Remark 9.2 Let us collect some observations about Procedure 9.1.

(a) Since p has only finitely many vertices and N increases at each iteration of Step 3 above, the procedure will always terminate after finitely many steps.

(b) The newly constructed broken line $\Sigma(p, \Theta)$ has the same endpoints as p, and each node of $\Sigma(p, \Theta)$ is a vertex of p.

(c) By construction, for any $k \in \{0, \dots, m\}$ the subpath of p between v_{s_k} and v_{t_k} contains no edge labelled by an element $h \in \mathcal{H}$ satisfying $|h|_X \geq \Theta$.

Figure 6 sketches an example of the output of Procedure 9.1.

In the next definition we describe paths that will serve as input for the above procedure.

Definition 9.3 (tamable broken line) Let $p = p_1 \cdots p_n$ be a broken line in $\Gamma(G, X \cup \mathcal{H})$, and let $B, C, \zeta \geq 0$ and $\Theta \in \mathbb{N}$. We say that p is (B, C, ζ, Θ)-*tamable* if all of the following conditions hold:

(i) $|p_i|_X \geq B$, for $i = 2, \dots, n-1$;

(ii) $\langle (p_i)_-, (p_{i+1})_+ \rangle^{\mathrm{rel}}_{(p_i)_+} \leq C$, for each $i = 1, \dots, n-1$;

(iii) whenever p has consecutive backtracking along $\mathcal{H}$-components $h_i, \dots, h_j$, of segments $p_i, \dots, p_j$, such that
$$\max\{|h_k|_X \mid k = i, \dots, j\} \geq \Theta,$$
it must be that $d_X((h_i)_-, (h_j)_+) \geq \zeta$.

The remainder of this section is devoted to showing the following result about quasigeodesicity of shortcuttings for tamable paths with appropriate constants.

Proposition 9.4 *Given arbitrary $c_0 \geq 14\delta$ and $\eta \geq 0$ there are constants $\lambda = \lambda(c_0) \geq 1$, $c = c(c_0) \geq 0$ and $\zeta = \zeta(\eta, c_0) \geq 1$ such that for any natural number $\Theta \geq \zeta$ there is $B_0 = B_0(\Theta, c_0) \geq 0$ satisfying the following.*

Let $p = p_1 \cdots p_n$ be a $(B_0, c_0, \zeta, \Theta)$-tamable broken line in $\Gamma(G, X \cup \mathcal{H})$ and let $\Sigma(p, \Theta)$ be the Θ-shortcutting, obtained by applying Procedure 9.1 to p, $\Sigma(p, \Theta) = f_0 e_1 f_1 \cdots f_{m-1} e_m f_m$. Then e_k is non-trivial, for each $k = 1, \ldots, m$, and $\Sigma(p, \Theta)$ is (λ, c)-quasigeodesic without backtracking.

Moreover, for any $k \in \{1, \ldots, m\}$, if we denote by e'_k the $\mathcal{H}$-component of $\Sigma(p, \Theta)$ containing e_k, then $|e'_k|_X \geq \eta$.

The idea of the proof will be to show that under the above assumptions the broken line $\Sigma(p, \Theta)$ satisfies the hypotheses of Proposition 5.19.

Notation 9.5 For the remainder of this section we fix arbitrary constants $c_0 \geq 14\delta$ and $\eta \geq 0$. We let $\rho = \kappa(4, c_3, 0)$, where $c_3 = c_3(c_0) \geq 0$ is the constant from Lemma 4.11 and $\kappa(4, c_3, 0)$ is the constant obtained by applying Proposition 5.17 to $(4, c_3)$-quasigeodesics. Let $\zeta_1 > 0$, $\lambda \geq 1$ and $c \geq 0$ be the constants given by Proposition 5.19, applied with constant ρ. Note that the constants λ and c only depend on c_0 and do not depend on η.

We now define the constant ζ by

$$\zeta = \max\{\zeta_1, \eta\} + 2\rho + 1. \tag{9-1}$$

Finally we take any natural number $\Theta \geq \zeta$ and

$$B_0 = \max\{(12c_0 + 12\delta + 1)\Theta, (4 + c_3)\Theta + 1\}. \tag{9-2}$$

The proof of Proposition 9.4 will consist of the following four lemmas. Throughout these lemmas we use the constants defined above and assume that $p = p_1 \cdots p_n$ is a $(B_0, c_0, \zeta, \Theta)$-tamable broken line in $\Gamma(G, X \cup \mathcal{H})$. As before, we write $v_0, \ldots, v_d$ for the set of vertices of p in the order of their appearance. We let $\Sigma(p, \Theta) = f_0 e_1 f_1 \cdots f_{m-1} e_m f_m$ be the Θ-shortcutting and $V(p, \Theta) = \{(s_0, t_0), \ldots, (s_m, t_m)\}$ be the set obtained by applying Procedure 9.1 to p.

Lemma 9.6 *For each $k = 1, \ldots, m$, we have $|e_k|_X \geq \zeta > 0$.*

Proof By the construction in Procedure 9.1, there are pairwise connected $\mathcal{H}$-components $h_1, \ldots, h_j$ of consecutive segments of p, such that $j \geq 1$, $(h_1)_- = (e_k)_-$, $(h_s)_+ = (e_k)_+$ and

$$\max\{|h_l|_X \mid l = 1, \ldots, j\} \geq \Theta.$$

If $j = 1$ we see that $|e_k|_X = |h_1|_X \geq \Theta \geq \zeta$, and if $j > 1$ then we know that $|e_k|_X \geq \zeta$ by property (iii) from Definition 9.3. □

Lemma 9.7 *The subpaths of p between v_{s_k} and v_{t_k}, for $k = 0, \ldots, m$, are $(4, c_3)$-quasigeodesic.*

Proof We write $c_1 = c_1(c_0) = 12c_0 + 12\delta + 1$, as in Lemma 4.11.

Choose any $k \in \{0, \ldots, m\}$ and denote by p' be the subpath of p starting at v_{s_k} and terminating at v_{t_k}. If v_{s_k} and v_{t_k} are both vertices of p_i, for some $i \in \{1, \ldots, n\}$, then p' is geodesic and we are done.

Otherwise $p' = p'_i p_{i+1} \cdots p_{j-1} p'_j$, for some $i, j \in \{1, \dots, n\}$, with $i < j$, where p'_i is a terminal segment of p_i and p'_j is an initial segment of p_j.

By Remark 9.2(c), the paths $p_{i+1}, \dots, p_{j-1}$ contain no $\mathcal{H}$-components h with $|h|_X \geq \Theta$. Since p is $(B_0, c_0, \zeta, \Theta)$-tamable, $|p_l|_X \geq B_0$ for each $l = i+1, \dots, j-1$ by condition (i). Thus we can combine Lemma 5.10 with (9-2) to obtain

$$d_{X \cup \mathcal{H}}((p_l)_-, (p_l)_+) = \ell(p_l) \geq \frac{1}{\Theta}|p_l|_X \geq \frac{B_0}{\Theta} \geq c_1 \quad \text{for each } l \in \{i+1, \dots, j-1\}.$$

Again, from the assumption that p is $(B_0, c_0, \zeta, \Theta)$-tamable, we have that

$$\langle (p_l)_-, (p_{l+1})_+ \rangle^{\mathrm{rel}}_{(p_l)_+} \leq c_0 \quad \text{for all } l = i, \dots, j-1,$$

using condition (ii). In view of Remark 4.6,

$$\langle (p'_i)_-, (p_{i+1})_+ \rangle^{\mathrm{rel}}_{(p'_i)_+} \leq c_0 \quad \text{and} \quad \langle (p_{j-1})_-, (p'_j)_+ \rangle^{\mathrm{rel}}_{(p_{j-1})_+} \leq c_0.$$

Therefore we can use Lemma 4.11 to conclude that p' is $(4, c_3)$-quasigeodesic, as required. □

Lemma 9.8 *If $k \in \{0, \dots, m-1\}$ and h is an $\mathcal{H}$-component of f_k or f_{k+1} that is connected to e_{k+1}, then $|h|_X \leq \rho$.*

Proof Arguing by contradiction, suppose that h is an $\mathcal{H}$-component of f_k connected to e_{k+1} and satisfying $|h|_X > \rho$ (the other case when h is an $\mathcal{H}$-component of f_{k+1} is similar). Remark 5.9 tells us that h is a single edge of f_k. Moreover, since h and e_{k+1} are connected and $(f_k)_+ = (e_{k+1})_-$, we have $d_{X \cup \mathcal{H}}(h_-, (f_k)_+) \leq 1$. The geodesicity of f_k in $\Gamma(G, X \cup \mathcal{H})$ now implies that h must in fact be the last edge of f_k, so that $h_+ = (f_k)_+ = v_{t_k}$.

Let $p' = p'_i p_{i+1} \cdots p_{j-1} p'_j$ be the subpath of p with $p'_- = v_{s_k}$ and $p'_+ = v_{t_k}$, where p'_i and p'_j are non-trivial subpaths of p_i and p_j respectively. By Lemma 9.7, p' is $(4, c_3)$-quasigeodesic.

Since $|h|_X > \rho = \kappa(4, c_3, 0)$ we may apply Proposition 5.17 to find that h is connected to an $\mathcal{H}$-component of p' (which may consist of multiple edges, each of which is an $\mathcal{H}$-component of a segment of p). We write h' for the final edge of this $\mathcal{H}$-component and denote by u the edge of p with endpoints v_{t_k} and v_{t_k+1} (see Figure 7). Procedure 9.1 and the assumption that h is connected to e_{k+1} imply that u is an $\mathcal{H}$-component of a segment of p and h' and u are connected as $\mathcal{H}$-subpaths of p.

Suppose, first, that p'_j is a proper subpath of p_j, so that u belongs to the segment p_j, as shown on Figure 7. Then there are the following possibilities.

Case 1 h' is an edge of p_j.

In this case h' and u are connected distinct $\mathcal{H}$-subpaths of p_j, which is a geodesic. This contradicts the observation of Remark 5.9, that geodesics are without backtracking and $\mathcal{H}$-components of geodesics are single edges.

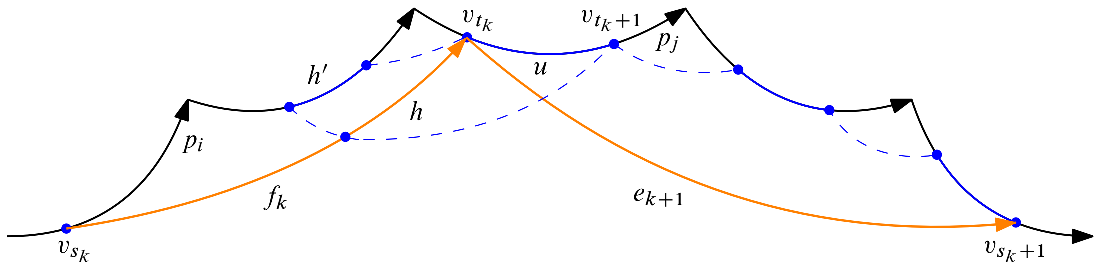


Figure 7: Illustration of Lemma 9.8.

Case 2 h' is an $\mathcal{H}$-component of p_{j-1}.

Let $t \in \{0, \dots, d\}$ be such that $v_t = h'_-$, and note that

$$s_k \le t < t_k. \tag{9-3}$$

By the construction from Procedure 9.1, there are pairwise connected $\mathcal{H}$-components $h_j, \dots, h_{j+l}$, of segments $p_j, \dots, p_{j+l}$, with $(e_{k+1})_- = (h_j)_- = v_{t_k}$ and $(e_{k+1})_+ = (h_{j+l})_+ = v_{s_{k+1}}$, such that

$$\max\{|h_j|_X, \dots, |h_{j+l}|_X\} \ge \Theta$$

and $l \in \{0, \dots, n-j\}$ is chosen to be maximal with this property. Then the components $h', h_j, \dots, h_{j+l}$ constitute a larger instance of consecutive backtracking, starting at $h'_- = v_t$, with

$$\max\{|h'|_X, |h_j|_X, \dots, |h_{j+l}|_X\} \ge \Theta.$$

In view of (9-3), this contradicts the choice of t_k and the inclusion of (s_k, t_k) in the set $V(p, \Theta)$ at Steps 2 and 3 of Procedure 9.1.

Case 3 h' is an $\mathcal{H}$-component of one of the paths $p'_i, p_{i+1}, \dots, p_{j-2}$.

Then the subpath q of p' from h'_+ to $p'_+ = v_{t_k}$ contains all of p_{j-1}. By Remark 9.2(c), p_{j-1} contains no $\mathcal{H}$-components q satisfying $|q|_X \ge \Theta$. Therefore, in view of Lemma 5.10 and the assumption that p is $(B_0, c_0, \zeta, \Theta)$-tamable, we can deduce that $\Theta\ell(p_{j-1}) \ge |p_{j-1}|_X \ge B_0$. Combining this with the $(4, c_3)$-quasigeodesicity of p', we obtain

$$d_{X\cup\mathcal{H}}(h'_+, p'_+) \ge \frac{1}{4}(\ell(q) - c_3) \ge \frac{1}{4}(\ell(p_{j-1}) - c_3) \ge \frac{B_0}{4\Theta} - \frac{c_3}{4} > 1,$$

where the last inequality follows from (9-2). On the other hand, the fact that h' and h are connected gives $d_{X\cup\mathcal{H}}(h'_+, p'_+) = d_{X\cup\mathcal{H}}(h'_+, h_+) \le 1$, contradicting the above.

In each case we arrive at a contradiction, so it is impossible that $|h|_X > \rho$ if p'_j is a proper subpath of p_j. If p'_j is instead the whole subpath p_j, we may carry out a similar analysis. In this situation it must be that u is an $\mathcal{H}$-component of the segment p_{j+1}. We now have only two relevant cases to consider: h' is an $\mathcal{H}$-component of p_j or h' is an $\mathcal{H}$-component of one of the paths $p'_i, p_{i+1}, \dots, p_{j-1}$. Both of them will lead to contradictions similarly to Cases 2 and 3 above.

Therefore it must be that $|h|_X \le \rho$, as required. □

Lemma 9.9 *For each $k \in \{1, \dots, m-1\}$, the $\mathcal{H}$-subpaths e_k and e_{k+1} of $\Sigma(p, \Theta)$ are not connected.*

Proof Suppose that e_k is connected to e_{k+1} for some $k \in \{1, \dots, m-1\}$. As before, according to Procedure 9.1, there exist two sets of pairwise connected $\mathcal{H}$-components of consecutive segments of p, $h_1, \dots, h_i$ and $q_1, \dots, q_j$, such that $(h_1)_- = (e_k)_-$, $(h_i)_+ = (e_k)_+$, $(q_1)_- = (e_{k+1})_-$, $(q_j)_+ = (e_{k+1})_+$ and

$$\max\{|h_1|_X, \dots, |h_i|_X\} \geq \Theta, \quad \max\{|q_1|_X, \dots, |q_j|_X\} \geq \Theta.$$

Since e_k and e_{k+1} are connected, h_i and q_1 will be connected $\mathcal{H}$-subpaths of p; in particular they cannot be contained in the same segment of the broken line p by Remark 5.9. If h_i and q_1 are $\mathcal{H}$-components of adjacent segments of p, then the components $h_1, \dots, h_i, q_1, \dots, q_j$ constitute a longer instance of consecutive backtracking in p, which contradicts the construction of e_k in Procedure 9.1.

Therefore it must be the case that the subpath p' of p between

$$(e_k)_+ = (h_i)_+ = v_{s_k} \quad \text{and} \quad (e_{k+1})_- = (q_1)_- = v_{t_k}$$

contains at least one full segment p_l (with $1 < l < n$). By Remark 9.2(c) the path p' has no $\mathcal{H}$-components h satisfying $|h|_X \geq \Theta$. Therefore we can combine Lemma 5.10 with the fact that p is $(B_0, c_0, \zeta, \Theta)$-tamable to deduce that

$$\ell(p') \geq \ell(p_l) \geq \frac{|p_l|_X}{\Theta} \geq \frac{B_0}{\Theta}. \tag{9-4}$$

Moreover, by Lemma 9.7 the path p' is $(4, c_3)$-quasigeodesic, so

$$\ell(p') \leq 4 d_{X \cup \mathcal{H}}((e_k)_+, (e_{k+1})_-) + c_3 \leq 4 + c_3,$$

where the last inequality is true because e_k and e_{k+1} are connected. Combined with (9-4), the above inequality gives $B_0 \leq (4 + c_3)\Theta$, which contradicts the choice of B_0 in (9-2).

Therefore e_k and e_{k+1} cannot be connected, for any $k \in \{1, \dots, m-1\}$. □

Proof of Proposition 9.4 The construction, together with Lemmas 9.6, 9.8 and 9.9, show that the Θ-shortcutting $\Sigma(p, \Theta) = f_0 e_1 f_1 \cdots f_{m-1} e_m f_m$ satisfies the hypotheses of Proposition 5.19 and e_k is non-trivial, for each $k = 1, \dots, m$. Therefore $\Sigma(p, \Theta)$ is (λ, c)-quasigeodesic without backtracking.

For the final claim of the proposition, consider any $k \in \{1, \dots, m\}$ and denote by e_k' the H_ν-component of $\Sigma(p, \Theta)$ containing e_k, for some $\nu \in \mathcal{N}$. Lemma 9.9 implies that e_k' is the concatenation $h_1 e_k h_2$, where h_1 is either trivial or it is an H_ν-component of f_{k-1}, and h_2 is either trivial or it is an H_ν-component of f_k. Combining the triangle inequality with Lemmas 9.6 and 9.8 and equation (9-1), we obtain

$$|e_k'|_X \geq |e_k|_X - |h_1|_X - |h_2|_X \geq \zeta - 2\rho \geq \eta,$$

as required. □

10 Metric quasiconvexity theorem

This section comprises a proof of Theorem 3.5, and, as usual, we work under Convention 6.1. First we will show that if some subgroups $Q' \leqslant Q$ and $R' \leqslant R$ satisfy conditions (C1)–(C5) with appropriately large constants, then minimal type path representatives of $\langle Q', R' \rangle$ meet the conditions of Proposition 9.4. We will then use the quasigeodesicity of shortcuttings of these path representatives to obtain properties (P1)–(P3).

Lemma 10.1 *Suppose that $Q' \leqslant Q$ and $R' \leqslant R$ satisfy* (C2) *with constant $B \geq 0$. Then*

$$\min_X((Q' \cup R') \setminus S) \geq B.$$

Proof Let $g \in (Q' \cup R') \setminus S$. If $g \in Q'$ then $g \notin R$ as $g \notin S$. Therefore $g \in Q' \setminus R \subseteq R\langle Q', R' \rangle R \setminus R$; hence $|g|_X \geq B$ by (C2). Similarly, if $g \in R'$ then $g \in Q\langle Q', R' \rangle Q \setminus Q$, and (C2) again implies that $|g|_X \geq B$. □

Notation 10.2 For the remainder of this section we fix the following notation:

- C_0 is the constant provided by Lemma 6.7;
- $c_0 = \max\{C_0, 14\delta\}$ and $c_3 = c_3(c_0)$ is the constant obtained by applying Lemma 4.11;
- $\lambda = \lambda(c_0)$ and $c = c(c_0)$ are the first two constants from Proposition 9.4;
- $C_1 \geq 0$ is the constant from Lemma 7.3;
- $\mathcal{P}_1$ is the finite family of parabolic subgroups of G defined by

$$\mathcal{P}_1 = \{tH_\nu t^{-1} \mid \nu \in \mathcal{N}, |t|_X \leq C_1\}.$$

Lemma 10.3 *For each $\eta \geq 0$ there are constants $C_3 = C_3(\eta) \geq 0$, $\zeta = \zeta(\eta) \geq 1$, $\Theta_1 = \Theta_1(\eta) \in \mathbb{N}$ and $B_1 = B_1(\eta) \geq 0$ such that the following is true.*

Suppose that $Q' \leqslant Q$ and $R' \leqslant R$ are subgroups satisfying conditions (C1)–(C5) *with constants $B \geq B_1$ and $C \geq C_3$ and family $\mathcal{P} \supseteq \mathcal{P}_1$. If $p = p_1 \cdots p_n$ is a minimal type path representative for an element $g \in \langle Q', R' \rangle$ then p is $(B, c_0, \zeta, \Theta_1)$-tamable.*

Moreover, let $\Sigma(p, \Theta_1) = f_0 e_1 f_1 \cdots f_{m-1} e_m f_m$ be the Θ_1-shortcutting of p obtained from Procedure 9.1 and let e'_k be the $\mathcal{H}$-component of $\Sigma(p, \Theta_1)$ containing e_k for $k = 1, \dots, m$. Then $\Sigma(p, \Theta_1)$ is a (λ, c)-quasigeodesic without backtracking and $|e'_k|_X \geq \eta$, for each $k = 1, \dots, m$.

Proof We define the following constants:

- $\zeta = \zeta(\eta, c_0) \geq 0$, the constant provided by Proposition 9.4;
- $C_3 = C_2(\zeta) \geq 0$, where $C_2(\zeta)$ is given by Proposition 8.5;

- $\Theta_1 = \max\{\Theta_0(\zeta), \zeta\}$, where Θ_0 is the constant of Lemma 7.6;
- $B_1 = \max\{B_0(\Theta_1, c_0), C_2(\zeta)\} \geq 0$, where B_0 is the remaining constant of Proposition 9.4.

Let $B \geq B_1$ and $C \geq C_3$. Suppose that Q', R', g and p are as in the statement of the lemma. In view of Remark 6.5, $\tilde{p}_i \in (Q' \cup R') \setminus S$, for every $i = 2, \dots, n-1$. Therefore, by Lemma 10.1, we have

$$|p_i|_X \geq B \quad \text{for each } i = 2, \dots, n-1. \tag{10-1}$$

On the other hand, Lemma 6.7 tells us that

$$\langle (p_i)_-, (p_{i+1})_+ \rangle^{\mathrm{rel}}_{(p_i)_+} \leq C_0 \leq c_0 \quad \text{for each } i = 1, \dots, n-1. \tag{10-2}$$

Now suppose that p has consecutive backtracking along $\mathcal{H}$-components $h_i, \dots, h_j$ of segments $p_i, \dots, p_j$ satisfying

$$\max\{|h_i|_X, \dots, |h_j|_X\} \geq \Theta_1.$$

If $j = i + 1$ then Lemma 7.6 and the choice of Θ_1 give that $d_X((h_i)_-, (h_j)_+) \geq \zeta$. Otherwise Proposition 8.5 gives the same inequality. The above together with (10-1) and (10-2) show that p is $(B, c_0, \zeta, \Theta_1)$-tamable.

The remaining claims of the lemma follow from Proposition 9.4. □

We can now deduce the relative quasiconvexity of $\langle Q', R' \rangle$ by applying Lemma 10.3 with $\eta = 0$.

Proposition 10.4 *Let $\beta_1 = B_1(0)$ and $\gamma_1 = C_3(0)$ be the constants provided by Lemma 10.3 applied to the case when $\eta = 0$.*

Suppose that $Q' \leqslant Q$ and $R' \leqslant R$ are relatively quasiconvex subgroups of G satisfying conditions (C1)–(C5) *with family $\mathcal{P} \supseteq \mathcal{P}_1$ and constants $B \geq \beta_1$ and $C \geq \gamma_1$. Then the subgroup $\langle Q', R' \rangle$ is relatively quasiconvex in G.*

Proof By assumption the subgroups Q' and R' are relatively quasiconvex, with some quasiconvexity constant $\varepsilon' \geq 0$. For any element $g \in \langle Q', R' \rangle$ consider a geodesic τ in $\Gamma(G, X \cup \mathcal{H})$ with $\tau_- = 1$ and $\tau_+ = g$. Let u be any vertex of τ.

Since $g \in \langle Q', R' \rangle$, it has a path representative $p = p_1 \cdots p_n$ of minimal type, with $p_- = 1$. Let $\Sigma(p, \Theta) = f_0 e_1 f_1 \cdots f_{m-1} e_m f_m$ be the Θ-shortcutting of p obtained from Procedure 9.1, where $\Theta = \Theta_1(0)$ is provided by Lemma 10.3. This lemma implies that p is (B, c_0, ζ, Θ)-tamable and $\Sigma(p, \Theta)$ is a (λ, c)-quasigeodesic without backtracking, where $\lambda \geq 1$ and $c \geq 0$ are the constants fixed in Notation 10.2. Therefore, by Proposition 5.17, there is a phase vertex v of $\Sigma(p, \Theta)$ with $d_X(u, v) \leq \kappa(\lambda, c, 0)$.

Since each e_i is a single edge, the vertex v lies on the geodesic subpath f_i of $\Sigma(p, \Theta)$, for some $i \in \{0, \dots, m\}$. The subpath of p sharing endpoints with f_i is $(4, c_3)$-quasigeodesic by Lemma 9.7. Hence there is a vertex w of p such that $d_X(v, w) \leq \kappa(4, c_3, 0)$, by Proposition 5.17.

Now w is a vertex of a subpath p_j of p, for some $j \in \{1, \dots, n\}$. Let $x = (p_j)_-$, and note that $x \in \langle Q', R' \rangle$. Without loss of generality, suppose that $\tilde{p}_j \in Q'$ (the case when $\tilde{p}_j \in R'$ can be treated similarly). Then by the relative quasiconvexity of Q', $d_X(w, xQ') \le \varepsilon'$, whence $d_X(w, \langle Q', R' \rangle) \le \varepsilon'$. Therefore

$$\begin{aligned} d_X(u, \langle Q', R' \rangle) &\le d_X(u, v) + d_X(v, w) + d_X(w, \langle Q', R' \rangle) \\ &\le \kappa(\lambda, c, 0) + \kappa(4, c_3, 0) + \varepsilon', \end{aligned}$$

so $\langle Q', R' \rangle$ is a relatively quasiconvex subgroup of G, with the quasiconvexity constant

$$\kappa(\lambda, c, 0) + \kappa(4, c_3, 0) + \varepsilon'. \qquad \square$$

We will next show that properties (P2) and (P3) will be satisfied if one chooses the constants B and C of (C1)–(C5) to be sufficiently large with respect to A.

Lemma 10.5 *For any $A \ge 0$ there exist constants $\beta_2 = \beta_2(A) \ge 0$ and $\gamma_2 = \gamma_2(A) \ge 0$ such that if $Q' \leqslant Q$ and $R' \leqslant R$ satisfy conditions* (C1)–(C5) *with constants $B \ge \beta_2$ and $C \ge \gamma_2$ and family $\mathcal{P} \supseteq \mathcal{P}_1$, then*

$$\min_X(\langle Q', R' \rangle \setminus S) \ge A.$$

Proof Given any $A \ge 0$ let $\eta = \eta(\lambda, c, A)$ be the constant provided by Lemma 5.12. Using Lemma 10.3, set

$$\Theta = \Theta_1(\eta), \quad \gamma_2 = C_3(\eta), \quad \beta_2 = \max\{B_1(\eta), (4A + c_3)\Theta\}.$$

Suppose that Q' and R' satisfy conditions (C1)–(C5) with constants $B \ge \beta_2$ and $C \ge \gamma_2$, and let $g \in \langle Q', R' \rangle$ be any element with $|g|_X < A$. Let $p = p_1 \cdots p_n$ be a path representative of g with minimal type. By Lemma 10.3, p is $(B, c_0, \zeta, \Theta_1)$-tamable, the Θ-shortcutting $\Sigma(p, \Theta) = f_0 e_1 f_1 \cdots f_{m-1} e_m f_m$ is (λ, c)-quasigeodesic without backtracking, and, for each $k = 1, \dots, m$, e'_k, the $\mathcal{H}$-component of $\Sigma(p, \Theta)$ containing e_k, is isolated and satisfies $|e'_k|_X \ge \eta$.

If $m \ge 1$, then, according to Lemma 5.12, $|g|_X = |\Sigma(p, \Theta)|_X \ge A$, contradicting our assumption. Therefore it must be the case that $m = 0$ and $\Sigma(p, \Theta) = f_0$. Since $p_- = (f_0)_-$ and $p_+ = (f_0)_+$, Lemma 9.7 tells us that p is $(4, c_3)$-quasigeodesic. Moreover, following Remark 9.2(c), we see that p_i has no $\mathcal{H}$-component h with $|h|_X \ge \Theta$, for each $i = 1, \dots, n$.

Now, arguing by contradiction, suppose that $g \notin S$. Then $\tilde{p}_1 \in (Q' \cup R') \setminus S$ (by Remark 6.5), so $|p_1|_X \ge B \ge \beta_2$, by Lemma 10.1. Lemma 5.10 now implies that

$$\ell(p_1) \ge \beta_2 / \Theta \ge 4A + c_3.$$

Since $\ell(p) \ge \ell(p_1)$, the $(4, c_3)$-quasigeodesicity of p yields

$$A > |g|_X \ge |g|_{X \cup \mathcal{H}} = |p|_{X \cup \mathcal{H}} \ge \tfrac{1}{4}(\ell(p) - c_3) \ge A,$$

which is a contradiction. Therefore $g \in S$ and the lemma is proved. $\square$

In order to prove that property (P3) holds for the subgroups Q' and R', we need to consider path representatives of elements $g \in Q\langle Q', R'\rangle R$. These path representatives will necessarily have to be slightly different from those in Definition 6.2.

Definition 10.6 (path representative, II) Let g be an element of $Q\langle Q', R'\rangle R$, and suppose that $p = qp_1 \cdots p_n r$ is a broken line in $\Gamma(G, X \cup \mathcal{H})$, satisfying all of the following conditions:

- $\tilde{p} = g$;
- $\tilde{q} \in Q$ and $\tilde{r} \in R$;
- $\tilde{p}_i \in Q' \cup R'$, for each $i \in \{1, \dots, n\}$.

Then we say that p is a *path representative* of g in the product $Q\langle Q', R'\rangle R$.

Similarly to Definition 6.3, we can define types for such path representatives.

Definition 10.7 (type of a path representative, II) Suppose that $p = qp_1 \cdots p_n r$ is a path representative of some $g \in Q\langle Q', R'\rangle R$, as described in Definition 10.6. Let Y denote the set of all $\mathcal{H}$-components of the segments of p. We define the *type* of the path representative p to be the triple

$$\tau(p) = \left(n, \ell(p), \sum_{y \in Y} |y|_X\right) \in \mathbb{N}_0{}^3.$$

Remark 10.8 Note that, by Definition 10.6, a path representative $p = qp_1 \cdots p_n r$, of an element $g \in Q\langle Q', R'\rangle R \setminus QR$, must necessarily satisfy $n > 0$. Moreover, if p has minimal type (so n is the smallest possible) then $\tilde{p}_1 \in R' \setminus S$, $\tilde{p}_n \in Q' \setminus S$ and the labels of $p_1, \dots, p_n$ will alternate between representing elements of $R' \setminus S$ and $Q' \setminus S$. It follows that the integer n must be even, so $n \geq 2$.

For example, if $g \in R'Q' \setminus QR$ then a minimal type path representative of g will have the form qp_1p_2r, where q and r are trivial paths, $\tilde{p}_1 \in R'$ and $\tilde{p}_2 \in Q'$.

It is not difficult to check that the results of Sections 6, 7, and 8 hold equally well for minimal type path representatives of the above form for elements $g \in Q\langle Q', R'\rangle R \setminus QR$, with only superficial adjustments to the proofs in those sections. It follows that Lemma 10.3 also remains valid in these settings.

Lemma 10.9 *In the statement of Lemma 10.5 we can add that*

$$\min_X(Q\langle Q', R'\rangle R \setminus QR) \geq A.$$

Proof For any $A \geq 0$ we define the constants η, Θ, γ_2 and β_2 exactly as in Lemma 10.5.

Suppose that for some element $g \in Q\langle Q', R'\rangle R \setminus QR$ we have $|g|_X < A$. Let $p = qp_1 \cdots p_n r$ be a minimal type path representative of g, of the form described in Definition 10.6.

Arguing in the same way as in Lemma 10.5, we can deduce that p is $(4, c_3)$-quasigeodesic and for each $i = 1, \dots, n$, p_i has no $\mathcal{H}$-component h with $|h|_X \geq \Theta$.

According to Remark 10.8, $n \geq 2$ and $\tilde{p}_1 \in R' \setminus S$. So, by Lemma 10.1, $|p_1|_X \geq B \geq \beta_2$. The same argument as in Lemma 10.5 now yields that $|g|_X \geq A$, leading to a contradiction. Therefore it must be that $|g|_X \geq A$ for any $g \in Q\langle Q', R'\rangle R \setminus QR$. □

We are finally able to prove Theorem 3.5.

Proof of Theorem 3.5 Choose $\mathcal{P}$ to be the finite family $\mathcal{P}_1$, defined in Notation 10.2. Given any $A \geq 0$, we apply Proposition 10.4 and Lemma 10.5 to define the constants

$$B = \max\{\beta_1, \beta_2(A)\} \quad \text{and} \quad C = \max\{\gamma_1, \gamma_2(A)\}.$$

Suppose that $Q' \leqslant Q$ and $R' \leqslant R$ are subgroups satisfying conditions (C1)–(C5) with constants B and C and the finite family of parabolic subgroups $\mathcal{P}$. Then property (P1) holds by Proposition 10.4, while properties (P2) and (P3) are satisfied by Lemmas 10.5 and 10.9 respectively. □

11 Using separability to establish the conditions of the quasiconvexity theorem

In this section we will show how one can prove the existence of finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$, satisfying the conditions (C1)–(C5) from Section 3.1, using certain separability assumptions. We start with finding such assumptions for establishing (C2) and (C3).

Proposition 11.1 *Let G be a group generated by a finite subset X, let $Q, R \leqslant G$ and $S = Q \cap R$, and let $\mathcal{P}$ be a finite collection of subgroups of G. Suppose that Q and R are separable in G and PS is separable in G, for each $P \in \mathcal{P}$.*

Then for any constants $B, C \geq 0$ there exists a finite-index subgroup $L \leqslant_f G$, with $S \subseteq L$, such that conditions (C2) *and* (C3) *are satisfied by arbitrary subgroups $Q' \leqslant Q \cap L$ and $R' \leqslant R \cap L$.*

Proof Combining the separability of Q and R in G with Lemma 4.16, we can find $E_1, E_2 \lhd_f G$ such that $\min_X(QE_1 \setminus Q) \geq B$ and $\min_X(RE_2 \setminus R) \geq B$. Set $N_0 = E_1 \cap E_2 \lhd_f G$ and observe that

$$QSN_0Q = QN_0Q = QQN_0 = QN_0 \subseteq QE_1,$$

as Q is a subgroup containing S and normalising N_0 in G. Similarly, $RSN_0R = RN_0 \subseteq RE_2$; therefore

$$\min_X(QSN_0Q \setminus Q) \geq B \quad \text{and} \quad \min_X(RSN_0R \setminus R) \geq B. \tag{11-1}$$

Let $\mathcal{P} = \{P_1, \dots, P_k\}$. The assumptions imply that for every $i \in \{1, \dots, k\}$ the double coset P_iS is separable in G; hence we can apply Lemma 4.16 again to find finite-index normal subgroups $N_i \lhd_f G$ satisfying

$$\min_X(P_iSN_i \setminus P_iS) \geq C \quad \text{for each } i = 1, \dots, k. \tag{11-2}$$

Now set $L = \bigcap_{i=0}^{k} SN_i \leqslant_f G$, and choose arbitrary subgroups $Q' \leqslant Q \cap L$ and $R' \leqslant R \cap L$. Then $S \subseteq L$ and $\langle Q', R' \rangle \subseteq L \subseteq SN_i$, for all $i = 0, \dots, k$, by construction; hence (C2) holds by (11-1) and (C3) holds by (11-2), as desired. $\square$

To establish condition (C5) we need to be able to lift certain finite-index subgroups of a maximal parabolic subgroup $P \leqslant G$ to finite-index subgroups of G in a controlled way. The next statement shows how a double coset separability assumption can help with this task.

Lemma 11.2 *Let G be a group, $P, Q \leqslant G$ be subgroups of G and let $K \leqslant_f P$ be a finite-index subgroup of P, with $Q \cap P \subseteq K$. If KQ is separable in G, then there is a finite-index subgroup $M \leqslant_f G$ such that $Q \subseteq M$ and $M \cap P \subseteq K$.*

Proof Let $P = K \cup Kh_1 \cup \dots \cup Kh_m$, where $h_1, \dots, h_m \in P \setminus K$. Note that $KQ \cap P = K(Q \cap P) = K$, so $h_1, \dots, h_m \notin KQ$. The double coset KQ is profinitely closed, so, by Lemma 4.16(a), there exists $N \lhd_f G$ such that

$$\{h_1, \dots, h_m\} \cap KQN = \varnothing.$$

Let $M = QN \leqslant_f G$, so that the above implies $Kh_i \cap M = \varnothing$, for each $i = 1, \dots, m$. We then have $Q \subseteq M$ and $M \cap P \subseteq K$, as required. $\square$

We are now in position to prove the main result of this section.

Theorem 11.3 *Assume that G is a group generated by a finite set X, $Q, R \leqslant G$ are subgroups of G, and denote $S = Q \cap R$. Let $\mathcal{P}$ be a finite collection of subgroups of G such that for every $P \in \mathcal{P}$ all of the following hold*:

- (S1) *Q and R are separable in G*;
- (S2) *the double coset PS is separable in G*;
- (S3) *for all $K \leqslant_f P$ and $T \leqslant_f Q$, satisfying $S \subseteq T$ and $T \cap P \subseteq K$, the double coset KT is separable in G*;
- (S4) *for all $U \leqslant_f Q \cap P$, with $S \cap P \subseteq U$, the double coset $U(R \cap P)$ is separable in P.*

Then, given arbitrary constants $B, C \geq 0$, there exist finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ such that conditions (C1)–(C5) *are all satisfied.*

More precisely, there exists $L \leqslant_f G$, with $S \subseteq L$, such that for any $L' \leqslant_f L$, satisfying $S \subseteq L'$, we can choose $Q' = Q \cap L' \leqslant_f Q$ and there exists $M \leqslant_f L'$, with $Q' \subseteq M$, such that for any $M' \leqslant_f M$, satisfying $Q' \subseteq M'$, we can choose $R' = R \cap M' \leqslant_f R$.

Proof The idea is that (S1) will take care of condition (C2), (S2) will take care of (C3), and (S3) and (S4) will take care of (C5). The subgroups Q' and R' will satisfy $Q' = Q \cap M'$ and $R = R \cap M'$, for some $M' \leqslant_f G$, with $S \subseteq M'$, which will immediately imply (C1) and (C4).

Let $\mathcal{P} = \{P_1, \dots, P_k\}$. Arguing just like in the proof of Proposition 11.1 (using the assumptions (S1) and (S2)), we can find finite-index normal subgroups $N_i \lhd_f G$, $i = 0, \dots, k$, such that

$$\min_X(QSN_0Q \setminus Q) \geq B, \quad \min_X(RSN_0R \setminus R) \geq B,$$
$$\min_X(P_iSN_i \setminus P_iS) \geq C \quad \text{for } i = 1, \dots, k.$$

We can now define a finite-index subgroup $L \leqslant_f G$ by $L = \bigcap_{i=0}^k SN_i$. Note that $S \subseteq L$ by construction, and for each $i \in \{1, \dots, k\}$ we have

$$\min_X(QLQ \setminus Q) \geq B, \quad \min_X(RLR \setminus R) \geq B, \quad \min_X(P_iL \setminus P_iS) \geq C. \tag{11-3}$$

Choose an arbitrary finite-index subgroup $L' \leqslant_f L$, with $S \subseteq L'$, and define $Q' = Q \cap L'$, so that $S \leqslant Q' \leqslant_f Q$.

To construct $R' \leqslant_f R$, consider any $i \in \{1, \dots, k\}$ and denote

$$Q_i = Q \cap P_i, \quad R_i = R \cap P_i, \quad Q_i' = Q' \cap P_i \leqslant_f Q_i.$$

Choose some elements $a_{i1}, \dots, a_{in_i} \in Q_i$ such that $Q_i = \bigsqcup_{j=1}^{n_i} a_{ij} Q_i'$. Condition (S4) implies that the subset $Q_i'R_i$ is separable in P_i; hence, by claim (c) of Lemma 4.16, there exists $F_i \lhd_f P_i$ such that

$$\min_X(a_{ij}Q_i'R_iF_i \setminus a_{ij}Q_i'R_i) \geq C \quad \text{for } j = 1, \dots, n_i. \tag{11-4}$$

Define $K_i = Q_i'F_i \leqslant_f P_i$. Then $Q' \cap P_i = Q_i' \subseteq K_i$ and $a_{ij}K_iR_i = a_{ij}Q_i'R_iF_i$, for each $j = 1, \dots, n_i$. Therefore, from (11-4) we can deduce that

$$\min_X(a_{ij}K_iR_i \setminus a_{ij}Q_i'R_i) \geq C \quad \text{for all } j = 1, \dots, n_i. \tag{11-5}$$

By (S3), the double coset K_iQ' is separable in G, so we can apply Lemma 11.2 to find $M_i \leqslant_f G$ such that $Q' \subseteq M_i$ and $M_i \cap P_i \subseteq K_i$.

We now let $M = \bigcap_{i=1}^k M_i \cap L'$ and observe that $Q' \leqslant M \leqslant_f L'$ and $M \cap P_i \subseteq K_i$ for each $i \in \{1, \dots, k\}$. Inequality (11-5) yields

$$\min_X\big(a_{ij}(M \cap P_i)R_i \setminus a_{ij}Q_i'R_i\big) \geq C \quad \text{for all } i = 1, \dots, k \text{ and } j = 1, \dots, n_i. \tag{11-6}$$

We can now choose an arbitrary finite-index subgroup $M' \leqslant_f M$, with $Q' \subseteq M'$, and define $R' = R \cap M'$. Observe that $M' \leqslant_f G$, by construction, hence $R' \leqslant_f R$.

Let us check that the subgroups Q' and R' obtained above satisfy conditions (C1)–(C5). Indeed, by construction, $S = Q \cap R \subseteq Q'$, so $S \subseteq R \cap M' = R'$; hence

$$S \subseteq Q' \cap R' \subseteq Q \cap R = S.$$

Thus (C1) holds. We also have $Q' = Q \cap L' = Q \cap M'$, as $Q' \subseteq M' \subseteq L'$; hence

$$Q' \subseteq Q \cap \langle Q', R' \rangle \subseteq Q \cap M' = Q'.$$

Thus $Q \cap \langle Q', R' \rangle = Q'$. After intersecting both sides of the latter equation with an arbitrary $P \in \mathcal{P}$, we get $Q_P \cap \langle Q', R' \rangle = Q'_P$; hence

$$Q'_P \subseteq Q_P \cap \langle Q'_P, R'_P \rangle \subseteq Q_P \cap \langle Q', R' \rangle = Q'_P.$$

Thus $Q_P \cap \langle Q'_P, R'_P \rangle = Q'_P$. Similarly, $R_P \cap \langle Q'_P, R'_P \rangle = R'_P$, so condition (C4) is satisfied.

Conditions (C2) and (C3) hold by (11-3), because $Q', R' \subseteq L$ by construction.

To prove (C5), take $P_i \in \mathcal{P}$ for any $i \in \{1, \dots, k\}$, and denote $Q_i = Q \cap P_i$, $Q'_i = Q' \cap P_i$, $R_i = R \cap P_i$ and $R'_i = R' \cap P_i$, as before. For any $q \in Q_i$ there exists $j \in \{1, \dots, n_i\}$ such that $q \in a_{ij} Q'_i$. It follows that

$$q \langle Q'_i, R'_i \rangle R_i = a_{ij} \langle Q'_i, R'_i \rangle R_i \quad \text{and} \quad q Q'_i R_i = a_{ij} Q'_i R_i. \tag{11-7}$$

Since $\langle Q'_i, R'_i \rangle \leqslant M \cap P_i$, we can combine (11-7) with (11-6) to deduce that

$$\min\nolimits_X (q \langle Q'_i, R'_i \rangle R_i \setminus q Q'_i R_i) \geq C,$$

which establishes condition (C5). □

12 Double coset separability in amalgamated free products

In this section we develop a method for establishing the separability assumptions (S2) and (S3) of Theorem 11.3 using amalgamated products. The idea is that when G is a relatively hyperbolic group, P is a maximal parabolic subgroup and Q is a relatively quasiconvex subgroup of G, we can apply the combination theorem of Martínez-Pedroza (Theorem 5.26) to find a finite-index subgroup $H \leqslant_f P$ such that $A = \langle H, Q \rangle \cong H *_{H \cap Q} Q$, so proving the separability of PQ in G can be reduced to proving the separability of HQ in the amalgamated free product A.

The next proposition gives a new criterion for showing separability of double cosets in amalgamated free products. This criterion may be of independent interest.

Proposition 12.1 *Let $A = B *_D C$ be an amalgamated free product, where we consider B, C and D as subgroups of A with $B \cap C = D$. Suppose that D is separable in A, and $U \subseteq B$ and $V \subseteq C$ are arbitrary subsets.*

If the product UD (respectively, DV) is separable in A then the product UC (respectively, BV) is separable in A.

Proof We will prove the statement in the case of UC, as the other case is similar.

If $U = \varnothing$ then $UC = \varnothing$, so we can suppose that U is non-empty. Take any $u \in U$. According to Remark 4.12, without loss of generality we can replace U with $u^{-1}U$ to assume that $1 \in U$.

Consider any element $g \in A \setminus UC$; since $1 \in U$, we deduce that $g \notin C$. We will construct a homomorphism from A to a finite group L which separates the image of g from the image of UC.

Since $g \notin D$, it has a reduced form $g = x_1 x_2 \cdots x_k$, where x_i belongs to one of the factors B or C, for each i, consecutive elements x_i and x_{i+1} belong to different factors, and $x_i \notin D$ for all $i = 1, \dots, k$ (see [35, page 187]).

Since D is separable in A, by Lemma 4.16(a) there is a finite group M and a homomorphism $\varphi\colon A \to M$ such that

$$\varphi(x_i) \notin \varphi(D) \text{ in } M \quad \text{for every } i = 1, \dots, k. \tag{12-1}$$

Denote by $\overline{B}$, $\overline{C}$ and $\overline{D}$ the φ-images if B, C and D in M respectively. We can then consider the amalgamated free product $\overline{A} = \overline{B} *_{\overline{D}} \overline{C}$, together with the natural homomorphism $\psi\colon A \to \overline{A}$, which is compatible with φ on B and C (in other words, $\psi|_B = \varphi|_B$ and $\psi|_C = \varphi|_C$). It follows that φ factors through ψ. That is, $\varphi = \bar{\varphi} \circ \psi$, where $\bar{\varphi}\colon \overline{A} \to M$ is the natural homomorphism extending the embeddings of $\overline{B}$ and $\overline{C}$ in M. Equation (12-1) now implies that

$$\psi(x_i) \notin \overline{D} \text{ in } \overline{A} \quad \text{for every } i = 1, \dots, k. \tag{12-2}$$

Denote $\bar{x}_i = \psi(x_i) \in \overline{A}$, $i = 1, \dots, k$. In view of (12-2), $\psi(g) = \bar{x}_1 \cdots \bar{x}_k$ is a reduced form in the amalgamated free product $\overline{A}$. We will now consider several cases.

Case 1 Assume that $k \geq 3$.

Then the above reduced form for $\psi(g)$ has length $k \geq 3$, so by the normal form theorem for amalgamated free products [35, Theorem IV.2.6], it cannot be equal to an element from $\psi(UC) = \psi(U)\overline{C} \subseteq \overline{B}\,\overline{C}$, which would necessarily have a reduced form of length at most 2 in $\overline{A}$. Therefore $\psi(g) \notin \psi(UC)$ in $\overline{A}$.

Since $\overline{B}$ and $\overline{C}$ are finite groups, their amalgamated free product $\overline{A}$ is residually finite (in fact, $\overline{A}$ is a virtually free group — see [55, Proposition 2.6.11]), so the finite subset $\psi(UC)$ is closed in the profinite topology on $\overline{A}$. Hence there is a finite group L and a homomorphism $\eta\colon \overline{A} \to L$ such that $\eta(\psi(g)) \notin \eta(\psi(UC))$ in L. The composition $\eta \circ \psi\colon A \to L$ is the required homomorphism separating the image of g from the image of UC, and the consideration of Case 1 is complete.

Case 2 Suppose that $k = 2$, $x_1 \in C \setminus D$ and $x_2 \in B \setminus D$.

Then $\bar{x}_1 \in \overline{C} \setminus \overline{D}$ and $\bar{x}_2 \in \overline{B} \setminus \overline{D}$ by (12-2), so $\psi(g) = \bar{x}_1 \bar{x}_2$ is a reduced form of length 2 in $\overline{A}$. Again, the normal form theorem for amalgamated free products implies that $\psi(g) \notin \overline{B}\,\overline{C}$ in $\overline{A}$; hence $\psi(g) \notin \psi(UC)$ and we can find the required finite quotient L of A as in Case 1.

Case 3 $g = bc$, where $b \in B \setminus UD$ and $c \in C$ (here we allow $c \in D$, so this case also covers the situation when $k = 1$).

This is the only case where we need to use the assumption that UD is separable in A. This assumption implies that we can find a finite group M and a homomorphism $\varphi\colon A \to M$ satisfying

$$\varphi(b) \notin \varphi(UD) \quad \text{in } M.$$

As above, we can construct the amalgamated free product $\bar{A} = \bar{B} *_{\bar{D}} \bar{C}$, together with the natural homomorphism $\psi\colon A \to \bar{A}$, such that φ factors through ψ. It follows that

$$\psi(b) \notin \psi(UD) = \psi(U)\bar{D} \quad \text{in } \bar{A}. \tag{12-3}$$

Observe that $\psi(g) \notin \psi(UC) = \psi(U)\bar{C}$ in $\bar{A}$. Indeed, otherwise we would have

$$\psi(b) = \psi(g)\psi(c^{-1}) \in \psi(U)\bar{C} \cap \bar{B} = \psi(U)(\bar{C} \cap \bar{B}) = \psi(U)\bar{D},$$

which would contradict (12-3) (in the first equality we used the fact that $\bar{B}$ is a subgroup of $\bar{A}$ containing the subset $\psi(U)$). We can now argue as in Case 1 above to find a homomorphism from A to a finite group L separating the image of g from the image of UC.

It is not hard to see that since $g \notin UC$ in A, the above three cases cover all possibilities; hence the proof is complete. □

In the next two corollaries we assume that $A = B *_D C$ is the amalgamated free product of its subgroups B and C, with $B \cap C = D$.

Corollary 12.2 *Suppose that D is a separable subgroup in A. Then B, C and BC are all separable in A.*

Proof The separability of C and B in A follows from Proposition 12.1, after choosing $U = \{1\}$ and $V = \{1\}$.

The separability of BC is also a consequence of Proposition 12.1, where we take $U = B$ (so that $UD = BD = B$). □

We will not need the next corollary in this paper, but it may be of independent interest and can be used to strengthen some of the statements proved in Section 13.

Corollary 12.3 *Suppose that $U \subseteq B$ and $V \subseteq C$ are subsets such that UD and DV are separable in A. Then the triple product UDV is separable in A.*

Proof If either U or V are empty then UDV is empty, and, hence, separable in A. Thus we can suppose that there exist some elements $u \in U$ and $v \in V$. By Remark 4.12. the subsets $u^{-1}UD \subseteq B$ and $DVv^{-1} \subseteq C$ are separable in A. Since both of them contain D, we see that $D = u^{-1}UD \cap DVv^{-1}$; thus D is separable in A.

By Proposition 12.1, the products UC and BV are separable in A, so the statement follows from the observation that

$$UC \cap BV = UDV \quad \text{in } A. \qquad \square$$

In the case when U and V are subgroups, the above corollary shows that we can use separability of double cosets UD and DV to deduce separability of the triple coset UDV. Moreover, if both U and V are subgroups containing D, Corollary 12.3 implies that the double coset $UV = UDV$ is separable in A, as long as U and V are separable in A.

13 Separability of double cosets when one factor is parabolic

Throughout this section we will assume that G is group generated by a finite subset X and hyperbolic relative to a collection of peripheral subgroups $\{H_\nu \mid \nu \in \mathcal{N}\}$, with $|\mathcal{N}| < \infty$.

Our goal in this section will be to establish separability of double cosets required by conditions (S2) and (S3) of Theorem 11.3. All statements in this section will assume that finitely generated relatively quasiconvex subgroups of G are separable — that is, G is QCERF (see Definition 1.1).

Lemma 13.1 *Suppose that G is QCERF. If A is a finitely generated relatively quasiconvex subgroup of G then every subset of A which is closed in $\mathcal{PT}(A)$ is also closed in $\mathcal{PT}(G)$.*

Proof By Lemma 5.22 every subgroup of finite index in A is finitely generated and relatively quasiconvex; hence it is separable in G as G is QCERF. The claim of the lemma now follows from Lemma 4.13(b). $\square$

The following statement is essentially a corollary of the combination theorem of Martínez-Pedroza (Theorem 5.26).

Proposition 13.2 *Suppose that G is QCERF. Let P be a maximal parabolic subgroup of G, let $Q \leqslant G$ be a finitely generated relatively quasiconvex subgroup and let $D = P \cap Q$. Then there exists a finite-index subgroup $H \leqslant_f P$ such that all of the following properties hold:*

- *$H \cap Q = D$;*
- *the subgroup $A = \langle H, Q\rangle$ is relatively quasiconvex in G;*
- *A is naturally isomorphic to $H *_D Q$;*
- *D is separable in A;*
- *every subset of A which is closed in $\mathcal{PT}(A)$ is also closed in $\mathcal{PT}(G)$.*

Proof Let $C \geq 0$ be the constant provided by Theorem 5.26, applied to the maximal parabolic subgroup P and the relatively quasiconvex subgroup Q. By QCERF-ness, Q is separable in G, so by Lemma 4.16 there exists $N \lhd_f G$ such that $\min_X(QN \setminus Q) \geq C$. Therefore, after setting $H = P \cap QN \leqslant_f P$, we get $\min_X(H \setminus D) = \min_X(H \setminus Q) \geq C$.

Note that since $D = P \cap Q \subseteq H \subseteq P$, we have $H \cap Q = D$. Hence we can apply Theorem 5.26 to conclude that $A = \langle H, Q\rangle$ is relatively quasiconvex in G and is naturally isomorphic to the amalgamated free product $H *_D Q$.

Recall, from Lemma 5.24 and Corollary 5.23, that P is finitely generated and relatively quasiconvex in G; hence it is separable in G by QCERF-ness. It follows that $D = P \cap Q$ is separable in G, which implies that it is separable in A by Lemma 4.13.

Observe that H and Q are both finitely generated, hence A is finitely generated and relatively quasiconvex in G. Therefore Lemma 13.1 yields the last assertion of the proposition, that every subset of A which is closed in $\mathcal{PT}(A)$ is also closed in $\mathcal{PT}(G)$. $\square$

By combining Proposition 13.2 with Proposition 12.1 we obtain the first double coset separability result when one of the factors is parabolic and the other one is finitely generated and relatively quasiconvex.

Proposition 13.3 *Assume that G is QCERF. Let P be a maximal parabolic subgroup of G, let $R \leqslant G$ be a finitely generated relatively quasiconvex subgroup of G. Suppose that $D \leqslant P$ is a subgroup satisfying the following condition*:

$$\text{for each } U \leqslant_f D \text{ the double coset } U(P \cap R) \text{ is separable in } P. \tag{13-1}$$

Then the double coset DR is separable in G.

Proof According to Proposition 13.2, there exists $H \leqslant_f P$ such that the subgroup $A = \langle H, R\rangle$ is naturally isomorphic to the amalgamated free product $H *_E R$, where $E = P \cap R = H \cap R$ is separable in A, and every closed subset from $\mathcal{PT}(A)$ is separable in G.

Denote $U = D \cap H \leqslant_f D$. By assumption (13-1), UE is separable in P. Since P is finitely generated and relatively quasiconvex in G, we can conclude that UE is separable in G by Lemma 13.1. As $UE \subseteq A \leqslant G$, UE will also be closed in $\mathcal{PT}(A)$, so we can apply Proposition 12.1 to deduce that the double coset UR is closed in $\mathcal{PT}(A)$. It follows that this double coset is separable in G and, since $U \leqslant_f D$, Lemma 4.14 implies that DR is separable in G, as desired. □

We can now prove that (S3) of Theorem 11.3 holds as long as the relatively hyperbolic group G is QCERF.

Corollary 13.4 *Suppose that G is QCERF, P is a maximal parabolic subgroup of G and $Q \leqslant G$ is a finitely generated relatively quasiconvex subgroup. Then for all finite-index subgroups $K \leqslant_f P$ and $T \leqslant_f Q$ the double coset KT is separable in G.*

Proof Note that T is finitely generated and relatively quasiconvex in G by Lemma 5.22. Hence, to apply Proposition 13.3 we simply need to check that for any $U \leqslant_f K$ the double coset $U(P \cap T)$ is separable in P. The latter is true because $U(P \cap T)$ is a basic closed set in $\mathcal{PT}(P)$, being a finite union of right cosets to $U \leqslant_f P$. Therefore KT is separable in G by Proposition 13.3. □

The proof of (S2) of Theorem 11.3 is slightly more involved because the intersection of two finitely generated relatively quasiconvex subgroups need not be finitely generated.

Proposition 13.5 *Let P be a maximal parabolic subgroup of G, let $Q, R \leqslant G$ be finitely generated relatively quasiconvex subgroups, let $S = Q \cap R$ and $D = P \cap Q$. Suppose that G is QCERF and condition* (13-1) *is satisfied. Then the double coset PS is separable in G.*

Proof Proposition 13.3 tells us that the double coset DR is separable in G, and G is QCERF so Q is separable in G. Now, observe that $DR \cap Q = D(R \cap Q) = DS$, because $D \leqslant Q$. It follows that the double coset DS is separable in G.

According to Proposition 13.2, there exists a finite-index subgroup $H \leqslant_f P$ such that $H \cap Q = D$, $A = \langle H, Q \rangle \cong H *_D Q$, D is separable in A and every closed subset in $\mathcal{PT}(A)$ is closed in $\mathcal{PT}(G)$. The double coset DS is separable in A by Lemma 4.13, so HS is closed in $\mathcal{PT}(A)$ by Proposition 12.1. It follows that HS is closed in $\mathcal{PT}(G)$, which implies that the double coset PS is separable in G by Lemma 4.14. □

14 Quasiconvexity of a virtual join from separability properties

In this section we will prove Theorems 1.2 and 1.3 from the introduction. The latter follows from the following result and the observation that a finite-index subgroup of a relatively quasiconvex subgroup is itself relatively quasiconvex (see Lemma 5.22).

Theorem 14.1 *Let G be a group generated by a finite set X and hyperbolic relative to a finite collection of abelian subgroups. Assume that G is QCERF. If $Q, R \leqslant G$ are relatively quasiconvex subgroups and $S = Q \cap R$ then for every $A \geq 0$ there exists a finite-index subgroup $L \leqslant_f G$, with $S \subseteq L$, such that properties* (P1)–(P3) *from Section 3.1 hold for arbitrary subgroups $Q' \leqslant Q \cap L$ and $R' \leqslant R \cap L$ satisfying $Q' \cap R' = S$.*

Proof By combining the assumptions with Lemma 5.24, we know that maximal parabolic subgroups of G are finitely generated abelian groups. Since such groups are slender, all relatively quasiconvex subgroups of G are finitely generated (see [30, Corollary 9.2]). Moreover, finitely generated abelian groups are LERF, and hence, they are double coset separable (because the product of two subgroups is again a subgroup). Therefore the double coset PS is separable in G for any maximal parabolic subgroup $P \leqslant G$ by Proposition 13.5.

In view of Proposition 11.1, for any finite collection $\mathcal{P}$, of maximal parabolic subgroups of G, and any $B, C \geq 0$ there exists $L \leqslant_f G$, with $S \subseteq L$, such that any subgroups $Q' \leqslant Q \cap L$ and $R' \leqslant R \cap L$ satisfy conditions (C1)–(C3), as long as $Q' \cap R' = S$. Remark 3.3 tells us that these subgroups automatically satisfy conditions (C4) and (C5). Thus we can obtain the desired statement by applying Theorem 3.5. □

Corollary 14.2 *Suppose that G is a QCERF group generated by a finite subset X and hyperbolic relative to a finite family $\{H_\nu \mid \nu \in \mathcal{N}\}$ of virtually abelian subgroups. Let $Q, R \leqslant G$ be relatively quasiconvex subgroups and let $S = Q \cap R$. Then there exists $L \leqslant_f G$ such that if $Q' \leqslant Q \cap L$ and $R' \leqslant R \cap L$ are relatively quasiconvex subgroups of G satisfying $Q' \cap R' = S \cap L$ then the subgroup $\langle Q', R' \rangle$ is also relatively quasiconvex in G.*

Proof By the assumptions for each $\nu \in \mathcal{N}$ there exists a finite-index abelian subgroup $K_\nu \leqslant_f H_\nu$. Since G is QCERF, each K_ν is separable in G (it is finitely generated by Lemma 5.24 and it is relatively quasiconvex by Corollary 5.23). Thus, in view of Lemma 4.17, for every $\nu \in \mathcal{N}$ there exists $L_\nu \leqslant_f G$ such that $L_\nu \cap H_\nu = K_\nu$.

Since $|\mathcal{N}| < \infty$, the intersection $\bigcap_{\nu \in \mathcal{N}} L_\nu$ has finite index in G, hence it contains a finite-index normal subgroup $G_1 \lhd_f G$. Note that for any $g \in G$ and any $\nu \in \mathcal{N}$ we have

$$G_1 \cap gH_\nu g^{-1} = g(G_1 \cap H_\nu)g^{-1} \subseteq g(L_\nu \cap H_\nu)g^{-1} = gK_\nu g^{-1}, \tag{14-1}$$

where the first equality follows from the normality of G_1, the middle inclusion follows from the fact that $G_1 \subseteq L_\nu$, and the last equality is due to the fact that $L_\nu \cap H_\nu = K_\nu$. By Lemma 5.22, G_1 is finitely generated and relatively quasiconvex in G; hence, by [30, Theorem 9.1] it is hyperbolic relative to representatives of G_1-conjugacy classes of the intersections $G_1 \cap gH_\nu g^{-1}$, $g \in G$. Thus, in view of (14-1), all peripheral subgroups in G_1 are abelian.

By [30, Corollary 9.3], a subgroup of G_1 is relatively quasiconvex in G_1 (with respect to the above family of peripheral subgroups) if and only if it is relatively quasiconvex in G. Therefore G_1 is QCERF and $Q_1 = Q \cap G_1 \leqslant_f Q$, $R_1 = R \cap G_1 \leqslant_f R$ are finitely generated relatively quasiconvex subgroups of G_1 by Lemma 5.22. After denoting $S_1 = S \cap G_1 = Q_1 \cap R_1$, we can apply Theorem 1.3 to find a finite-index subgroup $L \leqslant_f G_1$ such that $S_1 \subseteq L$ (thus, $S_1 = S \cap L$) and the subgroup $\langle Q', R' \rangle$ is relatively quasiconvex in G_1, for arbitrary $Q' \leqslant Q_1 \cap L = Q \cap L$ and $R' \leqslant R_1 \cap L = R \cap L$ satisfying $Q' \cap R' = Q_1 \cap R_1 = S_1$. We can use [30, Corollary 9.3] again to deduce that $\langle Q', R' \rangle$ is relatively quasiconvex in G. □

The following collects the results of the previous sections, allowing us to find subgroups Q' and R' to which Theorem 3.5 can be applied.

Proposition 14.3 *Let G be a finitely generated QCERF relatively hyperbolic group with double coset separable peripheral subgroups, and let Q and R be finitely generated relatively quasiconvex subgroups. Then for any $B, C \geq 0$, and finite family $\mathcal{P}$ of maximal parabolic subgroups of G, there are finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ satisfying* (C1)–(C5) *with constants B and C and family $\mathcal{P}$.*

More precisely, writing $S = Q \cap R$, there exists $L \leqslant_f G$ with $S \subseteq L$ such that for any $L' \leqslant_f L$ satisfying $S \subseteq L'$, we can choose $Q' = Q \cap L' \leqslant_f Q$ and there exists $M \leqslant_f L'$ with $Q' \subseteq M$ such that for any $M' \leqslant_f M$ satisfying $Q' \subseteq M'$, we can choose $R' = R \cap M' \leqslant_f R$.

Proof We check that all the assumptions of Theorem 11.3 are satisfied for every $P \in \mathcal{P}$. Indeed, (S1) holds because G is QCERF and (S3) is true by Corollary 13.4.

Note that the subgroups $D = Q \cap P$ and $R \cap P$ are finitely generated by Lemma 5.24, hence condition (13-1) follows from the double coset separability of P; thus (S4) is satisfied. Finally, (S2) holds by Proposition 13.5.

The statement now follows by applying Theorem 11.3. □

Theorem 14.4 *Let G be a group generated by a finite set X and hyperbolic relative to a finite collection of subgroups $\{H_\nu \mid \nu \in \mathcal{N}\}$. Suppose that G is QCERF and H_ν is double coset separable, for each $\nu \in \mathcal{N}$.*

If $Q, R \leqslant G$ are finitely generated relatively quasiconvex subgroups and $S = Q \cap R$ then for every $A \geq 0$ there exist finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ which satisfy properties (P1)–(P3).

More precisely, there exists $L \leqslant_f G$ with $S \subseteq L$ such that for any $L' \leqslant_f L$ satisfying $S \subseteq L'$, we can choose $Q' = Q \cap L' \leqslant_f Q$ and there exists $M \leqslant_f L'$ with $Q' \subseteq M$ such that for any $M' \leqslant_f M$ satisfying $Q' \subseteq M'$, we can choose $R' = R \cap M' \leqslant_f R$.

Proof Let $\mathcal{P}$ be the finite collection of maximal parabolic subgroups of G provided by Theorem 3.5. The statement follows immediately from a combination of Theorem 3.5 with Proposition 14.3. □

Recall that Q and R are said to have almost compatible parabolics if for every maximal parabolic subgroup $P \leqslant G$, either $Q \cap P \preccurlyeq R \cap P$ or $R \cap P \preccurlyeq Q \cap P$. We find that in the case when Q and R have almost compatible parabolics, it is actually not necessary to assume that the peripheral subgroups are double coset separable:

Theorem 14.5 *Suppose that G is a finitely generated QCERF relatively hyperbolic group, $Q, R \leqslant G$ are finitely generated relatively quasiconvex subgroups with almost compatible parabolics and $S = Q \cap R$. Then for every $A \geq 0$ there exist finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ which satisfy properties* (P1)–(P3).

More precisely, there exists $L \leqslant_f G$, with $S \subseteq L$, such that for any $L' \leqslant_f L$, satisfying $S \subseteq L'$, we can choose $Q' = Q \cap L' \leqslant_f Q$ and there exists $M \leqslant_f L'$, with $Q' \subseteq M$, such that for any $M' \leqslant_f M$, satisfying $Q' \subseteq M'$, we can choose $R' = R \cap M' \leqslant_f R$.

Proof As before, we will be verifying the assumptions of Theorem 11.3. Let P be an arbitrary maximal parabolic subgroup of G. Condition (S1) follows from the QCERF-ness of G and (S3) follows from Corollary 13.4.

Let $D = Q \cap P$ and $U \leqslant_f D$. Since Q and R have almost compatible parabolics and $Q \cap P \preccurlyeq U$, we know that either $U \preccurlyeq R \cap P$ or $R \cap P \preccurlyeq U$. Note that both U and $R \cap P$ are finitely generated by Lemma 5.24 and relatively quasiconvex by Corollary 5.23, so they are separable because G is QCERF. Lemma 4.15 now implies that the double coset $U(R \cap P)$ is separable in G, thus condition (13-1) is satisfied by Lemma 4.13. This shows that (S4) of Theorem 11.3 is satisfied; furthermore, (S2) holds by Proposition 13.5.

We can now deduce the theorem by combining Theorem 3.5 with Theorem 11.3. □

15 Separability of double cosets in QCERF relatively hyperbolic groups

In this section we prove Corollary 1.4 from the introduction.

Proof of Corollary 1.4 Let X be a finite generating set of G. Consider any $g \in G \setminus QR$, and set $A = |g|_X + 1$. By Theorem 14.4 there are subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ satisfying properties (P1) and (P3). The latter property, combined with the definition of A, implies that $g \notin Q\langle Q', R'\rangle R$.

On the other hand, property (P1) tells us that $H = \langle Q', R' \rangle$ is relatively quasiconvex in G. Clearly it is also finitely generated, hence it must be separable in G by QCERF-ness. Observe that since Q' and R' are finite-index subgroups in Q and R respectively,

$$QHR = \bigcup_{i=1}^{n} \bigcup_{j=1}^{m} a_i H b_j,$$

where $a_1, \dots, a_n$ are left coset representatives of Q' in Q, and $b_1, \dots, b_m$ are right coset representatives of R' in R. Recalling Remark 4.12, we see that the subset QHR is separable in G; thus it is a closed set containing QR but not containing g. Since we found such a set for an arbitrary $g \in G \setminus QR$, we can conclude that QR is closed in $\mathcal{PT}(G)$, as required. □

Corollary 1.6 from the introduction can be proved in the same way as Corollary 1.4, except that one needs to use Theorem 14.5 instead of Theorem 14.4.

Part III Separability of products of subgroups

This part of the paper is dedicated to proving Theorem 1.8 from the introduction. In order to do this we must generalise the discussion of path representatives in Sections 6–8, adapting the arguments there to deal with additional technicalities. Let us give a summary of the argument.

Let G be a QCERF finitely generated relatively hyperbolic group with a finite collection of peripheral subgroups $\{H_\nu \mid \nu \in \mathcal{N}\}$. Suppose that, for each $\nu \in \mathcal{N}$, the subgroup H_ν has property RZ_s. Let $F_1, \dots, F_s \leqslant G$ be finitely generated relatively quasiconvex subgroups. In order to show that the product $F_1 \cdots F_s$ is separable, we proceed by induction on s. The case that $s = 1$ is the QCERF condition and $s = 2$ is Corollary 1.4, so we may assume $s > 2$. For ease of reading we now relabel the subgroups $F_1 = Q$, $F_2 = R$, $F_3 = T_1, \dots, F_s = T_m$, where $m = s - 2 > 0$.

We approximate the product $QRT_1 \cdots T_m$ with sets of the form $Q\langle Q', R' \rangle RT_1 \cdots T_m$, where $Q' \leqslant_f Q$ and $R' \leqslant_f R$ are finite-index subgroups of Q and R respectively. Observe that we can write these sets as finite unions

$$Q\langle Q', R' \rangle RT_1 \cdots T_m = \bigcup_{i,j} a_i \langle Q', R' \rangle b_j T_1 \cdots T_m, \tag{15-1}$$

where the elements a_i and b_j are coset representatives of Q' and R' in Q and R respectively. Note that the products on the right-hand side of (15-1) now involve only $s - 1$ subgroups. By Theorem 1.2, the subgroups Q' and R' can be chosen so that $\langle Q', R' \rangle$ is relatively quasiconvex, hence we can apply the induction hypothesis to show that such products are separable in G.

It then remains to prove that the product $QRT_1 \cdots T_m$ is, in fact, an intersection of subsets of the form $Q\langle Q', R' \rangle RT_1 \cdots T_m$ as above. To this end, we study path representatives $qp_1 \cdots p_n r t_1 \cdots t_m$ of elements of $Q\langle Q', R'\rangle RT_1\cdots T_m$ in a similar manner to Part II. The main additional difficulty comes from controlling instances of multiple backtracking that involve segments in the $t_1\cdots t_m$ part of the path. We introduce new metric conditions (C2-m) and (C5-m) to deal with these technicalities.

16 Auxiliary definitions

Convention 16.1 We write G for a group generated by a finite set X and hyperbolic relative to a family of subgroups $\{H_\nu \mid \nu \in \mathcal{N}\}$, $|\mathcal{N}| < \infty$. Let $\mathcal{H} = \bigsqcup_{\nu\in\mathcal{N}}(H_\nu \setminus \{1\})$ and choose $\delta \in \mathbb{N}$ so that the Cayley graph $\Gamma(G, X \cup \mathcal{H})$ is δ-hyperbolic (see Lemma 5.4).

We will assume that $Q, R, T_1, \dots, T_m \leqslant G$ are fixed relatively quasiconvex subgroups of G, with quasiconvexity constant $\varepsilon \geq 0$, where $m \in \mathbb{N}_0$. Denote $S = Q \cap R$.

Throughout this section we use Q' and R' to denote subgroups of Q and R respectively. We will also assume that $Q' \cap R' = Q \cap R = S$ (that is, Q' and R' satisfy (C1)).

16.1 New metric conditions

Suppose $B, C \geq 0$ are some constants, $\mathcal{P}$ is a finite collection of maximal parabolic subgroups of G, and $\mathcal{U}$ is a finite family of finitely generated relatively quasiconvex subgroups of G. We will be interested in the following generalisations of conditions (C2) and (C5) to the multiple coset setting:

(C2-m) $\min_X(R\langle Q', R'\rangle RT_1\cdots T_j \setminus RT_1\cdots T_j) \geq B$, for each $j = 0, \dots, m$;

(C5-m) $\min_X\big(q\langle Q'_P, R'_P\rangle R_P(U_1)_P\cdots(U_j)_P \setminus qQ'_PR_P(U_1)_P\cdots(U_j)_P\big) \geq C$, for each $P \in \mathcal{P}$, all $q \in Q_P$, any $j \in \{0, \dots, m\}$ and arbitrary $U_1, \dots, U_j \in \mathcal{U}$, where $(U_i)_P = U_i \cap P \leqslant P$.

Remark 16.2 Let us make the following observations.

- When $j = 0$, the inequality from condition (C2-m) reduces to $\min_X(R\langle Q', R'\rangle R \setminus R) \geq B$, which is a part of (C2); on the other hand, the inequality from condition (C5-m) simply becomes (C5). In particular, for each $m \geq 0$, (C5-m) implies (C5).
- In our usage of (C5-m), the set $\mathcal{U}$ will consists of finitely many conjugates of $T_1, \dots, T_m$; in fact, $U_i = T_i^{a_i}$, for some $a_i \in G$, $i = 1, \dots, m$.

Remark 16.3 Similarly to conditions (C1)–(C5), the above conditions are best understood with a view towards the profinite topology.

- To prove separability of products of relatively quasiconvex subgroups we argue by induction on the number of factors. That is, we assume that the product of $m+1$ relatively quasiconvex subgroups is separable and then deduce the separability of the product of $m+2$ relatively quasiconvex subgroups.

The existence of finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ realising condition (C2-m) will be deduced from this inductive assumption.

- The existence of finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ realising condition (C5-m), given a finite family $\mathcal{U}$, will be deduced from the assumption that the peripheral subgroups $\{H_\nu \mid \nu \in \mathcal{N}\}$ of G each satisfy the property RZ_{m+2}.

16.2 Path representatives for products of subgroups

In this subsection we define path representatives for elements of $Q\langle Q', R'\rangle RT_1 \cdots T_m$ similarly to the path representatives for elements of $Q\langle Q', R'\rangle R$ from Definition 10.6 and discuss their properties.

Definition 16.4 (path representative, III) Let g be an element of $Q\langle Q', R'\rangle RT_1 \cdots T_m$. Suppose that $p = qp_1 \cdots p_n r t_1 \cdots t_m$ is a broken line in $\Gamma(G, X \cup \mathcal{H})$ satisfying the following properties:

- $\tilde{p} = g$;
- $\tilde{q} \in Q$ and $\tilde{r} \in R$;
- $\tilde{p}_i \in Q' \cup R'$ for each $i \in \{1, \dots, n\}$;
- $\tilde{t}_i \in T_i$ for each $i \in \{1, \dots, m\}$.

Then we say that p is a *path representative* of g in the product $Q\langle Q', R'\rangle RT_1 \cdots T_m$.

The type of a path representative is defined as before (cf Definitions 6.3 and 10.7).

Definition 16.5 (type and width of a path representative, III) Let $g \in Q\langle Q', R'\rangle RT_1 \cdots T_m$ and let $p = qp_1 \cdots p_n r t_1 \cdots t_m$ be a path representative of g in the sense of Definition 16.4. Denote by Y the set of all $\mathcal{H}$-components of the segments of p. We define the *width* of p as the integer n and the *type* of p as the triple

$$\tau(p) = \left(n, \ell(p), \sum_{y \in Y} |y|_X\right) \in \mathbb{N}_0{}^3.$$

The following observation will be useful.

Remark 16.6 Suppose $g \in Q\langle Q', R'\rangle RT_1 \cdots T_m$ can be written as a product

$$g = xy_1 \cdots y_n z u_1 \cdots u_m,$$

where $x \in Q$, $y_1, \dots, y_n \in Q' \cup R'$, $z \in R$ and $u_i \in T_i$, for each $i = 1, \dots, m$. Then g has a path representative of width n.

Similarly to path representatives of elements of $\langle Q', R'\rangle$ (in the sense defined in Section 6), we will be interested in path representatives whose type is minimal (as an element of $\mathbb{N}_0{}^3$ under the lexicographic ordering). Given an element $g \in Q\langle Q', R'\rangle RT_1 \cdots T_m$, such a path representative is always guaranteed to exist. Let us make the following observation (cf Remark 10.8).

Remark 16.7 Suppose that $p = qp_1 \cdots p_n r t_1 \cdots t_m$ is a minimal type path representative of an element $g \in Q\langle Q', R'\rangle RT_1 \cdots T_m$ such that $g \notin QRT_1 \cdots T_m$. Then $n > 0$, $\tilde{p}_1 \in R' \setminus S$, $\tilde{p}_n \in Q' \setminus S$ and the labels of $p_1, \dots, p_n$ alternate between representing elements of $R' \setminus S$ and $Q' \setminus S$. In particular, the integer n must be even.

Note that in Definition 16.4 the geodesic paths q, r and $t_1, \dots, t_m$ are always counted as segments of the path p, even if they end up being trivial paths. For example a minimal type path representative of an element $g \in R'Q'T_1 \cdots T_m \setminus QRT_1 \cdots T_m$ will be a broken line $p = qp_1p_2rt_1 \cdots t_m$ with $m+4$ segments, where q and r are trivial paths.

The proofs of the main results from Sections 6 and 7 can be easily adapted to apply to minimal type path representatives of elements $g \in Q\langle Q', R'\rangle RT_1 \cdots T_m \setminus QRT_1 \cdots T_m$ (in the sense of Definitions 16.4 and 16.5), with only superficial differences, so the proofs of the following generalisations of Lemmas 6.7, 7.3 and 7.6, respectively, will be omitted.

Lemma 16.8 *There is a constant $C_0 \geq 0$ such that the following holds.*

Assume that $Q' \leqslant Q$ and $R' \leqslant R$ are subgroups satisfying condition (C1). *Consider any element $g \in Q\langle Q', R'\rangle RT_1 \cdots T_m$ with $g \notin QRT_1 \cdots T_m$. Let $p = qp_1 \cdots p_n r t_1 \cdots t_m$ be a path representative of g of minimal type, with nodes $f_0, \dots, f_{n+m+2}$ (that is, $f_0 = q_-$, $f_i = (p_i)_-$, for each $i \in \{1, \dots, n\}$, $f_{n+1} = r_-$, $f_{n+1+j} = (t_j)_-$, for each $j \in \{1, \dots, m\}$, and $f_{n+m+2} = (t_m)_+$). Then $\langle f_{i-1}, f_{i+1}\rangle^{\mathrm{rel}}_{f_i} \leq C_0$, for all $i \in \{1, \dots, n+m+1\}$.*

Lemma 16.9 *There is a constant $C_1 \geq 0$ such that the following is true.*

Let $Q' \leqslant Q$ and $R' \leqslant R$ be subgroups satisfying condition (C1). *Consider a minimal type path representative $p = qp_1 \cdots p_n r t_1 \cdots t_m$ for an element $g \in Q\langle Q', R'\rangle RT_1 \cdots T_m \setminus QRT_1 \cdots T_m$. If a and b are adjacent segments of p, with $a_+ = b_-$, and h and k are connected $\mathcal{H}$-components of a and b respectively, then $d_X(h_+, a_+) \leq C_1$ and $d_X(a_+, k_-) \leq C_1$.*

Lemma 16.10 *For any $\zeta \geq 0$ there is $\Theta_0 = \Theta_0(\zeta) \in \mathbb{N}$ such that the following is true.*

Let $Q' \leqslant Q$ and $R' \leqslant R$ be subgroups satisfying condition (C1). *Consider a minimal type path representative $p = qp_1 \cdots p_n r t_1 \cdots t_m$ for an element $g \in Q\langle Q', R'\rangle RT_1 \cdots T_m \setminus QRT_1 \cdots T_m$. Suppose that a and b are adjacent segments of p, with $a_+ = b_-$, and h and k are connected $\mathcal{H}$-components of a and b respectively, such that*

$$\max\{|h|_X, |k|_X\} \geq \Theta_0.$$

Then $d_X(h_-, k_+) \geq \zeta$.

17 Multiple backtracking in product path representatives: two special cases

Just like in Theorem 3.5, the main difficulty in proving Theorem 1.8 consists in dealing with multiple backtracking in path representatives. In this section we will consider two of the possible cases. We will be working under Convention 16.1.

Throughout the rest of the paper we fix the following notation.

Notation 17.1 let C_1 be the larger of the two constants provided by Lemmas 7.3 and 16.9, and denote by $\mathcal{P}_1$ the finite collection of maximal parabolic subgroups of G given by

$$\mathcal{P}_1 = \{ {H_\nu}^b \mid \nu \in \mathcal{N}, |b|_X \le C_1 \}.$$

The following lemma is roughly analogous to Lemma 8.2.

Lemma 17.2 *For any $L \ge 0$ and any relatively quasiconvex subgroup $T \leqslant G$ there is a constant $L' = L'(L, T) \ge 0$ such that the following is true.*

Let $P = {H_\nu}^b \in \mathcal{P}_1$, for some $\nu \in \mathcal{N}$ and $b \in G$, with $|b|_X \le C_1$, and let t be a geodesic path in $\Gamma(G, X \cup \mathcal{H})$, with $\tilde{t} \in T$. Suppose that $v \in Pb = bH_\nu$ is a vertex of t and $u \in P$ is an element satisfying $d_X(u, t_-) \le L$. Denote $a = u^{-1} t_- \in G$. Then there is a geodesic path t' in $\Gamma(G, X \cup \mathcal{H})$ such that

- $t'_- = u$ *and* $d_X(t'_+, v) \le L'$;
- $\tilde{t}' \in T^a \cap P$;
- $(t'_+)^{-1} t_+ \in aT$.

Proof Let $K = \max\{C_1, \sigma + L\}$, where $\sigma \ge 0$ is a quasiconvexity constant for T. Denote

$$L' = \max\{ K'(P, T^a, K) \mid P \in \mathcal{P}_1,\ a \in G,\ |a|_X \le L \}, \tag{17-1}$$

where $K'(P, T^a, K)$ is obtained from Lemma 4.1.

The hypotheses that $v \in Pb$ and $|b|_X \le C_1$ imply that $d_X(v, P) \le |b|_X \le C_1$. As $u \in P$, we have $P = uP$ and so

$$d_X(v, uP) \le C_1. \tag{17-2}$$

Set $x = t_- = ua$. Since $\tilde{t} \in T$, we have $d_X(v, xT) \le \sigma$, as T is σ-quasiconvex. Hence

$$d_X(v, uT^a) = d_X(v, xTa^{-1}) \le d_X(v, xT) + |a|_X \le \sigma + L.$$

Combining the latter inequality with (17-2) allows us to apply Lemma 4.1 to find an element $z \in u(T^a \cap P)$ such that $d_X(v, z) \le L'$, where $L' \ge 0$ is the constant from (17-1). Now take t' to be any geodesic in

$\Gamma(G, X \cup \mathcal{H})$ with $t'_- = u$ and $t'_+ = z$. It is straightforward to verify that t' satisfies the first two of the required properties. For the last property, observe that

$$(t'_+)^{-1}t_+ = ((t'_+)^{-1}u)(u^{-1}t_-)(t_-^{-1}t_+) = \tilde{t}'^{-1}a\tilde{t} \in T^a aT = aT. \qquad \square$$

The following notation will be fixed for the remainder of the paper.

Notation 17.3 Let D be the constant from Lemma 8.2, corresponding to C_1 and $\mathcal{P}_1$ (from Notation 17.1) and subgroups Q, R. We define constants $L_1, \dots, L_{m+1}$ as

$$L_1 = D + C_1 \quad \text{and} \quad L_{i+1} = L'(L_i, T_i) + C_1 \quad \text{for each } i = 1, \dots, m,$$

where L' is obtained from Lemma 17.2.

We also define the family of subgroups

$$\mathcal{U}_1 = \bigcup_{i=1}^{m} \{T_i^g \mid i \in \{1, \dots, m\},\ g \in G,\ |g|_X \le L_i\},$$

consisting of finitely many conjugates of the subgroups $T_1, \dots, T_m$. Note that, by Lemma 5.22, each $U \in \mathcal{U}_1$ is a relatively quasiconvex subgroup of G.

The next proposition describes how we deal with consecutive backtracking that involves the $(t_1 \cdots t_m)$-part of a path representative of an element $g \in Q\langle Q', R'\rangle RT_1 \cdots T_m \setminus QRT_1 \cdots T_m$; it complements Proposition 8.4 which takes care of backtracking within the $qp_1 \cdots p_n r$-part.

Proposition 17.4 *Suppose that $p = qp_1 \cdots p_n r t_1 \cdots t_m$ is a path representative of minimal type for an element $g \in Q\langle Q', R'\rangle RT_1 \cdots T_m \setminus QRT_1 \cdots T_m$, where $Q' \leqslant Q$ and $R' \leqslant R$ are some subgroups satisfying* (C1). *Let $P = H_\nu{}^b \in \mathcal{P}_1$, for some $\nu \in \mathcal{N}$ and $b \in G$, with $|b|_X \le C_1$.*

Suppose that $h_1, \dots, h_j$ are connected H_ν-components of the segments $t_1, \dots, t_j$, respectively, with $j \in \{1, \dots, m\}$, such that $(h_1)_- \in Pb = bH_\nu$. If $u_1 \in P$ is an element satisfying $d_X(u_1, (t_1)_-) \le L_1$ then there exist elements $a_1, \dots, a_j \in G$ and a broken line $t'_1 \cdots t'_j$ in $\Gamma(G, X \cup \mathcal{H})$ such that

- (i) $(t'_1)_- = u_1$ *and* $d_X((t'_j)_+, (h_j)_+) \le L_{j+1}$;
- (ii) $a_{i+1} \in a_i T_i$, *for* $i = 1, \dots, j-1$;
- (iii) $a_i = (t'_i)_-^{-1}(t_i)_-$ *and* $|a_i|_X \le L_i$, *for each* $i = 1, \dots, j$;
- (iv) $\tilde{t}'_i \in T_i^{a_i} \cap P$, *for all* $i = 1, \dots, j$.

Proof We start by setting $a_1 = u_1^{-1}(t_1)_-$, so that $|a_1|_X = d_X(u_1, (t_1)_-) \le L_1$. Note that

$$(h_1)_+ = (h_1)_- \tilde{h}_1 \in bH_\nu = Pb.$$

Therefore we can apply Lemma 17.2 to find a geodesic path t'_1 in $\Gamma(G, X \cup \mathcal{H})$ such that $(t'_1)_- = u_1$, $d_X((t'_1)_+, (h_1)_+) \le L'(L_1, T_1)$, $\tilde{t}'_1 \in T_1^{a_1} \cap P$ and

$$(t'_1)_+^{-1}(t_1)_+ \in a_1 T_1. \tag{17-3}$$

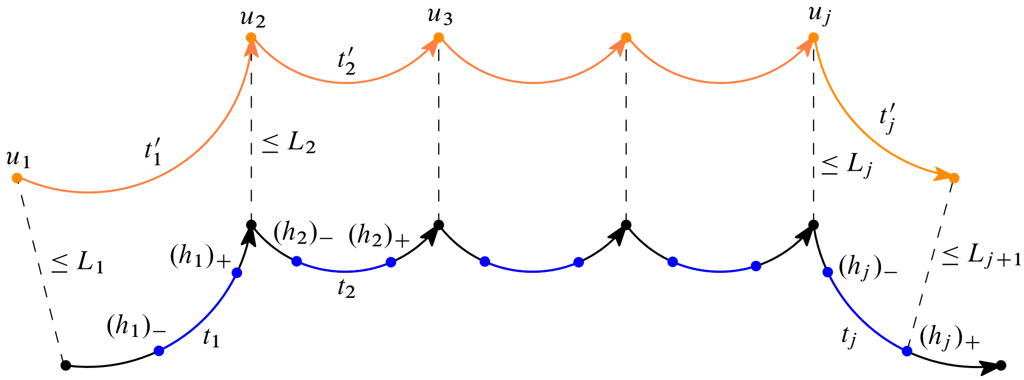


Figure 8: The new path $t'_1 \cdots t'_j$ constructed in Proposition 17.4.

It follows that properties (ii)–(iv) are satisfied for $i = 1$, while property (i) holds because $L_2 \geq L'(L_1, T_1)$ by definition. If $j = 1$ then property (ii) is vacuously true.

We can now suppose that $j > 1$. Then h_1 is connected to the component h_2 of t_2, so, according to Lemma 16.9, $d_X((h_1)_+, (t_1)_+) \leq C_1$. Set $u_2 = (t'_1)_+$ and $a_2 = u_2^{-1}(t_1)_+$. Note that $a_2 \in a_1 T_1$ by (17-3) and

$$|a_2|_X = d_X((t_1)'_+, (t_1)_+) \leq d_X((t'_1)_+, (h_1)_+) + d_X((h_1)_+, (t_1)_+) \leq L'(L_1, T_1) + C_1 = L_2.$$

Since $(t_2)_- = (t_1)_+$, we see that $a_2 = u_2^{-1}(t_2)_-$ and $d_X(u_2, (t_2)_-) = |a_2|_X \leq L_2$.

Now, observe that $u_2 = u_1 \tilde{t}'_1 \in P$ and $(h_2)_+ \in bH_\nu = Pb$, as h_2 is connected to h_1. This allows us to use Lemma 17.2 to find a geodesic path t'_2 in $\Gamma(G, X \cup \mathcal{H})$ such that $(t'_2)_- = u_2 = (t'_1)_+$, $d_X((t'_2)_+, (h_2)_+) \leq L'(L_2, T_2)$, $\tilde{t}'_2 \in T_2^{a_2} \cap P$ and $(t'_2)_+^{-1} t_+ \in a_2 T_2$ (see Figure 8).

If $j = 2$ then we are done, otherwise we construct the remaining elements $a_3, \ldots, a_j$ and the paths $t'_3, \ldots, t'_j$ inductively, similarly to the construction of a_2 and t'_2 above. □

The next two propositions prove that, under certain conditions, instances of multiple backtracking are long. Essentially, they generalise Proposition 8.5. The first of these shows how we can use condition (C5-m) to deal with particular instances of multiple backtracking.

Proposition 17.5 *For each $\zeta \geq 0$ there is a constant $C_2 = C_2(\zeta) \geq 0$ such that if $Q' \leqslant Q$ and $R' \leqslant R$ satisfy conditions* (C1), (C3) *and* (C5-m) *with constant $C \geq C_2$ and finite families $\mathcal{P}$ and $\mathcal{U}$, such that $\mathcal{P}_1 \subseteq \mathcal{P}$ and $\mathcal{U}_1 \subseteq \mathcal{U}$, then the following is true.*

Let $p = q p_1 \cdots p_n r t_1 \cdots t_m$ be a minimal type path representative for some $g \in Q\langle Q', R'\rangle R T_1 \cdots T_m$, with $g \notin Q R T_1 \cdots T_m$. Suppose that p has multiple backtracking along H_ν-components $h_1, \ldots, h_k$ of its segments, for some $\nu \in \mathcal{N}$, such that

- *h_1 is an H_ν-component of either q or p_i, for some $i \in \{1, \ldots, n-1\}$, with $\tilde{p}_i \in Q'$;*

- *h_k is an H_ν-component of a segment t_j, for some $j \in \{1, \dots, m\}$.*

Then $d_X((h_1)_-, (h_k)_+) \geq \zeta$.

Proof Take
$$C_2 = \max\{2C_1, D + \zeta + L_j \mid j = 1, \dots, m+1\} + 1,$$
where D and L_j are defined in Notation 17.3, and suppose that $C \geq C_2$.

The proof employs the same strategy as Proposition 8.5: we first construct a path whose endpoints are close to $(h_1)_-$ and $(h_k)_+$ and whose label represents an element of a parabolic subgroup. We will then obtain a contradiction with the minimality of the type of p, using condition (C5-m).

We will focus on the case when h_1 is an H_ν-component of p_i, for some $i \in \{1, \dots, n-1\}$ with $\tilde{p}_i \in Q'$, with the case when h_1 is an H_ν-component of q being similar. Note that since $g \notin QRT_1 \cdots T_m$, it must be that $n \geq 2$ by Remark 16.7. After translating by $(p_i)_+^{-1}$, we may assume that $(p_i)_+ = 1$. We write $b = (h_1)_+$ and note that, according to Lemma 16.9,

$$|b|_X = d_X((h_1)_+, (p_i)_+) \leq C_1. \tag{17-4}$$

Let $P = bH_\nu b^{-1} \in \mathcal{P}_1 \subseteq \mathcal{P}$. Since $h_1, \dots, h_k$ are pairwise connected, the vertices $(h_l)_+$ lie in the same left coset bH_ν, for all $l = 1, \dots, k$, thus

$$(h_l)_+ \in Pb \quad \text{for all } l = 1, \dots, k. \tag{17-5}$$

We construct a new broken line $p' = p'_i \cdots p'_n r' t'_1 \cdots t'_j$ in two steps. It will be used in conjunction with condition (C5-m) to obtain a path representative of g with lesser type than p.

Step 1 We start by constructing geodesic paths $p'_i, p'_{i+1}, \dots, p'_n$ and r' by using condition (C3) and applying Lemmas 8.2 and 8.3, in exactly the same way as in the proof of Proposition 8.4. The newly constructed paths will have the following properties:

- $\tilde{p}'_i \in Q_P$, $\tilde{p}'_l \in Q'_P \cup R'_P$, for each $l = i+1, \dots, n$, and $\tilde{r}' \in R_P$;
- $d_X((p'_i)_-, (h_1)_-) \leq D$ and $(p'_i)_+ = (p_i)_+ = 1$;
- $(p'_l)_+ = (p'_{l+1})_-$, for $l = i, \dots, n-1$;
- $r'_- = (p'_n)_+$ and $d_X(r'_+, (h_{k-j})_+) \leq D$;
- $(p'_l)_+^{-1}(p_l)_+ \in S$, for $l = i+1, \dots, n$.

Step 2 We now construct geodesic paths $t'_1, \dots, t'_j$ as follows. Set $u_1 = (r')_+$ and observe that since $(p'_{i+1})_- = (p'_i)_+ = 1$, we have
$$u_1 = \tilde{p}'_{i+1} \cdots \tilde{p}'_n \tilde{r}' \in P.$$

By Lemma 16.9, we have $d_X((h_{k-j})_+, (t_1)_-) = d_X((h_{k-j})_+, r_+) \leq C_1$. Moreover, by Step 1 above, $d_X(u_1, (h_{k-j})_+) \leq D$. Therefore
$$d_X(u_1, (t_1)_-) \leq C_1 + D = L_1.$$

Together with (17-5) this allows us to apply Proposition 17.4 to find elements $a_1,\dots,a_j \in G$ and a broken line $t_1't_2'\cdots t_j'$ in $\Gamma(G, X\cup\mathcal{H})$ such that

- $(t_1')_- = u_1$ and $d_X((t_j')_+,(h_k)_+) \le L_{j+1}$;
- $a_{l+1} \in a_l T_l$, for $l = 1,\dots,j-1$;
- $a_l = (t_l')_-^{-1}(t_l)_-$ and $|a_l|_X \le L_l$, for each $l = 1,\dots,j$;
- $\tilde{t}_l' \in T_l^{a_l} \cap P$, for all $l = 1,\dots,j$.

Observe that

(17-6) $a_1 = (t_1')_-^{-1}(t_1)_- = u_1^{-1}r_+ = (r_+'^{-1}r_-')(r_-'^{-1}r_-)(r_-^{-1}r_+) = \tilde{r}'^{-1}(p_n')_+^{-1}(p_n)_+\tilde{r} \in R_P S R \subseteq R.$

We now define a new broken line p' in $\Gamma(G, X\cup\mathcal{H})$ by

$$p' = p_i'\cdots p_n' r' t_1'\cdots t_j'.$$

Note that $d_X(p_-',(h_1)_-) \le D$, $d_X(p_+',(h_k)_+) \le L_{j+1}$ and $\tilde{p}' \in \tilde{p}_i'\langle Q_P', R_P'\rangle R_P (T_1^{a_1})_P\cdots(T_j^{a_j})_P$, where $\tilde{p}_i' \in Q_P$. Moreover, $T_l^{a_l} \in \mathcal{U}_1 \subseteq \mathcal{U}$, for each $l = 1,\dots,j$.

Now, suppose, for a contradiction, that $d_X((h_1)_-,(h_k)_+) < \zeta$. Then, by the triangle inequality,

$$|p'|_X \le D + \zeta + L_{j+1} < C_2.$$

Thus, as $C \ge C_2$, we can apply (C5-m) to deduce that $\tilde{p}' \in \tilde{p}_i' Q_P' R_P (T_1^{a_1})_P\cdots(T_j^{a_j})_P$. Therefore, there exist elements $z \in \tilde{p}_i' Q_P'$, $x \in R$ and $y_l \in T_l$, $l = 1,\dots,j$, such that $\tilde{p}' = zxy_1^{a_1}\cdots y_j^{a_j}$. By construction, for each $l = 1,\dots,j-1$ there is $b_l \in T_l$ such that $a_{l+1} = a_l b_l$, and so $a_l^{-1}a_{l+1} = b_l \in T_l$. Recalling that $(p_i')_+ = (p_i)_+ = 1$, the above yields

$$\tilde{p}' = zxy_1^{a_1}\cdots y_j^{a_j} = zxa_1y_1b_1y_2b_2\cdots b_{j-1}y_ja_j^{-1}. \tag{17-7}$$

Let α and β be geodesic segments in $\Gamma(G, X\cup\mathcal{H})$ connecting $(p_i)_-$ with $(p_i')_-$ and $(t_j')_+$ with $(t_j)_+$ respectively. Since $(p_i)_+ = (p_i')_+$, we have

$$\tilde{\alpha} = (p_i)_-^{-1}(p_i')_- = (p_i)_-^{-1}(p_i)_+(p_i')_+^{-1}(p_i')_- = \tilde{p}_i\tilde{p}_i'^{-1}. \tag{17-8}$$

On the other hand, it follows from the construction that

$$\tilde{\beta} = (t_j')_+^{-1}(t_j)_+ = \tilde{t}_j'^{-1}(t_j')_-^{-1}(t_j)_-\tilde{t}_j = \tilde{t}_j'^{-1}a_j\tilde{t}_j \in T_j^{a_j}a_jT_j = a_jT_j. \tag{17-9}$$

The broken lines p and $\gamma = qp_1\cdots p_{i-1}\alpha p'\beta t_{j+1}\cdots t_m$ have the same endpoints in $\Gamma(G, X\cup\mathcal{H})$. Hence, in view of (17-8) and (17-7), we obtain

$$\begin{aligned} g = \tilde{p} = \tilde{\gamma} &= \tilde{q}\tilde{p}_1\cdots\tilde{p}_{i-1}\tilde{\alpha}\tilde{p}'\tilde{\beta}\tilde{t}_{j+1}\cdots\tilde{t}_m \\ &= \tilde{q}\tilde{p}_1\cdots\tilde{p}_{i-1}(\tilde{p}_i\tilde{p}_i'^{-1})(zxa_1y_1b_1y_2b_2\cdots b_{j-1}y_ja_j^{-1})\tilde{\beta}\tilde{t}_{j+1}\cdots\tilde{t}_m \\ &= \tilde{q}\tilde{p}_1\cdots\tilde{p}_{i-1}(\tilde{p}_i\tilde{p}_i'^{-1}z)(xa_1)(y_1b_1)\cdots(y_{j-1}b_{j-1})(y_ja_j^{-1}\tilde{\beta})\tilde{t}_{j+1}\cdots\tilde{t}_m. \end{aligned} \tag{17-10}$$

Recall that $\tilde{q} \in Q$, $\tilde{p}_1, \ldots, \tilde{p}_{i-1} \in Q' \cup R'$ and $\tilde{t}_l \in T_l$, for $l = j+1, \ldots, m$, by definition. On the other hand, $\tilde{p}_i \tilde{p}_i'^{-1} z \in Q' \tilde{p}_i'^{-1} \tilde{p}_i' Q'_P = Q'$, $xa_1 \in R$ by (17-6) and $y_l b_l \in T_l$, for each $l = 1, \ldots, j-1$, by construction. Finally, $y_j a_j^{-1} \tilde{\beta} \in T_j a_j^{-1} a_j T_j = T_j$ by (17-9). Thus, following Remark 16.6, the product decomposition (17-10) for g gives us a path representative of g with width $i < n$. This contradicts the minimality of the type of p, so the proposition is proved. □

Condition (C2-m) can be used deal with another case of multiple backtracking.

Proposition 17.6 *For every $\zeta \geq 0$ there is a constant $B_1 = B_1(\zeta) \geq 0$ such that if $Q' \leqslant Q$ and $R' \leqslant R$ satisfy condition* (C2-m) *with constant $B \geq B_1$ then the following is true.*

Let $p = qp_1 \cdots p_n r t_1 \cdots t_m$ be a minimal type path representative for some $g \in Q\langle Q', R'\rangle RT_1 \cdots T_m$, with $g \notin QRT_1 \cdots T_m$, and let $\nu \in \mathcal{N}$. Suppose that p has multiple backtracking along H_ν-components $h_1, \ldots, h_k$ of its segments such that

- *h_1 is an H_ν-component of p_i, for some $i \in \{1, \ldots, n-1\}$, with $\tilde{p}_i \in R'$;*
- *h_k is an H_ν-component of t_j for some $j \in \{1, \ldots, m\}$.*

Then $d_X((h_1)_-, (h_k)_+) \geq \zeta$.

Proof Take $B_1 = \zeta + 2\varepsilon + 1$, where $\varepsilon \geq 0$ is a quasiconvexity constant for the subgroups R and $T_1, \ldots, T_m$ (as in Convention 16.1), and let $B \geq B_1$. Suppose, for a contradiction, that $d_X((h_1)_-, (h_k)_+) < \zeta$.

Since $\tilde{p}_i \in R'$, we have $d_X((h_1)_-, (p_i)_+ R) \leq \varepsilon$, by the quasiconvexity of R. Therefore there is a geodesic path p_i' in $\Gamma(G, X \cup \mathcal{H})$, such that $\tilde{p}_i' \in R$, $d_X((p_i')_-, (h_1)_-) \leq \varepsilon$ and $(p_i')_+ = (p_i)_+$. Similarly, using the quasiconvexity of T_j, we can find a geodesic path t_j' in $\Gamma(G, X \cup \mathcal{H})$, such that $\tilde{t}_j' \in T_j$, $(t_j')_- = (t_j)_-$ and $d_X((t_j')_+, (h_k)_+) \leq \varepsilon$. Let p' be the broken line $p_i' p_{i+1} \cdots p_n r t_1 \cdots t_{j-1} t_j'$.

Observe that $\tilde{p}' \in R\langle Q', R'\rangle RT_1 \cdots T_j$ and, by the triangle inequality, $|p'|_X \leq \zeta + 2\varepsilon$. Therefore we can apply condition (C2-m) to $\tilde{p}'$ to find that $\tilde{p}' = xy_1 \cdots y_j$, where $x \in R$ and $y_l \in T_l$, for each $l = 1, \ldots, j$.

The broken lines p and $\gamma = qp_1 \cdots p_i p_i'^{-1} p' t_j'^{-1} t_j \cdots t_m$ have the same endpoints; hence

$$
\begin{aligned}
g = \tilde{p} = \tilde{\gamma} &= \tilde{q}\tilde{p}_1 \cdots \tilde{p}_i \tilde{p}_i'^{-1} \tilde{p}' \tilde{t}_j'^{-1} \tilde{t}_j \cdots \tilde{t}_m \\
&= \tilde{q}\tilde{p}_1 \cdots \tilde{p}_{i-1} (\tilde{p}_i \tilde{p}_i'^{-1} x) y_1 \cdots y_{j-1} (y_j \tilde{t}_j'^{-1} \tilde{t}_j) \tilde{t}_{j+1} \cdots \tilde{t}_m.
\end{aligned} \tag{17-11}
$$

Note that $\tilde{p}_i \tilde{p}_i'^{-1} x \in R$ and $y_j \tilde{t}_j'^{-1} \tilde{t}_j \in T_j$. In view of Remark 16.6, the product decomposition of g from (17-11) can be used to obtain a path representative p'' of g with width $i - 1 < n$. Thus the type of p'' is strictly less than the type of p, which yields the desired contradiction. □

18 Multiple backtracking in product path representatives: general case

Propositions 8.5, 17.5 and 17.6 above show that for $g \in Q\langle Q', R'\rangle RT_1 \cdots T_m \setminus QRT_1 \cdots T_m$, instances of multiple backtracking in a minimal type path representative $p = qp_1 \cdots p_n r t_1 \cdots t_m$, that start at a component of $q, p_1, \ldots,$ or p_{n-1}, are long. We cannot draw the same conclusion in all cases since we have no control over the elements $\tilde{r}, \tilde{t}_1, \dots, \tilde{t}_m$. Therefore in this section we use a different approach. Proposition 18.3 below shows that in the remaining cases we can find a path representative with one of the segments from the tail section $rt_1 \cdots t_m$ being short with respect to the proper metric d_X. Note that the main constant $\xi_0 = \xi_0(Q', \zeta)$, produced in this proposition, will depend on Q' (unlike the constants $C_1, D, C_2(\zeta), B_1(\zeta), \dots$, defined previously) but will be independent of R'.

As before, we work under Convention 16.1. We will also keep using Notation 17.1 and 17.3. Let us start with the following elementary observation.

Lemma 18.1 *For any $\zeta \geq 0$ and any given subsets $A_1, \dots, A_k \subseteq G$, $k \geq 1$, there is a constant*

$$\xi = \xi(\zeta, A_1, \dots, A_k) \geq 0$$

such that if $g \in A_1 \cdots A_k$ and $|g|_X \leq \zeta$, then there exist $a_1 \in A_1, \dots, a_k \in A_k$ such that $g = a_1 \cdots a_k$ and $|a_i|_X \leq \xi$, for all $i \in \{1, \dots, k\}$.

Proof For each $g \in A_1 \cdots A_k$ fix some elements $a_{1,g} \in A_1, \dots, a_{k,g} \in A_k$ such that $g = a_{1,g} \cdots a_{k,g}$. Now we can define

$$\xi = \max\bigl\{|a_{1,g}|_X, \dots, |a_{k,g}|_X \mid g \in A_1 \cdots A_k,\ |g|_X \leq \zeta\bigr\} < \infty.$$

Clearly ξ has the required property. □

Definition 18.2 (tail height) Suppose that $Q' \leqslant Q$, $R' \leqslant R$ and $p = qp_1 \cdots p_n rt_1 \cdots t_m$ is a path representative of an element $g \in Q\langle Q', R'\rangle RT_1 \cdots T_m$. The *tail height* of p, $\mathrm{th}_X(p)$, is defined as

$$\mathrm{th}_X(p) = \min\{|r|_X, |t_1|_X, \dots, |t_{m-1}|_X\}.$$

Proposition 18.3 *For each $\zeta \geq 0$, let $C_2 = C_2(\zeta)$ be the larger of the two constants provided by Propositions 8.5 and 17.5, and let $B_1 = B_1(\zeta)$ be given by Proposition 17.6. Set $B_2 = B_2(\zeta) = \max\{C_2(\zeta), B_1(\zeta)\}$.*

Suppose that $Q' \leqslant Q$ is a relatively quasiconvex subgroup of G containing $S = Q \cap R$. Then there exists a constant $\xi_0 = \xi_0(Q', \zeta) \geq 0$ such that if $R' \leqslant R$ and Q' and R' satisfy conditions (C1)–(C4), (C2-m) *and* (C5-m), *with constants $B \geq B_2$ and $C \geq C_2$ and collections of subgroups $\mathcal{P} \supseteq \mathcal{P}_1$ and $\mathcal{U} \supseteq \mathcal{U}_1$, then the following is true.*

Let $p = qp_1 \cdots p_n rt_1 \cdots t_m$ be a minimal type path representative for some $g \in Q\langle Q', R'\rangle RT_1 \cdots T_m$, with $g \notin QRT_1 \cdots T_m$. Suppose that p has multiple backtracking along $\mathcal{H}$-components $h_1, \dots, h_k$ of its segments, with $k \geq 3$ and $d_X((h_1)_-, (h_k)_+) \leq \zeta$. Then $m \geq 1$ and there is a path representative p' for g (*not necessarily of minimal type*) *such that $\mathrm{th}_X(p') \leq \xi_0$.*

Proof Let $\varepsilon' \geq 0$ be a quasiconvexity constant for Q'. Take $\xi_0 = \xi_0(Q', \zeta) \geq 0$ to be the maximum, taken over all indices i and j satisfying $1 \leq i \leq j \leq m$, of the constants

$$\xi(\zeta + \varepsilon + \varepsilon', Q', R, T_1, \dots, T_j), \quad \xi(\zeta + 2\varepsilon, R, T_1, \dots, T_j), \quad \xi(\zeta + 2\varepsilon, T_i, \dots, T_j),$$

obtained from Lemma 18.1.

Suppose that $h_1, \dots, h_k$ are as in the statement, with $d_X((h_1)_-, (h_k)_+) \le \zeta$. There are four possible cases to consider, depending on the segments of p to which the $\mathcal{H}$-components h_1 and h_k belong to. If h_k is an $\mathcal{H}$-component of one of the segments $p_2, \dots, p_n$ or r, then one obtains a contradiction to the minimality of type of p by following the same argument as in Proposition 8.5 (recall that (C5-m) implies (C5) by Remark 16.2).

If h_1 is an $\mathcal{H}$-component of one of the segments $q, p_1, \dots, p_{n-1}$ and h_k is an $\mathcal{H}$-component of one of the segments $t_1, \dots, t_m$, we obtain a contradiction by applying either Proposition 17.5 or 17.6 (depending on whether h_1 is a component of a segment of p representing an element of Q or R, respectively).

It remains to consider the possibility when h_1 is an $\mathcal{H}$-component of one of the segments $p_n, r, t_1, \dots, t_m$. It follows that h_k is an $\mathcal{H}$-component of t_j, for some $j \in \{1, \dots, m\}$, in particular $m \ge 1$. For simplicity we treat only the case when h_1 is an $\mathcal{H}$-component of p_n; the remaining cases can be dealt with similarly.

Note that $\tilde{p}_n \in Q'$ by Remark 16.7. By the relative quasiconvexity of Q' and T_j there are geodesic paths α and β in $\Gamma(G, X \cup \mathcal{H})$ satisfying

$$d_X(\alpha_-, (h_1)_-) \le \varepsilon', \quad \alpha_+ = (p_n)_+, \quad \tilde{\alpha} \in Q',$$
$$\beta_- = (t_j)_-, \quad d_X(\beta_+, (h_k)_+) \le \varepsilon, \quad \tilde{\beta} \in T_j.$$

Let $\gamma = \alpha r t_1 \cdots t_{j-1} \beta$. Observe that $\tilde{\gamma} \in Q' R T_1 \cdots T_j$ and, by the triangle inequality,

$$|\gamma|_X = d_X(\alpha_-, \beta_+) \le \varepsilon' + \zeta + \varepsilon.$$

Thus, applying Lemma 18.1, we can find elements $x \in Q'$, $y \in R$, $z_1 \in T_1, \dots, z_j \in T_j$ such that $\tilde{\gamma} = x y z_1 \cdots z_j$ and

$$|y|_X \le \xi_0. \tag{18-1}$$

Therefore

$$\begin{aligned} g = \tilde{p} &= \tilde{q}\tilde{p}_1 \cdots \tilde{p}_n (\tilde{\alpha}^{-1}\tilde{\alpha}) \tilde{r} \tilde{t}_1 \cdots \tilde{t}_{j-1} (\tilde{\beta}\tilde{\beta}^{-1}) \tilde{t}_j \cdots \tilde{t}_m \\ &= \tilde{q}\tilde{p}_1 \cdots \tilde{p}_n \tilde{\alpha}^{-1} \tilde{\gamma} \tilde{\beta}^{-1} \tilde{t}_j \cdots \tilde{t}_m \\ &= \tilde{q}\tilde{p}_1 \cdots \tilde{p}_{n-1} (\tilde{p}_n \tilde{\alpha}^{-1} x) y z_1 \cdots z_{j-1} (z_j \tilde{\beta}^{-1} \tilde{t}_j) \tilde{t}_{j+1} \cdots \tilde{t}_m. \end{aligned} \tag{18-2}$$

Following Remark 16.6, the product decomposition (18-2) gives rise to a path representative

$$p' = q' p'_1 \cdots p'_n r' t'_1 \cdots t'_m$$

for g, where $\tilde{q}' = \tilde{q} \in Q$, $\tilde{p}'_i = \tilde{p}_i \in Q' \cup R'$, for $i = 1, \dots, n-1$, $\tilde{p}'_n = \tilde{p}_n \tilde{\alpha}^{-1} x \in Q'$, $\tilde{r}' = y \in R$, $\tilde{t}'_l = z_l \in T_l$, for $l = 1, \dots, j-1$, $\tilde{t}'_j = z_j \tilde{\beta}^{-1} \tilde{t}_j \in T_j$ and $\tilde{t}'_s = \tilde{t}_s \in T_s$, for $s = j+1, \dots, m$. In view of (18-1), we see that $\mathrm{th}_X(p') \le |y|_X \le \xi_0$, so the proof is complete. □

The following proposition is an analogue of Lemma 10.3. It employs the constant $c_0 = \max\{C_0, 14\delta\}$, where C_0 is provided by Lemma 16.8, and the constants $\lambda = \lambda(c_0) \ge 1$ and $c = c(c_0) \ge 0$, given by Proposition 9.4.

Proposition 18.4 *For any $\eta \geq 0$ there are constants $\zeta = \zeta(\eta) \geq 0$, $C_3 = C_3(\eta) \geq 0$, $\Theta_1 = \Theta_1(\eta) \in \mathbb{N}$ and $B_3 = B_3(\eta) \geq 0$ such that if $Q' \leqslant Q$ is a relatively quasiconvex subgroup of G and $B \geq B_3$, $C \geq C_3$ then there exists $E = E(\eta, Q', B) \geq 0$ such that the following holds.*

Suppose Q' and some subgroup $R' \leqslant R$ satisfy conditions (C1)–(C4), (C2-m) *and* (C5-m), *with constants B and C, and families $\mathcal{P} \supseteq \mathcal{P}_1$ and $\mathcal{U} \supseteq \mathcal{U}_1$. Let p be a minimal type path representative for an element $g \in Q\langle Q', R'\rangle RT_1 \cdots T_m \setminus QRT_1 \cdots T_m$. Assume that for any path representative p' for g we have $\mathrm{th}_X(p') \geq E$. Then p is $(B, c_0, \zeta, \Theta_1)$-tamable.*

Let $\Sigma(p, \Theta_1) = f_0 e_1 f_1 \cdots e_l f_l$ denote the Θ_1-shortcutting of p, obtained by applying Procedure 9.1, and let e'_j be the $\mathcal{H}$-component of $\Sigma(p, \Theta_1)$ containing e_j, $j = 1, \dots, l$. Then $\Sigma(p, \Theta_1)$ is a (λ, c)-quasigeodesic without backtracking and $|e'_j|_X \geq \eta$, for each $j = 1, \dots, l$.

Proof The proof is similar to the argument in Lemma 10.3. Let us define the necessary constants:

- $\zeta = \zeta(\eta, c_0)$ is the constant from Proposition 9.4;
- $\Theta_1 = \max\{\Theta_0(\zeta), \zeta\}$, where Θ_0 is the constant from Lemma 16.10;
- $B_2(\zeta)$ and $C_3 = C_2(\zeta)$ are the constants provided by Proposition 18.3;
- $B_3 = \max\{B_0(\Theta_1, c_0), B_2(\zeta)\}$, where $B_0(\Theta_1, c_0)$ is the constant from Proposition 9.4;

and, finally, for any given $B \geq B_3, C \geq C_3$, we set

- $E = \max\{B, \xi_0(\eta, Q') + 1\}$, where $\xi_0(\eta, Q')$ is the constant from Proposition 18.3.

Suppose that Q', R', g and $p = qp_1 \cdots p_n r t_1 \cdots t_m$ are as in the statement of the proposition. We will now show that p is $(B, c_0, \zeta, \Theta_1)$-tamable.

Since Q' and R' satisfy (C2), Lemma 10.1 together with Remark 16.7 imply that $|p_i|_X \geq B$, for each $i = 1, \dots, n$. Moreover, by assumption, $|r|_X, |t_1|_X, \dots, |t_{m-1}|_X \geq E \geq B$, so condition (i) of Definition 9.3 is satisfied. On the other hand, condition (ii) is satisfied by Lemma 16.8.

If condition (iii) of Definition 9.3 is not satisfied then p must have consecutive backtracking along $\mathcal{H}$-components $h_1, \dots, h_k$ of its segments, such that

$$\max\{|h_i|_X \mid i = 1, \dots, k\} \geq \Theta_1 \quad \text{and} \quad d_X((h_1)_-, (h_k)_+) < \zeta.$$

Lemma 16.10 rules out the case of adjacent backtracking ($k = 2$), so it must be that $k \geq 3$. That is, $h_1, \dots, h_k$ is an instance of multiple backtracking in p. Proposition 18.3 now applies, giving a path representative p' for g with $\mathrm{th}_X(p') \leq \xi_0(\eta, Q') < E$. This contradicts a hypothesis of the proposition, so p must also satisfy condition (iii).

Thus p is $(B, c_0, \zeta, \Theta_1)$-tamable, and we can apply Proposition 9.4 to achieve the desired conclusion. □

19 Using separability to establish conditions (C2-m) and (C5-m)

In this section we exhibit, under suitable assumptions on G, the existence of finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ satisfying conditions (C1)–(C4), (C2-m) and (C5-m).

Lemma 19.1 *Let G be a group generated by finite set X, let $Q, R, T_1, \dots, T_m \leqslant G$ be some subgroups, and let $S = Q \cap R$. Suppose that $RT_1 \cdots T_l$ is separable in G, for each $l = 0, \dots, m$. Then for any $B \geq 0$ there is a finite-index subgroup $N \leqslant_f G$, with $S \subseteq N$, such that arbitrary subgroups $Q' \leqslant Q \cap N$ and $R' \leqslant R \cap N$ satisfy condition* (C2-m) *with constant B.*

Proof For each $l \in \{0, \dots, m\}$ the product $RT_1 \cdots T_l$ is separable, so, by Lemma 4.16(b), there is a finite-index normal subgroup $M_l \lhd_f G$ such that

$$\min_X(RT_1 \cdots T_l M_l \setminus RT_1 \cdots T_l) \geq B \quad \text{for all } l = 0, \dots, m. \tag{19-1}$$

Define the subgroup $M = \bigcap_{l=0}^m M_l \lhd_f G$, and take $N = SM \leqslant_f G$. Observe that

$$RNRT_1 \cdots T_l = RSMRT_1 \cdots T_l = RSRT_1 \cdots T_l M = RT_1 \cdots T_l M \quad \text{for all } l = 0, \dots, m. \tag{19-2}$$

Now choose arbitrary subgroups $Q' \leqslant Q \cap N$ and $R' \leqslant R \cap N$, so that $\langle Q', R' \rangle \subseteq N$. Since $M \subseteq M_l$ for all l, we can combine (19-1) with (19-2) to draw the desired conclusion. □

The next statement is similar to Theorem 11.3.

Lemma 19.2 *Suppose that G is a group generated by finite set X and $m \in \mathbb{N}_0$. Let $Q, R \leqslant G$ be some subgroups, and let $\mathcal{P}$ and $\mathcal{U}$ be finite collections of subgroups of G such that*

1. *each $P \in \mathcal{P}$ has property* RZ_{m+2};
2. *the subgroups $Q \cap P$, $R \cap P$ and $U \cap P$ are finitely generated, for all $P \in \mathcal{P}$ and all $U \in \mathcal{U}$;*
3. *if $P \in \mathcal{P}$, $K \leqslant_f P$ and $L \leqslant_f Q$ then KL is separable in G.*

Then for any $C \geq 0$ and any finite-index subgroup $Q' \leqslant_f Q$, there is a finite-index subgroup $O \leqslant_f G$, with $Q' \subseteq O$, such for any $R' \leqslant R \cap O$ the subgroups Q' and R' satisfy (C5-m) *with constant C and collections $\mathcal{P}$ and $\mathcal{U}$.*

Proof As usual, for subgroups $H \leqslant G$ and $P \in \mathcal{P}$ we denote $H \cap P$ by H_P.

Fix an enumeration $\mathcal{P} = \{P_1, \dots, P_k\}$ and let $Q' \leqslant_f Q$ be a finite-index subgroup of Q. Given any $i \in \{1, \dots, k\}$, we choose some coset representatives $a_{i1}, \dots, a_{in_i} \in Q_{P_i}$ of Q'_{P_i}, so that

$$Q_{P_i} = \bigsqcup_{j=1}^{n_i} a_{ij} Q'_{P_i}.$$

Let $\mathbb{U}$ be the finite set consisting of all l-tuples $(U_1, \dots, U_l)$, where $l \in \{0, \dots, m\}$ and $U_1, \dots, U_l \in \mathcal{U}$.

Consider any $i \in \{1, \ldots, k\}$ and $\underline{u} = (U_1, \ldots, U_l) \in \mathbb{U}$, where $l \in \{0, \ldots, m\}$. Note that $Q'_{P_i} \leqslant_f Q_{P_i}$ is finitely generated, for each $i = 1, \ldots, k$, since Q_{P_i} is itself finitely generated by (2). Combining assumptions (1) and (2), the subset $Q'_{P_i} R_{P_i} (U_1)_{P_i} \cdots (U_l)_{P_i}$ is separable in P_i. Therefore, by Lemma 4.16(c), for any $C \geq 0$ there is $F_{i,\underline{u}} \lhd_f P_i$ such that

$$\min_X \bigl(a_{ij} Q'_{P_i} F_{i,\underline{u}} R_{P_i} (U_1)_{P_i} \cdots (U_l)_{P_i} \setminus a_{ij} Q'_{P_i} R_{P_i} (U_1)_{P_i} \cdots (U_l)_{P_i}\bigr) \geq C, \tag{19-3}$$

for all $j = 1, \ldots, n_i$.

Define $K_{i,\underline{u}} = Q'_{P_i} F_{i,\underline{u}} \leqslant_f P_i$. Then (19-3) implies that for every $j = 1, \ldots, n_i$ we have

$$\min_X \bigl(a_{ij} K_{i,\underline{u}} R_{P_i} (U_1)_{P_i} \cdots (U_l)_{P_i} \setminus a_{ij} Q'_{P_i} R_{P_i} (U_1)_{P_i} \cdots (U_l)_{P_i}\bigr) \geq C. \tag{19-4}$$

Assumption (3) tells us that the double coset $K_{i,\underline{u}} Q'$ is separable in G, and since $Q' \cap P_i = Q'_{P_i} \subseteq K_{i,\underline{u}}$, we can apply Lemma 11.2 to find a finite-index subgroup $O_{i,\underline{u}} \leqslant_f G$ such that $Q' \subseteq O_{i,\underline{u}}$ and $O_{i,\underline{u}} \cap P_i \subseteq K_{i,\underline{u}}$.

We can now define a finite-index subgroup O of G by

$$O = \bigcap_{i=1}^{k} \bigcap_{\underline{u} \in \mathbb{U}} O_{i,\underline{u}} \leqslant_f G.$$

Observe that $Q' \subseteq O$ and $O \cap P_i \subseteq K_{i,\underline{u}}$, for each $i = 1, \ldots, k$ and all $\underline{u} \in \mathbb{U}$. Consider any subgroup $R' \leqslant R \cap O$. Then $Q'_{P_i} \cup R'_{P_i} \subseteq O \cap P_i$, so (19-4) yields that

$$\min_X \bigl(a_{ij} \langle Q'_{P_i}, R'_{P_i} \rangle R_{P_i} (U_1)_{P_i} \cdots (U_l)_{P_i} \setminus a_{ij} Q'_{P_i} R_{P_i} (U_1)_{P_i} \cdots (U_l)_{P_i}\bigr) \geq C, \tag{19-5}$$

for arbitrary $i = 1, \ldots, k$, $l = 0, \ldots, m$, $U_1, \ldots, U_l \in \mathcal{U}$ and any $j = 1, \ldots, n_i$.

Given any $i \in \{1, \ldots, k\}$ and any $q \in Q_{P_i}$, there is $j \in \{1, \ldots, n_i\}$ such that $q Q'_{P_i} = a_{ij} Q'_{P_i}$. It follows that $q \langle Q'_{P_i}, R'_{P_i} \rangle = a_{ij} \langle Q'_{P_i}, R'_{P_i} \rangle$, which, combined with (19-5), shows that Q' and R' satisfy condition (C5-m), as required. $\square$

For the next result we will follow the notation of Convention 16.1.

Proposition 19.3 *Suppose that G is QCERF, the product $RT_1 \cdots T_l$ is separable in G, for every $l = 0, \ldots, m$, and the peripheral subgroup H_ν has property* RZ_{m+2}, *for each $\nu \in \mathcal{N}$. Let $\mathcal{P}_1$ be a finite collection of maximal parabolic subgroups and let $\mathcal{U}_1$ be a finite collection of finitely generated relatively quasiconvex subgroups in G.*

Then for any $B, C \geq 0$ there exist finite-index subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$ such that

- *$\langle Q', R' \rangle$ is relatively quasiconvex in G;*
- *Q' and R' satisfy conditions* (C1)–(C4), (C2-m) *and* (C5-m) *with constants B and C and collections $\mathcal{P}_1$ and $\mathcal{U}_1$.*

More precisely, there is $L_1 \leqslant_f G$, with $S \subseteq L_1$, such that for any $L' \leqslant_f L_1$, satisfying $S \subseteq L'$, we can take $Q' = Q \cap L' \leqslant_f Q$, and there is $M_1 \leqslant_f L'$, with $Q' \subseteq M_1$, such that for any $M' \leqslant_f M_1$, satisfying $Q' \subseteq M'$, the subgroups Q' and $R' = R \cap M' \leqslant_f R$ enjoy the above properties.

Proof Fix some constants $B, C \geq 0$. Let $\mathcal{P}_0$ be the finite collection of maximal parabolic subgroups of G provided by Theorem 3.5 and set $\mathcal{P} = \mathcal{P}_0 \cup \mathcal{P}_1$.

Note that maximal parabolic subgroups of G are double coset separable by the assumptions, as $m+2 \geq 2$. Therefore the argument from the proof of Theorem 14.4 shows that G, its subgroups Q, R and $S = Q \cap R$, and the finite collection $\mathcal{P}$ satisfy assumptions (S1)–(S4) of Theorem 11.3. Let $L \leqslant_f G$, with $S \subseteq L$, be the finite-index subgroup provided by this theorem.

By the hypothesis on G, the subsets $RT_1 \cdots T_l$ are separable in G, for each $l = 0, \dots, m$. We can therefore apply Lemma 19.1 to obtain a finite-index subgroup $N \leqslant_f G$ from its statement (in particular, $S \subseteq N$). Now we define the finite-index subgroup $L_1 \leqslant_f G$, from the statement of the proposition, by setting $L_1 = L \cap N$. Clearly L_1 contains S. Take any $L' \leqslant_f L_1$, with $S \subseteq L'$, and set $Q' = Q \cap L' \leqslant_f Q$. Let $M \leqslant_f L'$ be the subgroup provided by Theorem 11.3, with $Q' \subseteq M$.

Lemma 5.24 and Corollary 13.4 imply that all the assumptions of Lemma 19.2 are satisfied, so let $O \leqslant_f G$ be the subgroup given by this lemma, with $Q' \subseteq O$. We now define the finite-index subgroup $M_1 \leqslant_f L'$, from the statement of the proposition, by $M_1 = M \cap O$.

Evidently, M_1 contains Q'. Choose an arbitrary finite-index subgroup $M' \leqslant_f M_1$, with $Q' \subseteq M'$, and set $R' = R \cap M'$. Observe that $M' \leqslant_f G$, by construction, hence $R' \leqslant_f R$.

The combined statements of Theorems 11.3 and 3.5 and Lemmas 5.22, 19.1 and 19.2 now imply that the subgroups $Q' \leqslant_f Q$ and $R' \leqslant_f R$, obtained above, satisfy all of the required properties. □

20 Separability of quasiconvex products in QCERF relatively hyperbolic groups

In this section we prove Theorem 1.8 from the introduction.

Remark 20.1 Let G be a relatively hyperbolic group. Suppose that $s \in \mathbb{N}$ and the product of any s finitely generated relatively quasiconvex subgroups is separable in G. If $Q_1, \dots, Q_s$ are finitely generated quasiconvex subgroups of G and $a_0, \dots, a_s \in G$ are arbitrary elements, then the subset $a_0 Q_1 a_1 \cdots Q_s a_s$ is separable in G.

Indeed, observe that the subset

$$a_0 Q_1 a_1 \cdots Q_s a_s = Q_1^{a_0} Q_2^{a_0 a_1} \cdots Q_s^{a_0 \cdots a_{s-1}} a_0 \cdots a_s$$

is a translate of a product of conjugates of the subgroups $Q_1, \dots, Q_s$. Combining Lemma 5.22 with Remark 4.12 and the assumption on G yields the desired conclusion.

Proof of Theorem 1.8 We induct on s. The case $s = 1$ is equivalent to the QCERF property of G, while the case $s = 2$ follows from Corollary 1.4. Thus we can assume that $s > 2$ and the product of any $s-1$ finitely generated relatively quasiconvex subgroups is separable in G.

Let $F_1, \dots, F_s$ be finitely generated relatively quasiconvex subgroups of G. For ease of notation we write $m = s-2$, $Q = F_1$, $R = F_2$ and $T_i = F_{i+2}$, for $i \in \{1, \dots, m\}$. Choose a finite generating set X for G and let $\delta \in \mathbb{N}$ be a hyperbolicity constant for the Cayley graph $\Gamma(G, X \cup \mathcal{H})$, where $\mathcal{H} = \bigsqcup_{\nu \in \mathcal{N}} (H_\nu \setminus \{1\})$. Denote by $\varepsilon \geq 0$ a common quasiconvexity constant for $Q, R, T_1, \dots, T_m$.

Arguing by contradiction, suppose that the subset $QRT_1 \cdots T_m = F_1 \cdots F_s$ is not separable in G. Then there exists $g \in G \setminus QRT_1 \cdots T_m$ such that g belongs to the profinite closure of $QRT_1 \cdots T_m$ in G. Let us fix the following notation for the remainder of the proof:

- $c_0 = \max\{C_0, 14\delta\} \geq 0$, where C_0 is the constant obtained from Lemma 16.8;
- $c_3 = c_3(c_0) \geq 0$ is the constant obtained from Lemma 4.11;
- $\lambda = \lambda(c_0) \geq 1$ and $c = c(c_0) \geq 0$ are obtained from Proposition 9.4, applied with the constant c_0;
- $\mathcal{P}_1$ is the finite family of maximal parabolic subgroups of G from Notation 17.1;
- $\mathcal{U}_1$ is the finite collection of finitely generated relatively quasiconvex subgroups of G from Notation 17.3;
- $A = |g|_X + 1$ and $\eta = \eta(\lambda, c, A) \geq 0$ is obtained from Lemma 5.12;
- $\zeta = \zeta(\eta) \geq 0$, $\Theta_1 = \Theta_1(\eta) \geq 0$, $C_3 = C_3(\eta) \geq 0$ and $B_3 = B_3(\eta) \geq 0$ are the constants obtained from Proposition 18.4;
- $B = \max\{B_3(\eta), (4A + c_3)\Theta_1\}$ and $C = C_3(\eta)$.

Observe that, by the induction hypothesis, the product $RT_1 \cdots T_l$ is separable in G, for every $l = 0, \dots, m$. Let $L_1 \leqslant_f G$ be the finite-index subgroup obtained from Proposition 19.3, applied with finite families $\mathcal{P}_1, \mathcal{U}_1$ and constants B, C, given above. Note that $S \subseteq L_1$, and define $Q' = Q \cap L_1 \leqslant_f Q$. Again, by Proposition 19.3, there is a finite-index subgroup $M_1 \leqslant_f L_1$ such that $Q' \subseteq M_1$ and for any $M' \leqslant_f M_1$, with $Q' \subseteq M'$, the subgroups Q' and $R' = R \cap M' \leqslant_f R$ satisfy the conclusion of Proposition 19.3.

Let $E = E(\eta, Q', B) \geq 0$ be the constant provided by Proposition 18.4. Let $\{N_j \mid j \in \mathbb{N}\}$ be an enumeration of the finite-index subgroups of M_1 containing Q', and define the subgroups

$$M_i' = \bigcap_{j=1}^{i} N_j \leqslant_f L' \quad \text{and} \quad R_i' = M_i' \cap R \leqslant_f R, \quad i \in \mathbb{N}. \tag{20-1}$$

Note that for every $i \in \mathbb{N}$, $Q' \subseteq M_i'$, so the subgroups Q' and R_i' satisfy the conclusion of Proposition 19.3. In particular, the subgroup $\langle Q', R_i' \rangle$ is relatively quasiconvex (and finitely generated) in G, and Q' and R_i' satisfy conditions (C1)–(C4), (C2-m) and (C5-m) with constants B and C, and families $\mathcal{P}_1$ and $\mathcal{U}_1$, defined above. For each $i \in \mathbb{N}$, consider the subset

$$K_i = Q\langle Q', R_i' \rangle RT_1 \cdots T_m.$$

Choose coset representatives $x_1, \dots, x_a \in Q$ and $y_{i,1}, \dots, y_{i,b_i} \in R$ such that $Q = \bigcup_{j=1}^{a} x_j Q'$ and $R = \bigcup_{k=1}^{b_i} R'_i y_{i,k}$. Then

$$Q\langle Q', R'_i\rangle R = \bigcup_{j=1}^{a} \bigcup_{k=1}^{b_i} x_j \langle Q', R'_i\rangle y_{i,k};$$

hence K_i may be written as the finite union

$$K_i = \bigcup_{j=1}^{a} \bigcup_{k=1}^{b_i} x_j \langle Q', R'_i\rangle y_{i,k} T_1 \cdots T_m.$$

Therefore, for every $i \in \mathbb{N}$, K_i is separable in G by Remark 20.1 and the induction hypothesis. Since each K_i contains $QRT_1 \cdots T_m$ and g is in the profinite closure of $QRT_1 \cdots T_m$, it must be the case that $g \in K_i$, for every $i \in \mathbb{N}$. We will show that considering sufficiently large values of i leads to a contradiction.

For each $i \in \mathbb{N}$, let $\mathcal{S}_i$ be the set of path representatives of g in $K_i = Q\langle Q', R'_i\rangle R T_1 \cdots T_m$ (see Definition 16.4, where R' is replaced by R'_i). We will now consider two cases.

Case 1 There exists $i \in \mathbb{N}$ such that $\inf_{p' \in \mathcal{S}_i} \mathrm{th}_X(p') \geq E$.

Choose a path representative of minimal type $p = qp_1 \cdots p_n r t_1 \cdots t_m$ for g in K_i. Note that $n \geq 1$ and $\tilde{p}_1 \in R'_i \setminus S$ because $g \notin QRT_1 \cdots T_m$ (see Remark 16.7). By the assumptions of Case 1 and the above construction, we can apply Proposition 18.4 to conclude that p is $(B, c_0, \zeta, \Theta_1)$-tamable and the shortcutting $\Sigma(p, \Theta_1) = f_0 e_1 f_1 \cdots f_{l-1} e_l f_l$, obtained from Procedure 9.1, is (λ, c)-quasigeodesic without backtracking, with $|e'_k|_X \geq \eta$ for each $k = 1, \dots, l$ (where e'_k denotes the $\mathcal{H}$-component of $\Sigma(p, \Theta_1)$ containing e_k).

If $l > 0$, then applying Lemma 5.12 to the path $\Sigma(p, \Theta_1)$ gives

$$|g|_X = |p|_X = |\Sigma(p, \Theta_1)|_X \geq A > |g|_X,$$

by the choice of η, which gives a contradiction.

Therefore it must be that $l = 0$. Then p is $(4, c_3)$-quasigeodesic by Lemma 9.7 and, according to Remark 9.2(c), no segment of p contains an $\mathcal{H}$-component h with $|h|_X \geq \Theta_1$. By the quasigeodesicity of p and the fact that p_1 is a subpath of p, we have

$$|g|_{X \cup \mathcal{H}} = |p|_{X \cup \mathcal{H}} \geq \tfrac{1}{4}(\ell(p) - c_3) \geq \tfrac{1}{4}(\ell(p_1) - c_3). \tag{20-2}$$

Applying Lemma 5.10 to the geodesic p_1 in $\Gamma(G, X \cup \mathcal{H})$ we obtain

$$\ell(p_1) \geq \frac{1}{\Theta_1}|p_1|_X \geq \frac{B}{\Theta_1} \geq 4A + c_3, \tag{20-3}$$

where the second inequality follows from the fact that $\tilde{p}_1 \in R'_i \setminus S$ and Lemma 10.1. Combining (20-2) and (20-3), we get

$$|g|_X \geq |g|_{X \cup \mathcal{H}} \geq \tfrac{1}{4}(4A + c_3 - c_3) = A > |g|_X,$$

which is a contradiction.

Case 2 For all $i \in \mathbb{N}$ we have $\inf_{p' \in \mathcal{S}_i} \mathrm{th}_X(p') < E$.

Then for each $i \in \mathbb{N}$ there is a path representative $p_i = q_i p_{1,i} \cdots p_{n_i,i} r_i t_{1,i} \cdots t_{m,i} \in \mathcal{S}_i$ for g such that $\mathrm{th}(p_i) \le E$. It must either be the case that $\liminf_{i \to \infty} |r_i|_X \le E$ or $\liminf_{i \to \infty} |t_{j,i}|_X \le E$, for some $j \in \{1, \dots, m\}$. We will consider the former case, as the latter is very similar.

Since there are only finitely many elements $x \in G$ with $|x|_X \le E$, we may pass to a subsequence $(p_{i_k})_{k \in \mathbb{N}}$ such that $\tilde{r}_{i_k} = y \in R$ is some fixed element, for all $k \in \mathbb{N}$. It follows that

$$g = \tilde{p}_{i_k} \in Q\langle Q', R'_{i_k}\rangle y T_1 \cdots T_m, \quad \text{for each } k \in \mathbb{N}. \tag{20-4}$$

Now, $g \notin QyT_1 \cdots T_m$ (as $y \in R$), and the subset $QyT_1 \cdots T_m$ is separable in G by the induction hypothesis and Remark 20.1. By Lemma 4.16(a), there is a finite-index normal subgroup $O \lhd_f G$ such that $g \notin QOyT_1 \cdots T_m$. The subgroup $M_1 \cap QO$ has finite index in M_1 and contains Q'; therefore $M_1 \cap QO = N_{j_0}$, for some $j_0 \in \mathbb{N}$.

Choose $k \in \mathbb{N}$ such that $i_k \ge j_0$, so that $M'_{i_k} \subseteq N_{j_0} \subseteq QO$ (see (20-1)). Then $R'_{i_k} = M'_{i_k} \cap R \subseteq QO$; hence

$$Q\langle Q', R'_{i_k}\rangle y T_1 \cdots T_m \subseteq QOyT_1 \cdots T_m. \tag{20-5}$$

Since $g \notin QOyT_1 \cdots T_m$, inclusions (20-4) and (20-5) contradict each other.

We have arrived to a contradiction at each of the two cases; hence the proof is complete. □

21 New examples of product separable groups

In this section we prove Theorem 2.2, which will follow from the three propositions below.

Proposition 21.1 *Limit groups are product separable.*

Proof Dahmani [15] and, independently, Alibegović [3] proved that every limit group is hyperbolic relative to a collection of conjugacy class representatives of its maximal non-cyclic finitely generated abelian subgroups.

Moreover, Wilton [58] showed that limit groups are LERF and Dahmani [15] showed they are *locally quasiconvex* (that is, each of their finitely generated subgroups is relatively quasiconvex with respect to the given peripheral structure). Therefore our Theorem 1.8 yields that limit groups are product separable. □

Finitely generated Kleinian groups are not always locally quasiconvex, and we require the following two lemmas to deal with the case when one of the factors is not relatively quasiconvex.

Lemma 21.2 *Let N be a group and $n \ge 2$ be an integer. Suppose that $H_1, \dots, H_n$ are subgroups of N such that $H_i \lhd N$, for some $i \in \{1, \dots, n\}$, and the image of the product $H_1 \cdots H_{i-1} H_{i+1} \cdots H_n$ is separable in N/H_i. Then $H_1 \cdots H_n$ is separable in N.*

Proof Let $\varphi\colon N \to N/H_i$ denote the natural epimorphism. By the assumptions, the subset

$$S = \varphi(H_1 \cdots H_{i-1} H_{i+1} \cdots H_n)$$

is separable in N/H_i. Observe that

$$H_1 \cdots H_n = (H_1 \cdots H_{i-1} H_{i+1} \cdots H_n) H_i = \varphi^{-1}(S),$$

as $H_i \lhd N$; hence $H_1 \cdots H_n$ is closed in the profinite topology on N because group homomorphisms are continuous with respect to profinite topologies. □

Lemma 21.3 *Let G be a group with finitely generated subgroups $F_1, \dots, F_n \leqslant G$, $n \geq 2$. Suppose that there exists a finite-index subgroup $G' \leqslant_f G$ and an index $i \in \{1, \dots, n\}$ such that $F_i' = F_i \cap G' \lhd G'$ and G'/F_i' has property RZ_{n-1}. Then the product $F_1 \cdots F_n$ is separable in G.*

Proof Let $N \lhd_f G$ be a finite-index normal subgroup contained in G', and set $H_j = F_j \cap N$, for $j = 1, \dots, n$.

Since $|F_j : H_j| < \infty$, for each $j = 1, \dots, n$, the product $F_1 \cdots F_n$ can be written as a finite union of subsets of the form $h_1 H_1 h_2 H_2 \cdots h_n H_n$, where $h_1, \dots, h_n \in G$. Observe that

$$h_1 H_1 h_2 H_2 \cdots h_n H_n = H_1^{g_1} H_2^{g_2} \cdots H_n^{g_n} g_n,$$

where $g_j = h_1 \cdots h_j \in G$, $j = 1, \dots, n$. Thus, in view of Remark 4.12, in order to prove the separability of $F_1 \cdots F_n$ in G it is enough to show that the product $H_1^{g_1} H_2^{g_2} \cdots H_n^{g_n}$ is separable, for arbitrary $g_1, \dots, g_n \in G$.

Given any elements $g_1, \dots, g_n \in G$, the subgroups $H_1^{g_1}, H_2^{g_2}, \dots, H_n^{g_n} \leqslant G$ are finitely generated and are contained in N. Moreover, since the subgroup $H_i = F_i \cap N = F_i' \cap N$ is normal in N and $N \leqslant G'$ is normal in G, we see that $H_i^{g_i} \lhd N$ and

$$N/H_i^{g_i} = N^{g_i}/H_i^{g_i} \cong N/H_i \leqslant G'/F_i'.$$

Therefore the group $N/H_i^{g_i}$ has RZ_{n-1}, as a subgroup of G'/F_i', so the image of the product

$$H_1^{g_1} \cdots H_{i-1}^{g_{i-1}} H_{i+1}^{g_{i+1}} \cdots H_n^{g_n}$$

is separable in $N/H_i^{g_i}$. Lemma 21.2 now implies that $H_1^{g_1} H_2^{g_2} \cdots H_n^{g_n}$ is separable in N; hence it is also separable in G by Lemma 4.13(b). As we observed above, the latter yields the separability of $F_1 \cdots F_n$ in G, as required. □

Proposition 21.4 *Finitely generated Kleinian groups are product separable.*

Proof Let G be a finitely generated discrete subgroup of $\mathrm{Isom}(\mathbb{H}^3)$. We will first reduce the proof to the case when $G \backslash \mathbb{H}^3$ is a finite-volume manifold. This idea is inspired by the argument of Manning and Martínez-Pedroza used in the proof of [36, Corollary 1.5].

Using Selberg's lemma, we can find a torsion-free finite-index subgroup $K \leqslant G$. Since product separability of K implies that of G [52, Lemma 11.3.5], without loss of generality we can assume that G is torsion-free. It follows that G acts freely and properly discontinuously on $\mathbb{H}^3$, so that $M = G\backslash\mathbb{H}^3$ is a complete hyperbolic 3-manifold.

If M has infinite volume then, by [39, Theorem 4.10], G is isomorphic to a geometrically finite Kleinian group. Thus we can further assume that G is geometrically finite, which allows us to apply a theorem of Brooks [10, Theorem 2] to find an embedding of G into a torsion-free Kleinian group G^* such that $G^*\backslash\mathbb{H}^3$ is a finite-volume manifold. If G^* is product separable, then so is any subgroup of it; hence we have made the promised reduction.

Thus we can suppose that $G = \pi_1(M)$, for a hyperbolic 3-manifold M of finite volume. The tameness conjecture, proved by Agol [1] and Calegari and Gabai [11], combined with a result of Canary [12, Corollary 8.3], imply that any finitely generated subgroup $F \leqslant G$ is either geometrically finite or is a virtual fibre subgroup. The latter means that there is a finite-index subgroup $G' \leqslant_f G$ such that $F' = F \cap G' \lhd G'$ and $G'/F' \cong \mathbb{Z}$.

By [39, Theorem 3.7], G is a geometrically finite subgroup of $\mathrm{Isom}(\mathbb{H}^3)$; hence it is finitely generated and hyperbolic relative to a finite collection of finitely generated virtually abelian subgroups, each of which is product separable by [52, Lemma 11.3.5]. Moreover, by [30, Corollary 1.6], a subgroup of G is relatively quasiconvex if and only if it is geometrically finite. Finally, G is LERF (and, hence, QCERF) by [2, Corollary 9.4].

Let $F_1, \dots, F_n$ be finitely generated subgroups of G, $n \geq 2$. If F_j is geometrically finite, for all $j = 1, \dots, n$, then the product $F_1 \cdots F_n$ is separable in G by Theorem 1.8. Thus we can suppose that F_i is not geometrically finite, for some $i \in \{1, \dots, n\}$. By the above discussion, in this case F_i must be a virtual fibre subgroup of G. Since $\mathbb{Z}$ is product separable, we can apply Lemma 21.3 to conclude that $F_1 \cdots F_n$ is separable in G, completing the proof. □

Proposition 21.5 *Let G be the fundamental group of a finite graph of free groups with cyclic edge groups. If G is balanced then it is product separable.*

Limit groups and Kleinian groups are hyperbolic relative to virtually abelian subgroups. The peripheral subgroups from relatively hyperbolic structures on groups in Proposition 21.5 will be fundamental groups of graphs of cyclic groups, which motivates the next auxiliary lemma.

Lemma 21.6 *Suppose that G is the fundamental group of a finite graph of infinite cyclic groups. If G is balanced then it is product separable.*

Proof Suppose that $G = \pi_1(G_-, \Gamma)$, where (G_-, Γ) is a graph of groups, associated to a finite connected graph Γ with vertex set $V\Gamma$ and edge set $E\Gamma$. According to the assumptions, each vertex group G_v,

$v \in V\Gamma$, is infinite cyclic. As usual, we use G_e to denote the edge group corresponding to an edge $e \in E\Gamma$ (see Dicks and Dunwoody [16, Section I.3] for the definition and general theory of graphs of groups).

If $|E\Gamma| = 0$ then G is cyclic and, thus, product separable. Let us proceed by induction on $|E\Gamma|$.

Assume first that one of the edge groups G_e is trivial. If removing e disconnects Γ then G splits as a free product $G_1 * G_2$, where G_1 and G_2 are the fundamental groups of finite graphs of infinite cyclic groups corresponding to the two connected components of $\Gamma \setminus \{e\}$. Otherwise, $G \cong G_1 * G_2$, where G_1 the fundamental group of a finite graph of infinite cyclic groups corresponding to the graph $\Gamma \setminus \{e\}$ and G_2 is infinite cyclic. Moreover, G_1 and G_2 will be balanced as subgroups of a balanced group G. Hence G_1 and G_2 will be product separable by induction, so $G \cong G_1 * G_2$ will be product separable by Coulbois' theorem [14, Theorem 1].

Therefore we can assume that every edge group G_e is infinite cyclic. This means that G is a *generalised Baumslag–Solitar group*. The assumption that G is balanced now translates into the assumption that G is *unimodular*, using Levitt's terminology from [34]. We can now apply [34, Proposition 2.6] to deduce that G has a finite-index subgroup K isomorphic to the direct product $F \times \mathbb{Z}$, where F is a free group.

Now, $K \cong F \times \mathbb{Z}$ is product separable by You's result [61, Theorem 5.1]; hence G is product separable as a finite-index supergroup of K (see [52, Lemma 11.3.5]). □

Proof of Proposition 21.5 Suppose that G splits as the fundamental group of a finite graph of free groups (G_-, Γ) with cyclic edge groups.

Without loss of generality we can assume that each vertex group is a finitely generated free group (in particular, G is finitely generated). Indeed, otherwise $G \cong G_1 * F$, where G_1 is the fundamental group of a finite graph of finitely generated free groups with cyclic edge groups and F is free (this follows from the fact that any element of a free group is the product of only finitely many free generators). In this case we can deduce the product separability of G from the product separability of G_1 and F by [14, Theorem 1] (recall that F is product separable by Ribes and Zalesskii [53, Theorem 2.1]).

Now, for each vertex group G_v, choose and fix a finite family of maximal infinite cyclic subgroups $\mathbb{P}_v$ such that

(a) no two subgroups from $\mathbb{P}_v$ are conjugate in G_v;

(b) for every edge e incident to v in Γ, the image of the cyclic group G_e in G_v is conjugate into one of the subgroups from $\mathbb{P}_v$.

Condition (a) means that each G_v is hyperbolic relative to the finite family $\mathbb{P}_v$ (for example, by [8, Theorem 7.11]), and condition (b) means that each edge group of the given splitting of G is parabolic in the corresponding vertex groups. Therefore we can apply the work of Bigdely and Wise [6, Theorem 1.4] to conclude that G is hyperbolic relative to a finite collection of subgroups $\mathbb{Q}$, where each $Q \in \mathbb{Q}$ acts cocompactly on a *parabolic tree* (see [6, Definition 1.3]) with vertex stabilisers conjugate to elements of $\bigcup_{v \in V\Gamma} \mathbb{P}_v$ and edge stabilisers conjugate to elements of $\{G_e \mid e \in \Gamma\}$. The structure theorem for groups acting on trees [16, Section I.4.1] implies that every $Q \in \mathbb{Q}$ is isomorphic to the fundamental group of a finite graph of infinite cyclic groups. Since Q is balanced, being a subgroup of G, we can apply Lemma 21.6 to conclude that each $Q \in \mathbb{Q}$ is product separable. By Wise's result [59, Theorem 5.1] G is LERF, hence we can apply our Theorem 1.8 to deduce that the product of a finite number of finitely generated relatively quasiconvex subgroups is separable in G.

To establish the product separability of G it remains to show that it is locally quasiconvex. To achieve this we will again use the results of Bigdely and Wise. More precisely, according to [6, Theorem 2.6], a subgroup of G is relatively quasiconvex if it is *tamely generated*.

Let $H \leqslant G$ be a finitely generated subgroup. The splitting of G as the fundamental group of the graph of groups (G_-, Γ) induces a splitting of H as the fundamental group of a graph of groups (H_-, Δ), where for each vertex $u \in V\Delta$ the stabiliser H_u is equal to $H \cap {G_v}^g$, for some $v \in V\Gamma$ and some $g \in G$. Moreover, the graph Δ is finite, because H is finitely generated (see [16, Proposition I.4.13]). Note that every edge group from (H_-, Δ) is cyclic; hence each vertex group H_u, $u \in V\Delta$, must be finitely generated as H is finitely generated (see [6, Lemma 2.5]).

According to [6, Definition 0.1], H is tamely generated if for every $u \in V\Delta$ the subgroup $H_u = H \cap {G_v}^g$ is relatively quasiconvex in ${G_v}^g$, equipped with the peripheral structure ${\mathbb{P}_v}^g$. But the latter is true because ${G_v}^g$ is a finitely generated free group, so any finitely generated subgroup is undistorted, and hence it is relatively quasiconvex with respect to any peripheral structure on ${G_v}^g$, by [30, Theorem 1.5]. Thus every finitely generated subgroup $H \leqslant G$ is tamely generated, and so it is relatively quasiconvex in G by [6, Theorem 2.6]. $\square$

Remark 21.7 In the case when the graph of groups has two vertices and one edge (so that G is a free amalgamated product of two free groups over a cyclic subgroup), Proposition 21.5 was originally proved by Coulbois in his thesis; see [13, Theorem 5.18]. We can use similar methods to recover another result of Coulbois: if $G = H *_C F$, where H is product separable, F is free and C is a maximal cyclic subgroup in F then G is product separable [13, Theorem 5.4]. Indeed, in this case G will be hyperbolic relative to $\mathbb{Q} = \{H\}$ and will be LERF by Gitik's theorem [21, Theorem 4.4]. As in the proof of Proposition 21.5, the results from [6] imply that G is locally quasiconvex. Therefore G is product separable by Theorem 1.8.

Remark 21.8 Using recent work of Shepherd and Woodhouse [56, Theorem 1.2], Proposition 21.5 can be immediately extended to balanced groups G that split as fundamental groups of finite graphs of groups with virtually free vertex groups and virtually cyclic edge groups. In fact, by [56, Proposition 3.13], G has a torsion-free finite-index subgroup K. Then K is balanced and is isomorphic to the fundamental group of a finite graph of free groups with cyclic edge groups. So the product separability of G follows by combining Proposition 21.5 with [52, Lemma 11.3.5].

References

[1] **I Agol**, *Tameness of hyperbolic* 3*-manifolds*, preprint (2004) arXiv math/0405568

[2] **I Agol**, *The virtual Haken conjecture*, Doc. Math. 18 (2013) 1045–1087 MR Zbl

[3] **E Alibegović**, *A combination theorem for relatively hyperbolic groups*, Bull. London Math. Soc. 37 (2005) 459–466 MR Zbl

[4] **R B J T Allenby**, **R J Gregorac**, *On locally extended residually finite groups*, from "Conference on group theory" (R W Gatterdam, K W Weston, editors), Lecture Notes in Math. 319, Springer (1973) 9–17 MR Zbl

[5] **M Baker**, **D Cooper**, *A combination theorem for convex hyperbolic manifolds, with applications to surfaces in 3-manifolds*, J. Topol. 1 (2008) 603–642 MR Zbl

[6] **H Bigdely**, **D T Wise**, *Quasiconvexity and relatively hyperbolic groups that split*, Michigan Math. J. 62 (2013) 387–406 MR Zbl

[7] **B H Bowditch**, *Geometrical finiteness for hyperbolic groups*, J. Funct. Anal. 113 (1993) 245–317 MR Zbl

[8] **B H Bowditch**, *Relatively hyperbolic groups*, Internat. J. Algebra Comput. 22 (2012) art. id. 1250016 MR Zbl

[9] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR Zbl

[10] **R Brooks**, *Circle packings and co-compact extensions of Kleinian groups*, Invent. Math. 86 (1986) 461–469 MR Zbl

[11] **D Calegari**, **D Gabai**, *Shrinkwrapping and the taming of hyperbolic* 3*-manifolds*, J. Amer. Math. Soc. 19 (2006) 385–446 MR Zbl

[12] **R D Canary**, *A covering theorem for hyperbolic* 3*-manifolds and its applications*, Topology 35 (1996) 751–778 MR Zbl

[13] **T Coulbois**, *Propriétés de Ribes–Zalesskii, topologie profinie, produit libre et généralisations*, PhD thesis, Université Paris VII (2000) `https://www.i2m.univ-amu.fr/perso/thierry.coulbois/publications.html`

[14] **T Coulbois**, *Free product, profinite topology and finitely generated subgroups*, Internat. J. Algebra Comput. 11 (2001) 171–184 MR Zbl

[15] **F Dahmani**, *Combination of convergence groups*, Geom. Topol. 7 (2003) 933–963 MR Zbl

[16] **W Dicks**, **M J Dunwoody**, *Groups acting on graphs*, Cambridge Studies in Advanced Mathematics 17, Cambridge Univ. Press (1989) MR Zbl

[17] **C Druţu**, **M Sapir**, *Tree-graded spaces and asymptotic cones of groups*, Topology 44 (2005) 959–1058 MR Zbl

[18] **E Einstein**, **D Groves**, *Relatively geometric actions on* CAT(0) *cube complexes*, J. Lond. Math. Soc. 105 (2022) 691–708 MR Zbl

[19] **E Einstein**, **T Ng**, *Relative cubulation of small cancellation free products*, preprint (2021) arXiv 2111.03008

[20] **B Farb**, *Relatively hyperbolic groups*, Geom. Funct. Anal. 8 (1998) 810–840 MR Zbl

[21] **R Gitik**, *Graphs and separability properties of groups*, J. Algebra 188 (1997) 125–143 MR Zbl

[22] **R Gitik**, *On the profinite topology on negatively curved groups*, J. Algebra 219 (1999) 80–86 MR Zbl

[23] **R Gitik**, *Ping-pong on negatively curved groups*, J. Algebra 217 (1999) 65–72 MR Zbl

[24] **R Gitik**, **E Rips**, *On separability properties of groups*, Internat. J. Algebra Comput. 5 (1995) 703–717 MR Zbl

[25] **M Gromov**, *Hyperbolic groups*, from "Essays in group theory" (S M Gersten, editor), Math. Sci. Res. Inst. Publ. 8, Springer (1987) 75–263 MR Zbl

[26] **D Groves**, **J F Manning**, *Dehn filling in relatively hyperbolic groups*, Israel J. Math. 168 (2008) 317–429 MR Zbl

[27] **D Groves**, **J F Manning**, *Specializing cubulated relatively hyperbolic groups*, J. Topol. 15 (2022) 398–442 MR Zbl

[28] **F Haglund**, **D T Wise**, *Special cube complexes*, Geom. Funct. Anal. 17 (2008) 1551–1620 MR Zbl

[29] **E Hamilton**, **H Wilton**, **P A Zalesskii**, *Separability of double cosets and conjugacy classes in* 3*-manifold groups*, J. Lond. Math. Soc. 87 (2013) 269–288 MR Zbl

[30] **G C Hruska**, *Relative hyperbolicity and relative quasiconvexity for countable groups*, Algebr. Geom. Topol. 10 (2010) 1807–1856 MR Zbl

[31] **T Hsu**, **D T Wise**, *Cubulating graphs of free groups with cyclic edge groups*, Amer. J. Math. 132 (2010) 1153–1188 MR Zbl

[32] **O Kharlampovich**, **A Myasnikov**, *Elementary theory of free non-abelian groups*, J. Algebra 302 (2006) 451–552 MR Zbl

[33] **J C Lennox**, **J S Wilson**, *On products of subgroups in polycyclic groups*, Arch. Math. (Basel) 33 (1979/80) 305–309 MR Zbl

[34] **G Levitt**, *On the automorphism group of generalized Baumslag–Solitar groups*, Geom. Topol. 11 (2007) 473–515 MR Zbl

[35] **R C Lyndon**, **P E Schupp**, *Combinatorial group theory*, Ergebnisse der Math. 89, Springer (1977) MR Zbl

[36] **J F Manning**, **E Martínez-Pedroza**, *Separation of relatively quasiconvex subgroups*, Pacific J. Math. 244 (2010) 309–334 MR Zbl

[37] **E Martínez-Pedroza**, *Combination of quasiconvex subgroups of relatively hyperbolic groups*, Groups Geom. Dyn. 3 (2009) 317–342 MR Zbl

[38] **E Martínez-Pedroza**, **A Sisto**, *Virtual amalgamation of relatively quasiconvex subgroups*, Algebr. Geom. Topol. 12 (2012) 1993–2002 MR Zbl

[39] **K Matsuzaki**, **M Taniguchi**, *Hyperbolic manifolds and Kleinian groups*, Oxford Univ. Press (1998) MR Zbl

[40] **C McClellan**, *Separable at birth*: *products of full relatively quasi-convex subgroups*, PhD thesis, University of Illinois at Chicago (2019) MR `https://www.proquest.com/docview/2342583959`

[41] **A Minasyan**, *On residualizing homomorphisms preserving quasiconvexity*, Comm. Algebra 33 (2005) 2423–2463 MR Zbl

[42] **A Minasyan**, *Some properties of subsets of hyperbolic groups*, Comm. Algebra 33 (2005) 909–935 MR Zbl

[43] **A Minasyan**, *Separable subsets of GFERF negatively curved groups*, J. Algebra 304 (2006) 1090–1100 MR Zbl

[44] **G A Niblo**, *Separability properties of free groups and surface groups*, J. Pure Appl. Algebra 78 (1992) 77–84 MR Zbl

[45] **D V Osin**, *Relative Dehn functions of amalgamated products and HNN-extensions*, from "Topological and asymptotic aspects of group theory" (R Grigorchuk, M Mihalik, M Sapir, Z Šunić, editors), Contemp. Math. 394, Amer. Math. Soc., Providence, RI (2006) 209–220 MR Zbl

[46] **D V Osin**, *Relatively hyperbolic groups: intrinsic geometry, algebraic properties, and algorithmic problems*, Mem. Amer. Math. Soc. 843, Amer. Math. Soc., Providence, RI (2006) MR Zbl

[47] **D V Osin**, *Peripheral fillings of relatively hyperbolic groups*, Invent. Math. 167 (2007) 295–326 MR Zbl

[48] **J-E Pin**, *On a conjecture of Rhodes*, Semigroup Forum 39 (1989) 1–15 MR Zbl

[49] **J-E Pin**, **C Reutenauer**, *A conjecture on the Hall topology for the free group*, Bull. London Math. Soc. 23 (1991) 356–362 MR Zbl

[50] **J G Ratcliffe**, *Foundations of hyperbolic manifolds*, 2nd edition, Graduate Texts in Math. 149, Springer (2006) MR Zbl

[51] **E Reyes**, *On cubulated relatively hyperbolic groups*, Geom. Topol. 27 (2023) 575–640 MR Zbl

[52] **L Ribes**, *Profinite graphs and groups*, Ergebnisse der Math. (3) 66, Springer (2017) MR Zbl

[53] **L Ribes**, **P A Zalesskii**, *On the profinite topology on a free group*, Bull. London Math. Soc. 25 (1993) 37–43 MR Zbl

[54] **Z Sela**, *Diophantine geometry over groups, VI: The elementary theory of a free group*, Geom. Funct. Anal. 16 (2006) 707–730 MR Zbl

[55] **J-P Serre**, *Trees*, Springer (1980) MR Zbl

[56] **S Shepherd**, **D J Woodhouse**, *Quasi-isometric rigidity for graphs of virtually free groups with two-ended edge groups*, J. Reine Angew. Math. 782 (2022) 121–173 MR Zbl

[57] **H Short**, *Quasiconvexity and a theorem of Howson's*, from "Group theory from a geometrical viewpoint" (E Ghys, A Haefliger, A Verjovsky, editors), World Sci. Publ., River Edge, NJ (1991) 168–176 MR Zbl

[58] **H Wilton**, *Hall's theorem for limit groups*, Geom. Funct. Anal. 18 (2008) 271–303 MR Zbl

[59] **D T Wise**, *Subgroup separability of graphs of free groups with cyclic edge groups*, Q. J. Math. 51 (2000) 107–129 MR Zbl

[60] **W Yang**, *Combination of fully quasiconvex subgroups and its applications*, preprint (2012) arXiv 1205.2994

[61] **S You**, *The product separability of the generalized free product of cyclic groups*, J. Lond. Math. Soc. 56 (1997) 91–103 MR Zbl

School of Mathematical Sciences, University of Southampton
Southampton, United Kingdom

Mathematical Institute, University of Bonn
Bonn, Germany

aminasyan@gmail.com, lawk@math.uni-bonn.de

Received: 28 September 2022 Revised: 21 April 2023

Algebraic & Geometric Topology **25**:1 (2025) 489–561
DOI: 10.2140/agt.2025.25.489
Published: 24 March 2025

Mapping tori of A_∞-autoequivalences and Legendrian lifts of exact Lagrangians in circular contactizations

ADRIAN PETR

We study mapping tori of quasi-autoequivalences $\tau\colon \mathcal{A} \to \mathcal{A}$ which induce a free action of $\mathbb{Z}$ on objects. More precisely, we compute the mapping torus of τ when it is strict and acts bijectively on hom-sets, or when the A_∞-category $\mathcal{A}$ is directed and there is a bimodule map $\mathcal{A}(-,-) \to \mathcal{A}(-,\tau(-))$ satisfying some hypotheses. Then we apply these results in order to link together the Fukaya A_∞-category of a family of exact Lagrangians, and the Chekanov–Eliashberg DG-category of Legendrian lifts in the circular contactization.

53D42, 53D37, 18G70

Introduction

Legendrian contact homology was introduced by Chekanov [8] and Eliashberg [20], and it fits into the symplectic field theory as introduced by Eliashberg, Givental and Hofer [21]. It has been rigorously defined in the contactization of a Liouville manifold by Ekholm, Etnyre and Sullivan in [16] following [14]. The importance of Legendrian contact homology goes beyond its applications to the Legendrian isotopy problem: for example, it was used by Bourgeois, Ekholm and Eliashberg in [5] to compute symplectic invariants of Weinstein manifolds, and in a different way by Chantraine, Dimitroglou Rizell, Ghiggini and Golovko in [7] to prove a generation result for the wrapped Fukaya category of Weinstein manifolds.

The motivation for this paper is the study of Legendrian contact homology in subcritically fillable and Boothby–Wang contact manifolds, the latter being named after [4]. This has been done combinatorially in dimension three by Ekholm and Ng in [18] for the subcritically fillable case, and by Sabloff in [34] for the Boothby–Wang case. The importance of the first kind of manifolds comes from the fact that every Weinstein manifold is obtained from a subcritical Weinstein manifold (of the form $\mathbb{C} \times P$ for some Weinstein manifold P) by attaching handles along Legendrian submanifolds in its boundary at infinity. The importance of the second kind of manifolds comes from a theorem of Donaldson in [11], which states that any integral symplectic manifold (X, ω) admits a symplectic submanifold $D \subset X$ of codimension 2, such that $X \setminus D$ is a Liouville manifold whose boundary at infinity is a Boothby–Wang contact manifold. The first step before attacking both cases presented above is to study Legendrian contact homology in the circular contactization of a Liouville manifold. In fact, both subcritically fillable and Boothby–Wang contact manifolds can be seen as compactifications of such spaces. This paper links together the Fukaya A_∞-category of a family of connected compact exact Lagrangians in a Liouville manifold (P, λ), and the Chekanov–Eliashberg DG-category of Legendrian lifts in the circular contactization $(S^1 \times P, \ker(d\theta - \lambda))$.

The strategy we follow is to lift the situation to the usual contactization $\mathbb{R} \times P$ which has been much more studied. This naturally leads to consider an A_∞-category whose objects are the lifts in $\mathbb{R} \times P$ of our starting Legendrians, and morphisms spaces are generated by Reeb chords. Moreover, the deck transformations of the cover $\mathbb{R} \to S^1$ induce an A_∞-autoequivalence of this category. The rest of the proof has two main ingredients:

(1) Functorial properties of the Legendrian invariants, which are used to bring us in a situation where we can apply the correspondence result of Dimitroglou Rizell [10] between discs in the symplectization $\mathbb{R} \times \mathbb{R} \times P$ and polygons in P.

(2) Two algebraic results of independent interest about mapping tori of A_∞-autoequivalences, that allow us to bridge the gaps between the algebraic invariants we are interested in.

We now proceed to describe the organization of the paper and state our main results.

Algebra In Section 1, we briefly recall the definitions of A_∞-(co)categories and give references for standard notions that we do not recall, such as (co)bar, graded dual and Koszul dual constructions. On the other hand, we discuss in some detail the notions of modules over A_∞-categories, as well as the Grothendieck construction and homotopy pushout associated to a diagram of A_∞-categories following Ganatra, Pardon and Shende [24, Section A.4]. We use it to introduce the notion of "cylinder object for an A_∞-category", which is supposed to mimic the corresponding notion in homotopy theory.

Mapping torus of an A_∞-autoequivalence In Section 2,[1] we define the mapping torus associated to a quasi-autoequivalence τ of an A_∞-category $\mathcal{A}$ as the A_∞-category

$$\mathrm{MT}(\tau) := \operatorname{hocolim} \left(\begin{tikzcd} \mathcal{A} \sqcup \mathcal{A} \arrow[r] \arrow[d] & \mathcal{A} \\ \mathcal{A} & \end{tikzcd} \right).$$

[1] In Section 2, A_∞-categories are always assumed to be *strictly unital* (see Paragraph (2a) in Seidel's work[36]).

Observe that this terminology was also used by Kartal in [26], but we do not know if the two notions coincide. When considering an A_∞-autoequivalence $\tau\colon \mathcal{A} \to \mathcal{A}$, we always assume that $\mathcal{A}$ is equipped with a $\mathbb{Z}$-splitting of $\mathrm{ob}(\mathcal{A})$ compatible with τ, which is a bijection

$$\mathbb{Z} \times \mathcal{E} \xrightarrow{\sim} \mathrm{ob}(\mathcal{A}), \quad (n, E) \mapsto X^n(E),$$

such that $\tau(X^n(E)) = X^{n+1}(E)$ for every $n \in \mathbb{Z}$ and $E \in \mathcal{E}$ (see Definition 2.2). This naturally turns $\mathcal{A}$ into an Adams-graded A_∞-category, where the Adams degree of a morphism in $\mathcal{A}(X^i(E), X^j(E'))$ is defined to be $j - i$. It then follows that the mapping torus of τ is also Adams-graded.

Section 2 contains two results about mapping tori of A_∞-autoequivalences: we choose to only state the most important ones in this introduction. We denote by $\mathbb{F}[t_m]$ the augmented Adams-graded associative algebra generated by a variable t_m of bidegree $(m, 1)$. Observe that if $\mathcal{C}$ is a subcategory of an A_∞-category $\mathcal{D}$ with $\mathrm{ob}(\mathcal{C}) = \mathrm{ob}(\mathcal{D})$, then $\mathcal{C} \oplus (t_m\mathbb{F}[t_m] \otimes \mathcal{D})$ is naturally an Adams-graded A_∞-category, where the Adams degree of $t_m^k \otimes x$ equals k. Besides, if $\mathcal{C}$ is an A_∞-category equipped with a $\mathbb{Z}$-splitting of $\mathrm{ob}(\mathcal{C})$, we denote by $\mathcal{C}^0$ the full A_∞-subcategory of $\mathcal{C}$ whose set of objects corresponds to $\{0\} \times \mathcal{E}$. Finally, we use the functor $\mathcal{C} \mapsto \mathcal{C}_m$ of Definition 1.27.

Theorem A *Let τ be a quasi-autoequivalence of an A_∞-category $\mathcal{A}$, weakly directed with respect to some compatible $\mathbb{Z}$-splitting of* $\mathrm{ob}(\mathcal{A})$. *Assume that there exists a closed degree* 0 *bimodule map $f\colon \mathcal{A}_m(-,-) \to \mathcal{A}_m(-,\tau(-))$ such that $f\colon \mathcal{A}_m(X^i(E), X^j(E')) \to \mathcal{A}_m(X^i(E), X^{j+1}(E'))$ is a quasi-isomorphism for every $i < j$ and $E, E' \in \mathcal{E}$. Then there is a quasi-equivalence of Adams-graded A_∞-categories*

$$\mathrm{MT}(\tau) \simeq \mathcal{A}_m^0 \oplus (t_m\mathbb{F}[t_m] \otimes \mathcal{A}_m[f(\mathrm{units})^{-1}]^0).$$

Remark (1) In Ganatra's work [22], the chain complex of $\mathcal{A}$-bimodule maps from the diagonal bimodule $\mathcal{A}(-,-)$ to some $\mathcal{A}$-bimodule $\mathcal{B}$ is called the two-pointed complex for Hochschild cohomology of $\mathcal{A}$ with coefficients in $\mathcal{B}$. According to [22, Proposition 2.5], this complex is quasi-isomorphic to the (ordinary) Hochschild cochain complex of $\mathcal{A}$ with coefficients in $\mathcal{B}$. In particular, the bimodule map f in Theorem A defines a class in the Hochschild cohomology of $\mathcal{A}_m$ with coefficients in $\mathcal{A}_m(-,\tau(-))$.

(2) The A_∞-category which computes the mapping torus in Theorem A is very similar to the categories studied by Seidel in [35], with main difference the presence of curvature in Seidel's setting.

(3) The use of the functor $\mathcal{C} \mapsto \mathcal{C}_m$ in Theorem A is not of any deep importance. It was convenient for us to introduce it here for our application to Legendrian contact homology (see Theorem B).

Chekanov–Eliashberg DG-algebra In Section 3, we recall the definition and functorial properties of the Chekanov–Eliashberg DG-category associated to a family of Legendrians in a hypertight contact manifold.

Legendrian lifts of exact Lagrangians in the circular contactization In Section 4, we start with a family

$$\boldsymbol{L} = (L(E))_{E \in \mathcal{E}}, \quad \mathcal{E} = \{1, \dots, N\},$$

of mutually transverse compact connected exact Lagrangian submanifolds in a Liouville manifold (P,λ), and we study a Legendrian lift of $\boldsymbol{L}$ in the circular contactization $(S^1\times P, \ker(d\theta-\lambda))$. More precisely, we assume[2] that there are primitives $f_E\colon L(E)\to\mathbb{R}$ of $\lambda|_{L(E)}$ such that $0\le f_1<\cdots<f_N\le\frac{1}{2}$, and we consider the family of Legendrians

$$\boldsymbol{\Lambda}^{\circ} := (\Lambda^{\circ}(E))_{E\in\mathcal{E}}, \quad \text{where } \Lambda^{\circ}(E)=\{(f_E(x),x)\in(\mathbb{R}/\mathbb{Z})\times P \mid x\in L(E)\}.$$

We denote by $\mathrm{CE}(\boldsymbol{\Lambda}^{\circ})$ the Chekanov–Eliashberg category of $\boldsymbol{\Lambda}^{\circ}$, by $\mathcal{F}uk(\boldsymbol{L})$ the full subcategory of $\mathcal{F}uk(P)$ (see for example [36, Chapter 2]) with objects the Lagrangians $L(E)$, and by $\overrightarrow{\mathcal{F}uk}(\boldsymbol{L})$ its directed subcategory (see [36, Paragraph (5n)]).

In order for the latter algebraic objects to be $\mathbb{Z}$-graded, we assume that $H_1(P)$ is free, that the first Chern class of P (equipped with any almost complex structure compatible with $(-d\lambda)$) is 2-torsion, and that the Maslov class of the Lagrangians $L(E)$ vanish. As explained in Section 3.1, the grading on $\mathrm{CE}(\boldsymbol{\Lambda}^{\circ})$ depends on the choice of a symplectic trivialization of the contact structure along a fiber $h_0=S^1\times\{a_0\}$. We denote by $\mathrm{CE}^{r}_{-*}(\boldsymbol{\Lambda}^{\circ})$ the Chekanov–Eliashberg DG-category of $\boldsymbol{\Lambda}^{\circ}$ with grading induced by the trivialization

$$(\xi^{\circ}|_{h_0}, d\alpha^{\circ}) \xrightarrow{\sim} (h_0\times\mathbb{C}^n, dx\wedge dy), \quad ((\theta,a_0),(\lambda_{a_0}(v),v))\mapsto((\theta,a_0),e^{2i\pi r\theta}\psi(v)),$$

where $\psi\colon (T_{a_0}P,-d\lambda_{a_0})\xrightarrow{\sim}(\mathbb{C}^n,dx\wedge dy)$ is a symplectic isomorphism.

In this setting, $\mathrm{CE}^{r}_{-*}(\boldsymbol{\Lambda}^{\circ})$ is augmented (with the trivial augmentation) and Adams-graded (by the number of times a Reeb chord winds around the fiber). As above, we denote by $\mathbb{F}[t_m]$ the augmented Adams-graded associative algebra generated by a variable t_m of bidegree $(m,1)$. Moreover, we denote by $E(-)=B(-)^{\#}$ (graded dual of bar construction) the Koszul dual functor (see work by Lu, Palmieri, Wu and Zhang [29, Section 2] or Ekholm and Lekili [17, Section 2.3]). We say that Koszul duality holds for an augmented Adams-graded A_∞-category A if the natural map $A\to E(E(A))$ is a quasi-isomorphism (see [29, Theorem 2.4] or [17, Definition 17]).

Theorem B *Koszul duality holds for* $\mathrm{CE}^{r}_{-*}(\boldsymbol{\Lambda}^{\circ})$*, and there is a quasi-equivalence of augmented Adams-graded* A_∞*-categories*

$$E(\mathrm{CE}^{r}_{-*}(\boldsymbol{\Lambda}^{\circ}))\simeq\overrightarrow{\mathcal{F}uk}(\boldsymbol{L})\oplus(t_{2r}\mathbb{F}[t_{2r}]\otimes\mathcal{F}uk(\boldsymbol{L})).$$

Remark Koszul duality has many important consequences, see for example [29] or [17]. In particular, Theorem B implies that there is a quasi-equivalence of augmented Adams-graded DG-categories

$$\mathrm{CE}^{r}_{-*}(\boldsymbol{\Lambda}^{\circ})\simeq E\big(\overrightarrow{\mathcal{F}uk}(\boldsymbol{L})\oplus(t_{2r}\mathbb{F}[t_{2r}]\otimes\mathcal{F}uk(\boldsymbol{L}))\big).$$

Observe that in the particular case when the Lagrangians are spheres, this formula is closely related to Conjecture 6.3 in [35], which was also discussed by Ganatra and Maydanskiy in the appendix of [5].

[2]This can always be achieved by applying the Liouville flow in backwards time.

We now give a corollary of the latter result. If B is a (unpointed) space, we consider its one-point compactification B^* and view it as a pointed space (with basepoint the point at infinity). If moreover X is a pointed space, we consider the half-smash product of B and X,

$$X \rtimes B := X \wedge B^*$$

(where $\wedge$ denotes the smash product of pointed spaces). Finally, if Y is a pointed space, we denote by ΩY its based loop space.

Corollary *If L is a connected compact exact Lagrangian and Λ° is a Legendrian lift of L in the circular contactization, then there is a quasi-equivalence of augmented DG-algebras*

$$\mathrm{CE}^1_{-*}(\Lambda^\circ) \simeq C_{-*}(\Omega(\mathbb{CP}^\infty \rtimes L)).$$

Acknowledgments This work is part of my PhD thesis that I did at Nantes Université under the supervision of Paolo Ghiggini and Vincent Colin, who I thank for their guidance and support. I also thank Baptiste Chantraine, Georgios Dimitroglou Rizell and Tobias Ekholm for helpful discussions. Finally, I thank the anonymous referee for suggestions that helped improve the exposition of the paper. The work presented in this article is supported by Villum Fonden, Villum Investigator grant 37814.

1 Algebra

In the following, $\mathbb{F}$ denotes the field $\mathbb{Z}/2\mathbb{Z}$. Vector spaces are always over $\mathbb{F}$.

Definition 1.1 An A_∞-category $\mathcal{A}$ is the data of

(1) a collection of objects $\operatorname{ob}\mathcal{A}$,

(2) for every objects X, Y, a graded vector space of morphisms $\mathcal{A}(X, Y)$,

(3) a family of degree $2-d$ linear maps

$$\mu^d : \mathcal{A}(X_0, X_1) \otimes \cdots \otimes \mathcal{A}(X_{d-1}, X_d) \to \mathcal{A}(X_0, X_d)$$

indexed by the sequences of objects $(X_0, \dots, X_d)$, $d \geq 1$, such that

$$\sum_{0 \leq i < j \leq d} \mu^{d-(j-i)+1} \circ (\mathbf{1}^i \otimes \mu^{j-i} \otimes \mathbf{1}^{d-j}) = 0,$$

for all $d \geq 1$.

Definition 1.2 An A_∞-cocategory $\mathcal{C}$ is the data of

(1) a collection of objects $\operatorname{ob}\mathcal{C}$,

(2) for every objects X, Y, a graded vector space of morphisms $\mathcal{C}(X, Y)$,

(3) a family of degree $2-d$ linear maps

$$\delta^d : \mathcal{C}(X_0, X_d) \to \bigoplus_{d \geq 1} \bigoplus_{X_1, \dots, X_{d-1}} \mathcal{C}(X_0, X_1) \otimes \cdots \otimes \mathcal{C}(X_{d-1}, X_d)$$

indexed by the sequences of objects $(X_0, \dots, X_d)$, $d \geq 1$, such that

- for all $d \geq 1$,
$$\sum_{0 \leq i < j \leq d} (\mathbf{1}^i \otimes \delta^{j-i} \otimes \mathbf{1}^{d-j}) \circ \delta^{d-(j-i)+1} = 0,$$
- the map
$$C \to \prod_{d \geq 1} C^{\otimes d}, \quad x \mapsto (\delta^d(x))_{d \geq 1},$$
factors through the inclusion $\bigoplus_{d \geq 1} C^{\otimes d} \to \prod_{d \geq 1} C^{\otimes d}$.

Remark If $\mathcal{E}$ is some set, denote by $\mathbb{F}_{\mathcal{E}}$ the semisimple algebra over $\mathbb{F}$ generated by elements e_X, $X \in \mathcal{E}$, such that
$$e_X \cdot e_Y = \begin{cases} e_X & \text{if } X = Y, \\ 0 & \text{if } X \neq Y. \end{cases}$$
To any A_∞-category $\mathcal{A}$ with $\mathrm{ob}(\mathcal{A}) = \mathcal{E}$, we can associate an A_∞-algebra over $\mathbb{F}_{\mathcal{E}}$ where

- the underlying graded vector space is $\bigoplus_{X, Y \in \mathcal{E}} \mathcal{A}(X, Y)$,
- given $x \in \mathcal{A}(X_0, Y_0)$,
$$e_X \cdot x = \begin{cases} x & \text{if } X = X_0, \\ 0 & \text{if } X \neq X_0, \end{cases} \qquad \text{and} \qquad x \cdot e_Y = \begin{cases} x & \text{if } Y = Y_0, \\ 0 & \text{if } Y \neq Y_0, \end{cases}$$
- operations are the same as on $\mathcal{A}$.

Conversely, to any A_∞-algebra over $\mathbb{F}_{\mathcal{E}}$, one can associate an A_∞-category with $\mathrm{ob}(\mathcal{A}) = \mathcal{E}$. Note that the above discussion also applies to A_∞-cocategories. As a result, the theory of A_∞-(co)categories with $\mathcal{E}$ as set of objects is equivalent to the theory of A_∞-(co)algebras over $\mathbb{F}_{\mathcal{E}}$.

In this paper, we will appeal to several standard notions in the theory of A_∞-(co)categories that we choose not to recall: instead, we list them and give corresponding references.

- For A_∞-(co)maps, (co)augmentations and (co)bar, graded dual, Koszul dual constructions, see [17, Section 2] (where everything is written in the language of A_∞-(co)algebras over $\mathbb{F}_{\mathcal{E}}$).
- For general definitions and results about A_∞-categories (in particular about homotopy between A_∞-functors, homological perturbation theory, directed (sub)categories and twisted complexes), see [36, Chapter 1].
- For quotient of A_∞-categories, see [30], and for localization of A_∞-categories, see [23, Section 3.1.3].

Finally, we will use the following notion.

Definition 1.3 An Adams-graded vector space is a $\mathbb{Z}\times\mathbb{Z}$-graded vector space: if x is an element in the (i, j) component, we say that i is the cohomological degree of x, and j is the Adams degree of x. An Adams-graded A_∞-(co)category is an A_∞-(co)category enriched over Adams-graded vector spaces, where the operations are required to be of degree 0 with respect to the Adams grading. See [29] for a treatment of Koszul duality in the context of Adams-graded A_∞-algebras.

1.1 Modules over A_∞-categories

Let $\mathcal{C},\mathcal{D}$ be two A_∞-categories, and let $\mathcal{A},\mathcal{B}$ be two full subcategories of $\mathcal{C},\mathcal{D}$, respectively.

Definition 1.4 A $(\mathcal{C},\mathcal{D})$-bimodule $\mathcal{M}$ consists of the following data:

(1) for every pair $(X,Y)\in\mathrm{ob}(\mathcal{C})\times\mathrm{ob}(\mathcal{D})$, a graded vector space $\mathcal{M}(X,Y)$,

(2) a family of degree $1-p-q$ linear maps

$$\mu_{\mathcal{M}}\colon\mathcal{C}(X_0,X_1)\otimes\cdots\otimes\mathcal{C}(X_{p-1},X_p)\otimes\mathcal{M}(X_p,Y_q)\otimes\mathcal{D}(Y_q,Y_{q-1})\otimes\cdots\otimes\mathcal{D}(Y_1,Y_0)\to\mathcal{M}(X_0,Y_0)$$

indexed by the sequences

$$(X_0,\dots,X_p,Y_0,\dots,Y_q)\in\mathrm{ob}(\mathcal{C})^{p+1}\times\mathrm{ob}(\mathcal{D})^{q+1},$$

which satisfy the relations

$$\sum\mu_{\mathcal{M}}(\dots,\mu_{\mathcal{C}}(\dots),\dots,u\dots)+\sum\mu_{\mathcal{M}}(\dots,\mu_{\mathcal{M}}(\dots,u,\dots),\dots)\\+\sum\mu_{\mathcal{M}}(\dots,u,\dots,\mu_{\mathcal{D}}(\dots),\dots)=0.$$

A degree s morphism $t\colon\mathcal{M}_1\to\mathcal{M}_2$ between two $(\mathcal{C},\mathcal{D})$-bimodules consists of a family of degree $s-p-q$ linear maps

$$t\colon\mathcal{C}(X_0,X_1)\otimes\cdots\otimes\mathcal{C}(X_{p-1},X_p)\otimes\mathcal{M}_1(X_p,Y_q)\otimes\mathcal{D}(Y_q,Y_{q-1})\otimes\cdots\otimes\mathcal{D}(Y_1,Y_0)\to\mathcal{M}_2(X_0,Y_0)$$

indexed by the sequences

$$(X_0,\dots,X_p,Y_0,\dots,Y_q)\in\mathrm{ob}(\mathcal{C})^{p+1}\times\mathrm{ob}(\mathcal{D})^{q+1}.$$

The differential of such a morphism is defined by

$$\mu^1_{\mathrm{Mod}_{\mathcal{C},\mathcal{D}}}(t)(\dots,u,\dots)=\sum t(\dots,\mu_{\mathcal{C}}(\dots),\dots,u,\dots)+\sum t(\dots,\mu_{\mathcal{M}_1}(\dots,u,\dots),\dots)\\+\sum t(\dots,u,\dots,\mu_{\mathcal{D}}(\dots),\dots)+\sum\mu_{\mathcal{M}_2}(\dots,t(\dots,u,\dots),\dots).$$

Finally, the composition of $t_1\colon\mathcal{M}_1\to\mathcal{M}_2$ and $t_2\colon\mathcal{M}_2\to\mathcal{M}_3$ is such that

$$\mu^2_{\mathrm{Mod}_{\mathcal{C}}}(t_1,t_2)(\dots,u,\dots)=\sum t_2(\dots,t_1(\dots,u,\dots),\dots).$$

We denote by $\mathrm{Mod}_{\mathcal{C},\mathcal{D}}$ the DG-category of $(\mathcal{C},\mathcal{D})$-bimodules.

Definition 1.5 Let $\Phi_1,\Phi_2\colon\mathcal{C}\to\mathcal{D}$ be two A_∞-functors. Then there is a $\mathcal{C}$-bimodule $\mathcal{D}(\Phi_1(-),\Phi_2(-))$ defined as follows:

(1) On objects, it sends (X_1,X_2) to $\mathcal{D}(\Phi_1X_1,\Phi_2X_2)$.

(2) On morphisms, it sends a sequence $(\dots,y,\dots)$ in

$$\mathcal{C}(X_0,X_1)\times\cdots\times\mathcal{C}(X_{p-1},X_p)\times\mathcal{D}(\Phi_1X_p,\Phi_2X_{p+1})\times\mathcal{C}(X_{p+1},X_{p+2})\times\cdots\times\mathcal{C}(X_{p+q},X_{p+q+1})$$

to

$$\mu_{\mathcal{D}(\Phi_1(-),\Phi_2(-))}(\dots,y,\dots)=\sum\mu_{\mathcal{D}}\big(\Phi_1(\dots),\dots,\Phi_1(\dots),y,\Phi_2(\dots),\dots,\Phi_2(\dots)\big).$$

In the following, we will focus on *left* $\mathcal{C}$-modules, which correspond to $(\mathcal{C}, \mathbb{F})$-bimodules. We denote by $\mathrm{Mod}_{\mathcal{C}}$ the DG-category of (left) $\mathcal{C}$-modules.

Definition 1.6 Let $t\colon \mathcal{M}_1 \to \mathcal{M}_2$ be a degree 0 closed $\mathcal{C}$-module map. We say that t is a quasi-isomorphism if the induced chain map $t\colon \mathcal{M}_1(X) \to \mathcal{M}_2(X)$ is a quasi-isomorphism for every object X in $\mathcal{C}$. (See [24, Section A.2] for a discussion on quasi-isomorphisms between A_∞-modules.)

Definition 1.7 Let $t, t'\colon \mathcal{M}_1 \to \mathcal{M}_2$ be two degree 0 closed morphisms of $\mathcal{C}$-modules. A homotopy between t and t' is a $\mathcal{C}$-module map $h\colon \mathcal{M}_1 \to \mathcal{M}_2$ such that

$$t + t' = \mu^1_{\mathrm{Mod}_{\mathcal{C}}}(h).$$

Definition 1.8 (see [36, Paragraph (1l); 24, Section A.1]) There is an A_∞-functor

$$\mathcal{C} \to \mathrm{Mod}_{\mathcal{C}}, \quad Y \mapsto \mathcal{C}(-, Y),$$

called the Yoneda A_∞-functor, defined as follows. For every object X,

$$\mathcal{C}(-, Y)(X) = \mathcal{C}(X, Y).$$

Also, a sequence

$$(x_0, \dots, x_{d-1}) \in \mathcal{C}(X_0, X_1) \times \cdots \times \mathcal{C}(X_{d-1}, X_d)$$

acts on an element u in $\mathcal{C}(X_d, Y)$ via the operations

$$\mu_{\mathcal{C}(-,Y)}(x_0, \dots, x_{d-1}, u) = \mu_{\mathcal{C}}(x_0, \dots, x_{d-1}, u).$$

Finally, let

$$\boldsymbol{y} = (y_0, \dots, y_{p-1}) \in \mathcal{C}(Y_0, Y_1) \times \cdots \times \mathcal{C}(Y_{p-1}, Y_p)$$

be a sequence of morphisms in $\mathcal{C}$. Then the Yoneda functor gives a morphism of $\mathcal{C}$-modules

$$t_{\boldsymbol{y}} : \mathcal{C}(-, Y_0) \to \mathcal{C}(-, Y_p)$$

which sends every sequence $(x_0, \dots, x_{d-1}, u)$ as above to

$$\mu_{\mathcal{C}}(x_0, \dots, x_{d-1}, u, y_0, \dots, y_{p-1}) \in \mathcal{C}(X_0, Y_p).$$

We have the following important result.

Proposition 1.9 (Yoneda lemma) *The Yoneda A_∞-functor*

$$\mathcal{C} \to \mathrm{Mod}_{\mathcal{C}}, \quad Y \mapsto \mathcal{C}(-, Y),$$

is cohomologically full and faithful.

Proof This is Lemma 2.12 in [36], and also Lemma A.1 in [24]. □

The Yoneda lemma has the following easy consequence. We state it for future reference.

Corollary 1.10 *Every closed $\mathcal{C}$-module map $f : \mathcal{C}(-, X) \to \mathcal{C}(-, Y)$ is homotopic to the $\mathcal{C}$-module map $t_{f(e_X)}$ induced by $f(e_X) \in \mathcal{C}(X, Y)$. (see Definition 1.8).*

Proof According to the Yoneda lemma, f is homotopic to t_x for some closed x in $\mathcal{C}(X,Y)$. Thus, there exists a $\mathcal{C}$-module map $h\colon \mathcal{C}(-,X)\to\mathcal{C}(-,Y)$ such that

$$f = t_x + \mu^1_{\mathrm{Mod}_{\mathcal{C}}}(h).$$

Evaluating the latter relation at the unit $e_X \in \mathcal{C}(X,X)$ gives

$$f(e_X) = x + \mu^1_{\mathcal{C}}(he_X).$$

Therefore, x is homotopic to $f(e_X)$, and this implies that t_x is homotopic to $t_{f(e_X)}$ by the Yoneda lemma. Finally, we have that f is homotopic to $t_{f(e_X)}$. □

Pullback of A_∞-modules

Definition 1.11 (see [36, Paragraph (1k)]) Let $\Phi\colon \mathcal{C}\to\mathcal{D}$ be an A_∞-functor. Then there is a DG-functor

$$\Phi^*\colon \mathrm{Mod}_{\mathcal{D}} \to \mathrm{Mod}_{\mathcal{C}}, \quad \mathcal{N}\mapsto \Phi^*\mathcal{N},$$

defined as follows. Let $\mathcal{N}$ be a $\mathcal{D}$-module. For every object X,

$$\Phi^*\mathcal{N}(X) = \mathcal{N}(\Phi X).$$

Also, a sequence

$$(x_0,\dots,x_{d-1}) \in \mathcal{C}(X_0,X_1)\times\cdots\times\mathcal{C}(X_{d-1},X_d)$$

acts on an element $u\in\Phi^*\mathcal{N}(X_d)$ via the operations

$$\mu_{\Phi^*\mathcal{N}}(x_0,\dots,x_{d-1},u) = \sum \mu_{\mathcal{N}}(\Phi(x_0,\dots,x_{i_1-1}),\dots,\Phi(x_{d-i_r},\dots,x_{d-1}),u).$$

Finally, let $t\colon\mathcal{N}_1\to\mathcal{N}_2$ be a $\mathcal{D}$-module map. Then the above functor gives a $\mathcal{C}$-module map

$$\Phi^*t\colon \Phi^*\mathcal{N}_1\to\Phi^*\mathcal{N}_2$$

which sends every sequence $(x_0,\dots,x_{d-1},u)$ as above to

$$\Phi^*t(x_0,\dots,x_{d-1},u) = \sum t(\Phi(x_0,\dots,x_{i_1-1}),\dots,\Phi(x_{d-i_r},\dots,x_{d-1}),u).$$

Remark 1.12 Let $\Phi\colon\mathcal{C}\to\mathcal{D}$ be an A_∞-functor, and let $\Psi\colon\mathcal{D}\to\mathcal{E}$ be another A_∞-functor towards a third A_∞-category $\mathcal{E}$. Then $\Phi^*\circ\Psi^* = (\Psi\circ\Phi)^*$ as DG-functors.

Definition 1.13 Let Y be an object of $\mathcal{C}$, and let $\Phi\colon\mathcal{C}\to\mathcal{D}$ be an A_∞-functor. Then there is a degree 0 closed $\mathcal{C}$-module map $t_\Phi\colon\mathcal{C}(-,Y)\to\Phi^*\mathcal{D}(-,\Phi(Y))$ which sends any sequence

$$(x_0,\dots,x_{d-1},u)\in\mathcal{C}(X_0,X_1)\times\cdots\times\mathcal{C}(X_{d-1},X_d)\times\mathcal{C}(X_d,Y)$$

to

$$t_\Phi(x_0,\dots,x_{d-1},u) = \Phi(x_0,\dots,x_{d-1},u)\in\mathcal{D}(\Phi X_0,\Phi Y).$$

Quotient of A_∞-modules

Definition 1.14 (see [23, Section 3.1.3]) There is a DG-functor

$$\mathrm{Mod}_{\mathcal{C}} \to \mathrm{Mod}_{\mathcal{C}/\mathcal{A}}, \quad \mathcal{M} \mapsto {}_{\mathcal{A}\backslash}\mathcal{M},$$

defined as follows. Let $\mathcal{M}$ be a $\mathcal{C}$-module. For every object X,

$${}_{\mathcal{A}\backslash}\mathcal{M}(X) = \mathcal{M}(X) \oplus \Biggl(\bigoplus_{\substack{p\geq 1\\ A_1,\dots,A_p \in \mathcal{A}}} \mathcal{C}(X, A_1)[1] \otimes \cdots \otimes \mathcal{C}(A_{p-1}, A_p)[1] \otimes \mathcal{M}(A_p) \Biggr).$$

Also, a sequence

$$\boldsymbol{x}_i = (x_i^0, \dots, x_i^{p_i-1}) \in \mathcal{C}/\mathcal{A}(X_i, X_{i+1}) \quad (0 \leq i \leq d-1)$$

acts on an element

$$\boldsymbol{u} = (x_d^0, \dots, x_d^{p_d-1}, u) \in {}_{\mathcal{A}\backslash}\mathcal{M}(X_d)$$

via the operations

$$\begin{aligned}\mu_{\mathcal{A}\backslash\mathcal{M}}(\boldsymbol{x}_0, \dots, \boldsymbol{x}_{d-1}, \boldsymbol{u}) &= \sum_{\substack{0\leq i\leq p_0,\, 1\leq j\leq p_d\\ i<j \text{ if } d=0}} x_0^0 \otimes \cdots \otimes x_0^{i-1} \otimes \mu_{\mathcal{C}}(x_0^i, \dots, x_d^{j-1}) \otimes x_d^j \otimes \cdots \otimes x_d^{p_d-1} \otimes u \\ &\qquad + \sum_{0\leq i\leq p_0} x_0^0 \otimes \cdots \otimes x_0^{i-1} \otimes \mu_{\mathcal{M}}(x_0^i, \dots, x_d^{p_d-1}, u).\end{aligned}$$

Finally, let $t\colon \mathcal{M}_1 \to \mathcal{M}_2$ be a $\mathcal{C}$-module map. Then the above functor gives a $\mathcal{C}/\mathcal{A}$-module map ${}_{\mathcal{A}\backslash}t\colon {}_{\mathcal{A}\backslash}\mathcal{M}_1 \to {}_{\mathcal{A}\backslash}\mathcal{M}_2$ which sends every sequence $(\boldsymbol{x}_0, \dots, \boldsymbol{x}_{d-1}, \boldsymbol{u})$ as above to

$${}_{\mathcal{A}\backslash}t(\boldsymbol{x}_0, \dots, \boldsymbol{x}_{d-1}, \boldsymbol{u}) = \sum_{0\leq i\leq p_0} x_0^0 \otimes \cdots \otimes x_0^{i-1} \otimes t(x_0^i, \dots, x_d^{p_d-1}, u).$$

Relations between pullback and quotient of A_∞-modules

Definition 1.15 Let $\Phi\colon \mathcal{C} \to \mathcal{D}$ be an A_∞-functor such that $\Phi(\mathcal{A})$ is contained in $\mathcal{B}$, and let X be a fixed object of $\mathcal{C}$. Then, for each $\mathcal{D}$-module $\mathcal{N}$, there is a chain map ${}_{\mathcal{A}\backslash}(\Phi^*\mathcal{N})(X) \to {}_{\mathcal{B}\backslash}\mathcal{N}(\Phi X)$ which sends an element

$$\boldsymbol{u} = (x^0, \dots, x^{p-1}, u) \in {}_{\mathcal{A}\backslash}(\Phi^*\mathcal{N})(X)$$

to

$$\sum \Phi(x^0, \dots, x^{i_1-1}) \otimes \cdots \otimes \Phi(x^{i_r}, \dots, x^{p-1}) \otimes u \in {}_{\mathcal{B}\backslash}\mathcal{N}(\Phi X).$$

This defines a natural transformation between the functors $\mathcal{N} \mapsto {}_{\mathcal{A}\backslash}(\Phi^*\mathcal{N})(X)$ and $\mathcal{N} \mapsto {}_{\mathcal{B}\backslash}\mathcal{N}(\Phi X)$ from $\mathrm{Mod}_{\mathcal{D}}$ to Ch. In other words, for every $\mathcal{D}$-module map $t\colon \mathcal{N}_1 \to \mathcal{N}_2$, the following diagram of chain complexes commutes:

$$\begin{tikzcd} {}_{\mathcal{A}\backslash}(\Phi^*\mathcal{N}_1)(X) \arrow[r] \arrow[d, "{}_{\mathcal{A}\backslash}(\Phi^* t)"] & {}_{\mathcal{B}\backslash}\mathcal{N}_1(\Phi X) \arrow[d, "{}_{\mathcal{B}\backslash}t"] \\ {}_{\mathcal{A}\backslash}(\Phi^*\mathcal{N}_2)(X) \arrow[r] & {}_{\mathcal{B}\backslash}\mathcal{N}_2(\Phi X) \end{tikzcd}$$

Remark 1.16 Let Y be an object of $\mathcal{C}$, and let $\Phi\colon \mathcal{C} \to \mathcal{D}$ be an A_∞-functor such that $\Phi(\mathcal{A})$ is contained in $\mathcal{B}$. Let $\widetilde{\Phi}\colon \mathcal{C}/\mathcal{A} \to \mathcal{D}/\mathcal{B}$ be the A_∞-functor induced by Φ (see [30, Section 3]). Localize the morphism $t_\Phi\colon \mathcal{C}(-,Y) \to \Phi^*\mathcal{D}(-,\Phi Y)$ of Definition 1.13 at $\mathcal{A}$ and evaluate at X to get a chain map

$$\mathcal{C}/\mathcal{A}(X,Y) = {}_{\mathcal{A}\backslash}\mathcal{C}(-,Y)(X) \xrightarrow{\mathcal{A}\backslash t_\Phi} {}_{\mathcal{A}\backslash}(\Phi^*\mathcal{D}(-,\Phi Y))(X).$$

Then the composition of this map with the chain map

$${}_{\mathcal{A}\backslash}(\Phi^*\mathcal{D}(-,\Phi Y))(X) \to {}_{\mathcal{B}\backslash}\mathcal{D}(-,\Phi Y)(\Phi X) = \mathcal{D}/\mathcal{B}(\Phi X,\Phi Y)$$

of Definition 1.15 is the chain map $\widetilde{\Phi}\colon \mathcal{C}/\mathcal{A}(X,Y) \to \mathcal{D}/\mathcal{B}(\Phi X,\Phi Y)$.

Proposition 1.17 *Let $\Phi\colon \mathcal{C}_1 \to \mathcal{C}_2$ be an A_∞-functor such that $\Phi(\mathcal{A}_1)$ is contained in $\mathcal{A}_2$, and let $\widetilde{\Phi}\colon \mathcal{C}_1/\mathcal{A}_1 \to \mathcal{C}_2/\mathcal{A}_2$ be the A_∞-functor induced by Φ.*

Let Y_1 be an object of $\mathcal{C}_1$ and set $Y_2 := \Phi(Y_1)$. Assume that there exists a $\mathcal{C}_i$-module $\mathcal{M}_{\mathcal{C}_i}$, a degree 0 closed $\mathcal{C}_i$-module map $t_{\mathcal{C}_i}\colon \mathcal{C}_i(-,Y_i) \to \mathcal{M}_{\mathcal{C}_i}$ and a degree 0 closed $\mathcal{C}_1$-module map $t_0\colon \mathcal{M}_{\mathcal{C}_1} \to \Phi^*\mathcal{M}_{\mathcal{C}_2}$ such that the following diagram of $\mathcal{C}_1$-modules commutes:

$$\begin{tikzcd} \mathcal{C}_1(-,Y_1) \arrow[r, "t_\Phi"] \arrow[d, "t_{\mathcal{C}_1}"] & \Phi^*\mathcal{C}_2(-,Y_2) \arrow[d, "\Phi^* t_{\mathcal{C}_2}"] \\ \mathcal{M}_{\mathcal{C}_1} \arrow[r, "t_0"] & \Phi^*\mathcal{M}_{\mathcal{C}_2} \end{tikzcd}$$

(*see Definition 1.13 for the map t_Φ*). *Then for every object X in $\mathcal{C}_1$, there is a chain map*

$$u\colon {}_{\mathcal{A}_1\backslash}\mathcal{M}_{\mathcal{C}_1}(X) \to {}_{\mathcal{A}_2\backslash}\mathcal{M}_{\mathcal{C}_2}(\Phi X)$$

such that the following diagram of chain complexes commutes:

$$\begin{tikzcd} \mathcal{C}_1/\mathcal{A}_1(X,Y_1) \arrow[r, "\widetilde{\Phi}"] \arrow[d, "\mathcal{A}_1\backslash t_{\mathcal{C}_1}"] & \mathcal{C}_2/\mathcal{A}_2(\Phi X,Y_2) \arrow[d, "\mathcal{A}_2\backslash t_{\mathcal{C}_2}"] \\ {}_{\mathcal{A}_1\backslash}\mathcal{M}_{\mathcal{C}_1}(X) \arrow[r, "u"] & {}_{\mathcal{A}_2\backslash}\mathcal{M}_{\mathcal{C}_2}(\Phi X) \\ \mathcal{M}_{\mathcal{C}_1}(X) \arrow[r, "t_0"] \arrow[u, hook] & \mathcal{M}_{\mathcal{C}_2}(\Phi X) \arrow[u, hook] \end{tikzcd}$$

(*the two lowest vertical maps are the inclusions*). *If*, *moreover*,

1. *for every objects A in $\mathcal{A}_i$, the complexes $\mathcal{M}_{\mathcal{C}_i}(A)$ are acyclic*,
2. *the maps ${}_{\mathcal{A}_i\backslash}t_{\mathcal{C}_i}\colon {}_{\mathcal{A}_i\backslash}\mathcal{C}_i(X,Y_i) \to {}_{\mathcal{A}_i\backslash}\mathcal{M}_{\mathcal{C}_i}(X)$ are quasi-isomorphisms*, *and*
3. *the map $t_0\colon \mathcal{M}_{\mathcal{C}_1}(X) \to \Phi^*\mathcal{M}_{\mathcal{C}_2}(X)$ is a quasi-isomorphism*,

then the map $\widetilde{\Phi}\colon \mathcal{C}_1/\mathcal{A}_1(X,Y_1) \to \mathcal{C}_2/\mathcal{A}_2(\Phi X,Y_2)$ is a quasi-isomorphism.

Proof We apply the functor $\mathcal{P} \mapsto {}_{\mathcal{A}_1\backslash}\mathcal{P}$ to the first diagram, we evaluate at X and we use the natural map of Definition 1.15 to get the commutative diagram of chain complexes

$$\begin{tikzcd} {}_{\mathcal{A}_1\backslash}\mathcal{C}_1(-,Y_1)(X) \arrow[r, "{}_{\mathcal{A}_1\backslash}t_\Phi"] \arrow[d, "{}_{\mathcal{A}_1\backslash}t_{\mathcal{C}_1}"] & {}_{\mathcal{A}_1\backslash}(\Phi^*\mathcal{C}_2(-,Y_2))(X) \arrow[r] \arrow[d, "{}_{\mathcal{A}_1\backslash}(\Phi^* t_{\mathcal{C}_2})"] & {}_{\mathcal{A}_2\backslash}\mathcal{C}_2(-,Y_2)(\Phi X) \arrow[d, "{}_{\mathcal{A}_2\backslash}t_{\mathcal{C}_2}"] \\ {}_{\mathcal{A}_1\backslash}\mathcal{M}_{\mathcal{C}_1}(X) \arrow[r, "{}_{\mathcal{A}_1\backslash}t_0"] & {}_{\mathcal{A}_1\backslash}(\Phi^*\mathcal{M}_{\mathcal{C}_2})(X) \arrow[r] & {}_{\mathcal{A}_2\backslash}\mathcal{M}_{\mathcal{C}_2}(\Phi X) \end{tikzcd}$$

Then we compose the horizontal maps and we use Remark 1.16 to get a commutative diagram of chain complexes

$$\begin{tikzcd} \mathcal{C}_1/\mathcal{A}_1(X,Y_1) \arrow[r, "\widetilde{\Phi}"] \arrow[d, "{}_{\mathcal{A}_1\backslash}t_{\mathcal{C}_1}"] & \mathcal{C}_2/\mathcal{A}_2(\Phi X,Y_2) \arrow[d, "{}_{\mathcal{A}_2\backslash}t_{\mathcal{C}_2}"] \\ {}_{\mathcal{A}_1\backslash}\mathcal{M}_{\mathcal{C}_1}(X) \arrow[r, "u"] & {}_{\mathcal{A}_2\backslash}\mathcal{M}_{\mathcal{C}_2}(\Phi X) \end{tikzcd}$$

This proves the first part of the proposition because the following diagram of chain complexes commutes:

$$\begin{tikzcd} {}_{\mathcal{A}_1\backslash}\mathcal{M}_{\mathcal{C}_1}(X) \arrow[r, "u"] & {}_{\mathcal{A}_2\backslash}\mathcal{M}_{\mathcal{C}_2}(\Phi X) \\ \mathcal{M}_{\mathcal{C}_1}(X) \arrow[r, "t_0"] \arrow[u, hook] & \mathcal{M}_{\mathcal{C}_2}(\Phi X) \arrow[u, hook] \end{tikzcd}$$

The second part of the proposition follows directly with [23, Lemma 3.13]. □

Cone of module maps

Definition 1.18 Let $t\colon \mathcal{M}_1 \to \mathcal{M}_2$ be a degree 0 closed morphism of $\mathcal{C}$-modules. We denote by

$$\operatorname{Cone}(\mathcal{M}_1 \xrightarrow{t} \mathcal{M}_2) = \begin{bmatrix} \mathcal{M}_1 \\ \big\downarrow t \\ \mathcal{M}_2 \end{bmatrix}$$

the $\mathcal{C}$-module $\mathcal{M}$ defined as follows. For every object X in $\mathcal{C}$,

$$\mathcal{M}(X) = \mathcal{M}_1(X)[1] \oplus \mathcal{M}_2(X)$$

as graded vector space, and any sequence

$$(x_0,\dots,x_{d-1}) \in \mathcal{C}(X_0,X_1)\times\cdots\times\mathcal{C}(X_{d-1},X_d)$$

acts on an element $u_1 \oplus u_2$ in $\mathcal{M}(X_d)$ via the operations

$$\begin{aligned} &\mu_{\mathcal{M}}(x_0,\dots,x_{d-1},u_1\oplus u_2) \\ &\qquad = \mu_{\mathcal{M}_1}(x_0,\dots,x_{d-1},u_1) \oplus \big(\mu_{\mathcal{M}_2}(x_0,\dots,x_{d-1},u_2) + t(x_0,\dots,x_{d-1},u_1)\big). \end{aligned}$$

If we have two $\mathcal{C}$-module maps $t\colon \mathcal{M}_1 \to \mathcal{M}_2$ and $t'\colon \mathcal{M}_1 \to \mathcal{M}_2'$, then we set

$$\left[\begin{tikzcd} & \mathcal{M}_1 \arrow[dl, "t"'] \arrow[dr, "t'"] & \\ \mathcal{M}_2 & & \mathcal{M}_2' \end{tikzcd}\right] := \begin{bmatrix} \mathcal{M}_1 \\ \big\downarrow (t,t') \\ \mathcal{M}_2 \oplus \mathcal{M}_2' \end{bmatrix}.$$

Proposition 1.19 *Consider a diagram of $\mathcal{C}$-modules*

$$\begin{tikzcd} \mathcal{M}_1 \arrow[r, "t_1"] \arrow[d, "t_1'"] & \mathcal{M}_2 \arrow[d, "t_2"] \\ \mathcal{M}_2' \arrow[r, "t_2'"] & \mathcal{M}_3 \end{tikzcd}$$

where all the morphisms are of degree 0 *and closed. Then any homotopy $h\colon \mathcal{M}_1 \to \mathcal{M}_3$ between*

$$t := \mu^2_{\mathrm{Mod}_{\mathcal{C}}}(t_1, t_2) \quad \textit{and} \quad t' := \mu^2_{\mathrm{Mod}_{\mathcal{C}}}(t_1', t_2')$$

induces a degree 0 *closed $\mathcal{C}$-module map*

$$t_h\colon \left[\begin{tikzcd}[column sep=small] & \mathcal{M}_1 \arrow[dl, "t_1"'] \arrow[dr, "t_1'"] & \\ \mathcal{M}_2 & & \mathcal{M}_2' \end{tikzcd}\right] \to \mathcal{M}_3$$

defined by

$$t_h(x_0, \dots, x_{d-1}, u_1 \oplus u_2 \oplus u_2') = h(x_0, \dots, x_{d-1}, u_1) + t_2(x_0, \dots, x_{d-1}, u_2) + t_2'(x_0, \dots, x_{d-1}, u_2').$$

Proof The only thing to check is that $\mu^1_{\mathrm{Mod}_{\mathcal{C}}}(t_h) = 0$, which is straightforward. $\square$

Remark If $t\colon \mathcal{M}_1 \to \mathcal{M}_2$ is a degree 0 closed $\mathcal{C}$-module map, then

$${}_{\mathcal{A}\backslash}\mathrm{Cone}(\mathcal{M}_1 \xrightarrow{t} \mathcal{M}_2) = \mathrm{Cone}({}_{\mathcal{A}\backslash}\mathcal{M}_1 \xrightarrow{\mathcal{A}\backslash t} {}_{\mathcal{A}\backslash}\mathcal{M}_2).$$

1.2 Grothendieck construction and homotopy pushout

An exposition on Grothendieck constructions and homotopy colimits in the context of A_∞-categories can be found in [24, Appendix A]. We recall here definitions and basic facts that will serve us. In this section, A_∞-categories are always assumed to be *strictly unital* (see [36, Paragraph (2a)]).

Definition 1.20 Consider a diagram of A_∞-categories

$$\begin{tikzcd} \mathcal{C} \arrow[r, "\Phi_1"] \arrow[d, "\Phi_2"'] & \mathcal{D}_1 \\ \mathcal{D}_2 & \end{tikzcd}$$

The Grothendieck construction of this diagram is the A_∞-category $\mathcal{G}$ such that:

(1) The set of objects is $\mathrm{ob}(\mathcal{C}) \sqcup \mathrm{ob}(\mathcal{D}_1) \sqcup \mathrm{ob}(\mathcal{D}_2)$.

(2) The space of morphisms between two objects X and Y is given by

$$\mathcal{G}(X, Y) = \begin{cases} \mathcal{C}(X, Y) & \text{if } X, Y \in \mathrm{ob}(\mathcal{C}), \\ \mathcal{D}_i(X, Y) & \text{if } X, Y \in \mathrm{ob}(\mathcal{D}_i), \\ \mathcal{D}_i(\Phi_i X, Y) & \text{if } X \in \mathrm{ob}(\mathcal{C}) \text{ and } Y \in \mathrm{ob}(\mathcal{D}_i), \\ 0 & \text{otherwise.} \end{cases}$$

(3) The operations involving only objects of $\mathcal{C}$, respectively of $\mathcal{D}_i$, are the same as in $\mathcal{C}$, respectively in $\mathcal{D}_i$, and for every sequence

$$(x_0,\dots,x_{p-1},y,z_0,\dots,z_{q-1}) \\ \in \mathcal{C}(X_0,X_1)\otimes\cdots\otimes\mathcal{C}(X_{p-1},X_p)\otimes\mathcal{G}(X_p,Y_0)\otimes\mathcal{D}_i(Y_0,Y_1)\otimes\cdots\otimes\mathcal{D}_i(Y_{q-1},Y_q),$$

we have

$$\mu_{\mathcal{G}}(x_0,\dots,x_{p-1},y,z_0,\dots,z_{q-1}) \\ =\sum\mu_{\mathcal{D}_i}(\Phi_i(x_0,\dots,x_{i_1-1}),\dots,\Phi_i(x_{p-i_r},\dots,x_{p-1}),y,z_0,\dots,z_{q-1}).$$

An *adjacent unit* of $\mathcal{G}$ is any morphism in $\mathcal{G}(X,\Phi_i(X))$ which corresponds to the unit in $\mathcal{D}_i(\Phi_i(X),\Phi_i(X))$. The *homotopy colimit* $\mathcal{H}$ of the above diagram is the localization of $\mathcal{G}$ at its adjacent units.

Proposition 1.21 *Let $\mathcal{G}$ be the Grothendieck construction of a diagram*

$$\begin{tikzcd} \mathcal{C} \arrow[r, "\Phi_1"] \arrow[d, "\Phi_2"'] & \mathcal{D}_1 \\ \mathcal{D}_2 \end{tikzcd}$$

Then any strictly commutative square

$$\begin{tikzcd} \mathcal{C} \arrow[r, "\Phi_1"] \arrow[d, "\Phi_2"'] & \mathcal{D}_1 \arrow[d, "\Psi_1"] \\ \mathcal{D}_2 \arrow[r, "\Psi_2"] & \mathcal{E} \end{tikzcd}$$

induces a functor $\sigma\colon\mathcal{G}\to\mathcal{E}$ defined as follows. On the objects, σ acts on $\mathcal{D}_i$ as Ψ_i, and on $\mathcal{C}$ as $\Psi_1\circ\Phi_1=\Psi_2\circ\Phi_2$; on the morphisms, σ acts on $\mathcal{D}_i$ as Ψ_i, on $\mathcal{C}$ as $\Psi_1\circ\Phi_1=\Psi_2\circ\Phi_2$, and it sends any sequence

$$(x_0,\dots,x_{p-1},y,z_0,\dots,z_{q-1}) \\ \in \mathcal{C}(X_0,X_1)\otimes\cdots\otimes\mathcal{C}(X_{p-1},X_p)\otimes\mathcal{G}(X_p,Y_0)\otimes\mathcal{D}_i(Y_0,Y_1)\otimes\cdots\otimes\mathcal{D}_i(Y_{q-1},Y_q)$$

to

$$\sigma(x_0,\dots,x_{p-1},y,z_0,\dots,z_{q-1}) \\ =\sum\Psi_i\bigl(\Phi_i(x_0,\dots,x_{i_1-1}),\dots,\Phi_i(x_{p-i_r},\dots,x_{p-1}),y,z_0,\dots,z_{q-1}\bigr).$$

Proof This is a straightforward verification. □

Proposition 1.22 [24, Lemma A.5] *A strictly commutative diagram of A_∞-categories*

$$\begin{tikzcd} \mathcal{B}_1 \arrow[d] & \mathcal{A} \arrow[l] \arrow[r] \arrow[d] & \mathcal{B}_2 \arrow[d] \\ \mathcal{D}_1 & \mathcal{C} \arrow[l] \arrow[r] & \mathcal{D}_2 \end{tikzcd}$$

induces an A_∞-functor from the Grothendieck construction of the top line to the Grothendieck construction of the bottom line which preserves adjacent units. If moreover each vertical arrow is a quasi-equivalence, then the induced functor

$$\operatorname{hocolim}\begin{pmatrix}\mathcal{A} \longrightarrow \mathcal{B}_1\\ \big\downarrow \\ \mathcal{B}_2\end{pmatrix} \to \operatorname{hocolim}\begin{pmatrix}\mathcal{C} \longrightarrow \mathcal{D}_1\\ \big\downarrow \\ \mathcal{D}_2\end{pmatrix}$$

is a quasi-equivalence.

Proposition 1.23 *Consider two diagrams of A_∞-categories*

$$\begin{tikzcd} \mathcal{C} \arrow[r, "\Phi_1"] \arrow[d, "\Phi_2"'] & \mathcal{D}_1 \\ \mathcal{D}_2 \end{tikzcd} \quad \textit{and} \quad \begin{tikzcd} \mathcal{C} \arrow[r, "\Psi_1"] \arrow[d, "\Psi_2"'] & \mathcal{D}_1 \\ \mathcal{D}_2 \end{tikzcd}$$

If Φ_i and Ψ_i (for $i \in \{1,2\}$) are homotopic (see [36, Paragraph (1h)]), *then the homotopy colimits of the diagrams above are quasi-equivalent.*

Proof Let $\mathcal{G}_0$ and $\mathcal{G}_1$ be the Grothendieck constructions of the above diagrams.

Let T_i be a homotopy from Φ_i to Ψ_i. This means that

$$\Phi_i+\Psi_i=\sum T_i(\dots,\mu_{\mathcal{C}}(\dots),\dots)+\sum \mu_{\mathcal{D}_i}\big(\Psi_i(\dots),\dots,\Psi_i(\dots),T_i(\dots),\Phi_i(\dots),\dots,\Phi_i(\dots)\big).$$

We consider the functor $\kappa\colon \mathcal{G}_0 \to \mathcal{G}_1$ such that

$$\kappa|_{\mathcal{C}} = \mathrm{id}_{\mathcal{C}}, \quad \kappa|_{\mathcal{D}_i} = \mathrm{id}_{\mathcal{D}_i},$$

and which sends every sequence

$$(\dots,y,\dots) \in \mathcal{C}(X_0,X_1)\times\dots\times\mathcal{C}(X_{p-1},X_p)\times\mathcal{G}_0(X_p,Y_0)\times\mathcal{D}_i(Y_0,Y_1)\times\dots\times\mathcal{D}_i(Y_{q-1},Y_q)$$

to

$$\kappa(\dots,y,\dots)=\sum \mu_{\mathcal{D}_i}(\Psi_i(\dots),\dots,\Psi_i(\dots),T_i(\dots),\Phi_i(\dots),\dots,\Phi_i(\dots),y,\dots)$$

if p is positive, and to

$$\Phi(y,\dots)=\mathrm{id}_{\mathcal{D}_i}(y,\dots)$$

otherwise. Using the facts that Φ_i,Ψ_i are A_∞-functors, that T_i is a homotopy from Φ_i to Ψ_i, and gathering the terms depending on if they contain $T_i^k(\dots)$ or y, we conclude that κ satisfies the A_∞-relations. This proves the result because κ is a quasi-equivalence sending the adjacent units of $\mathcal{G}_0$ onto those of $\mathcal{G}_1$. □

1.3 Cylinder object and homotopy

Let $\mathcal{A}_\perp$, $\mathcal{A}_I$ and $\mathcal{A}_\top$ be three copies of an A_∞-category $\mathcal{A}$. We denote by $\mathcal{C}$ the Grothendieck construction of the diagram

$$\begin{tikzcd} \mathcal{A}_I \arrow[r, "\mathrm{id}"] \arrow[d, "\mathrm{id}"'] & \mathcal{A}_\top \\ \mathcal{A}_\perp & \end{tikzcd}$$

and we let $\iota_\perp, \iota_I, \iota_\top : \mathcal{A} \to \mathcal{C}$ be the strict inclusions with images $\mathcal{A}_\perp$, $\mathcal{A}_I$, $\mathcal{A}_\top$ respectively. Finally, we denote by $W_\mathcal{C}$ the set of adjacent units in $\mathcal{C}$, and we let $\mathcal{C}yl_\mathcal{A} = \mathcal{C}[W_\mathcal{C}^{-1}]$ be the homotopy colimit of the diagram above. We say that $\mathcal{C}yl_\mathcal{A}$ is a cylinder object for $\mathcal{A}$.

We denote by $\pi : \mathcal{C} \to \mathcal{A}$ the A_∞-functor induced by the commutative square

$$\begin{tikzcd} \mathcal{A} \arrow[r, "\mathrm{id}"] \arrow[d, "\mathrm{id}"'] & \mathcal{A} \arrow[d, "\mathrm{id}"] \\ \mathcal{A} \arrow[r, "\mathrm{id}"] & \mathcal{A} \end{tikzcd}$$

(see Proposition 1.21).

Proposition 1.24 *The following diagram of A_∞-categories commutes*:

$$\begin{tikzcd} \mathcal{A} \sqcup \mathcal{A} \arrow[r, "\iota_\perp \sqcup \iota_\top"] \arrow[rr, bend right, "\mathrm{id} \sqcup \mathrm{id}"'] & \mathcal{C} \arrow[r, "\pi"] & \mathcal{A} \end{tikzcd}$$

Moreover, π sends $W_\mathcal{C}$ to the set of units in $\mathcal{A}$, and the induced A_∞-functor $\tilde{\pi} : \mathcal{C}yl_\mathcal{A} \to \mathcal{A}[\{\text{units}\}^{-1}]$ is a quasi-equivalence.

Proof The facts that $\pi \circ (\iota_\perp \sqcup \iota_\top) = \mathrm{id} \sqcup \mathrm{id}$ and that π sends $W_\mathcal{C}$ to the set of units in $\mathcal{A}$ are clear. We now show that $\tilde{\pi} : \mathcal{C}yl_\mathcal{A} \to \mathcal{A}[\{\text{units}\}^{-1}]$ is a quasi-equivalence.

First observe that it is enough to show that the map

$$\tilde{\pi} : \mathcal{C}yl_\mathcal{A}(X, Y) \to \mathcal{A}[\{\text{units}\}^{-1}](\pi X, \pi Y)$$

is a quasi-isomorphism for every objects X, Y in $\mathcal{A}_\perp$ because every object of $\mathcal{C}$ can be related to one of $\mathcal{A}_\perp$ by a zigzag of morphisms in $W_\mathcal{C}$, which are quasi-isomorphisms in $\mathcal{C}yl_\mathcal{A}$ (see [23, Lemma 3.12]). Our strategy is to apply Proposition 1.17. Let $Y = \iota_\perp(Z)$ be an object in $\mathcal{A}_\perp$. For the $\mathcal{C}$-module we take

$$\mathcal{M}_\mathcal{C} = \left[\begin{tikzcd} & \mathcal{C}(-, \iota_I(Z)) \arrow[dl, "t_{I\perp}"] \arrow[dr, "t_{I\top}"] & \\ \mathcal{C}(-, \iota_\perp(Z)) & & \mathcal{C}(-, \iota_\top(Z)) \end{tikzcd} \right],$$

where

$$t_{I\triangle} : \mathcal{C}(-, \iota_I(Z)) \to \mathcal{C}(-, \iota_\triangle(Z)), \quad \triangle \in \{\perp, \top\},$$

is the $\mathcal{C}$-module map induced by the adjacent unit in $\mathcal{C}(\iota_I(Z), \iota_\triangle(Z))$ (see Definition 1.8)). For the $\mathcal{A}$-module we simply take $\mathcal{A}(-, Z)$. Besides, we let $t_\mathcal{C} : \mathcal{C}(-, \iota_\perp(Z)) \to \mathcal{M}_\mathcal{C}$ be the $\mathcal{C}$-module inclusion, and we let $t_{\mathcal{A}}\colon \mathcal{A}(-,Z)\to\mathcal{A}(-,Z)$ be the identity map. We now define the morphism $t_0\colon \mathcal{M}_{\mathcal{C}}\to\pi^*\mathcal{A}(-,Z)$. Consider the diagram of $\mathcal{C}$-modules

$$\begin{tikzcd}
\mathcal{C}(-,\iota_I(Z)) \arrow[r, "t_{I\perp}"] \arrow[d, "t_{I\top}"] & \mathcal{C}(-,\iota_\top(Z)) \arrow[d, "t_\pi"] \\
\mathcal{C}(-,\iota_\perp(Z)) \arrow[r, "t_\pi"] & \pi^*\mathcal{A}(-,Z)
\end{tikzcd}$$

Observe that this diagram is commutative, and thus it induces a strict $\mathcal{C}$-module map $t_0\colon \mathcal{M}_{\mathcal{C}}\to\pi^*\mathcal{A}(-,Z)$ according to Proposition 1.19. It is then easy to see that the following diagram commutes:

$$\begin{tikzcd}
\mathcal{C}(-,\iota_\triangle(Z)) \arrow[r, "t_\pi"] \arrow[d, "t_{\mathcal{C}}"] & \pi^*\mathcal{A}(-,Z) \arrow[d, "\pi^*t_{\mathcal{A}}"] \\
\mathcal{M}_{\mathcal{C}} \arrow[r, "t_0"] & \pi^*\mathcal{A}(-,Z)
\end{tikzcd}$$

To conclude the proof, it suffices to check the three items of Proposition 1.17. Observe that the pair $(\mathcal{A}(-,Z),t_{\mathcal{A}})$ trivially satisfies the two first items.

We check that $\mathcal{M}_{\mathcal{C}}$ satisfies the first item of Proposition 1.17. Let Z' be an object in $\mathcal{A}$ and let w be the adjacent unit in $\mathcal{C}(\iota_I(Z'),\iota_\perp(Z'))$ (the proof is the same for the adjacent unit in $\mathcal{C}(\iota_I(Z'),\iota_\top(Z'))\cap W_{\mathcal{C}}$). Then

$$\mathcal{M}_{\mathcal{C}}(\mathrm{Cone}\,w)=\mathrm{Cone}\big(\mathcal{M}_{\mathcal{C}}(\iota_\perp(Z'))\xrightarrow{\mu^2_{\mathcal{C}}(w,-)}\mathcal{M}_{\mathcal{C}}(\iota_I(Z'))\big)=\mathrm{Cone}\big(\mathcal{C}(\iota_\perp(Z'),\iota_\perp(Z))\xrightarrow{\mu^2_{\mathcal{C}}(w,-)}K\big),$$

where

$$K=\left[\begin{tikzcd}[column sep=small]
 & \mathcal{C}(\iota_I(Z'),\iota_I(Z)) \arrow[dl, "t_{I\perp}"'] \arrow[dr, "t_{I\top}"] & \\
\mathcal{C}(\iota_I(Z'),\iota_\perp(Z)) & & \mathcal{C}(\iota_I(Z'),\iota_\top(Z))
\end{tikzcd}\right].$$

Observe that $\mu^2_{\mathcal{C}}(w,-)\colon \mathcal{C}(\iota_\perp(Z'),\iota_\perp(Z))\to K$ is injective so its cone is quasi-isomorphic to its cokernel, which is the cone of $t_{I\top}\colon \mathcal{C}(\iota_I(Z'),\iota_I(Z))\to\mathcal{C}(\iota_I(Z'),\iota_\top(Z))$. The latter map is a quasi-isomorphism, so $\mathcal{M}_{\mathcal{C}}(\mathrm{Cone}\,w)$ is acyclic.

We now check that $(\mathcal{M}_{\mathcal{C}},t_{\mathcal{C}})$ satisfies the second item of Proposition 1.17. Observe that

$$W_{\mathcal{C}}^{-1}\mathcal{M}_{\mathcal{C}}=\left[\begin{tikzcd}[column sep=small]
 & W_{\mathcal{C}}^{-1}\mathcal{C}(-,\iota_I(Z)) \arrow[dl, "W_{\mathcal{C}}^{-1}t_{I\perp}"'] \arrow[dr, "W_{\mathcal{C}}^{-1}t_{I\top}"] & \\
W_{\mathcal{C}}^{-1}\mathcal{C}(-,\iota_\perp(Z)) & & W_{\mathcal{C}}^{-1}\mathcal{C}(-,\iota_\top(Z))
\end{tikzcd}\right]$$

and $W_{\mathcal{C}}^{-1}t_{\mathcal{C}}\colon W_{\mathcal{C}}^{-1}\mathcal{C}(-,\iota_\perp(Z))\to W_{\mathcal{C}}^{-1}\mathcal{M}_{\mathcal{C}}$ is the inclusion. Thus if X is some object of $\mathcal{C}$, the cone of $W_{\mathcal{C}}^{-1}t_{\mathcal{C}}\colon \mathcal{C}yl_{\mathcal{A}}(X,\iota_\perp(Z))\to W_{\mathcal{C}}^{-1}\mathcal{M}_{\mathcal{C}}(X)$ is quasi-isomorphic to the cone of the multiplication in $\mathcal{C}yl_{\mathcal{A}}$ by an element of $W_{\mathcal{C}}$, which is a quasi-isomorphism. Thus the map $W_{\mathcal{C}}^{-1}t_{\mathcal{C}}\colon W_{\mathcal{C}}^{-1}\mathcal{C}(-,\iota_\perp(Z))\to W_{\mathcal{C}}^{-1}\mathcal{M}_{\mathcal{C}}$ indeed is a quasi-isomorphism.

It remains to check the third item of Proposition 1.17, which is that the map $t_0\colon \mathcal{M}_{\mathcal{C}}(X) \to \pi^*\mathcal{A}(-,Z)(X)$ is a quasi-isomorphism when X is in $\mathcal{A}_\perp$. This is true because $\mathcal{M}_{\mathcal{C}}(\iota_\perp(Z')) = \mathcal{C}(\iota_\perp(Z'),\iota_\perp(Z)) = \mathcal{A}(Z',Z)$, and

$$t_0\colon \mathcal{A}(Z',Z) = \mathcal{M}_{\mathcal{C}}(\iota_\perp(Z')) \to \pi^*\mathcal{A}(-,Z)(\iota_\perp(Z')) = \mathcal{A}(Z',Z)$$

is the identity. □

Remark Proposition 1.24 can be thought as saying that $\mathcal{C}yl_{\mathcal{A}}$ is a cylinder object for $\mathcal{A}$.

Proposition 1.25 *If two A_∞-functors $\Phi,\Psi\colon \mathcal{A}\to\mathcal{B}$ are homotopic* (*see* [36, Paragraph (1h)]), *then there is an A_∞-functor $\eta\colon \mathcal{C}\to\mathcal{B}$ which sends the adjacent units of $\mathcal{C}$ to the units in $\mathcal{B}$ and such that the following diagram commutes*:

$$\begin{tikzcd} \mathcal{A} \arrow[d, "\iota_\top"'] \arrow[dr, bend left, "\Phi"] & \\ \mathcal{C} \arrow[r, "\eta"] & \mathcal{B} \\ \mathcal{A} \arrow[u, "\iota_\perp"] \arrow[ur, bend right, "\Psi"'] & \end{tikzcd}$$

Proof On the objects, we set $\eta(X_\triangle) = \Phi(X) = \Psi(X)$ for every object X of $\mathcal{A}$ and $\triangle\in\{\perp, I, \top\}$. On the morphisms, we set

$$\eta|_{\mathcal{A}_\perp} = \eta|_{\mathcal{A}_I} = \Psi, \quad \eta|_{\mathcal{A}_\top} = \Phi$$

and ask for the restriction of η to

$$\mathcal{A}_I(X_0,X_1)\otimes\cdots\otimes\mathcal{A}_I(X_{p-1},X_p)\otimes\mathcal{C}(X_p,X_{p+1})\otimes\mathcal{A}_\perp(X_{p+1},X_{p+2})\otimes\cdots\otimes\mathcal{A}_\perp(X_{p+q},X_{p+q+1})$$

to be Ψ. It remains to define η for

$$\begin{aligned}&(\dots,x,\dots)\\ &\in \mathcal{A}_I(X_0,X_1)\otimes\cdots\otimes\mathcal{A}_I(X_{p-1},X_p)\otimes\mathcal{C}(X_p,X_{p+1})\otimes\mathcal{A}_\top(X_{p+1},X_{p+2})\otimes\cdots\otimes\mathcal{A}_\top(X_{p+q},X_{p+q+1}).\end{aligned}$$

For this we take a homotopy T between Φ and Ψ, which means that

$$\Phi+\Psi = \sum T(\dots,\mu_{\mathcal{A}}(\dots),\dots) + \sum \mu_{\mathcal{B}}\big(\Phi(\dots),\dots,\Phi(\dots),T(\dots),\Psi(\dots),\dots,\Psi(\dots)\big).$$

Then we let

$$\begin{aligned}&\eta(\dots,x,\dots)\\ &\quad= \sum \mu_{\mathcal{B}}\big(\Phi(\dots),\dots,\Phi(\dots),T(\dots),\Psi(\dots),\dots,\Psi(\dots),\Psi(\dots,x,\dots)\Psi(\dots),\dots,\Psi(\dots)\big)\end{aligned}$$

if p is positive, and $\eta(x,\dots) = \Psi(x,\dots)$ otherwise. □

1.4 Adjunctions between Adams-graded and non-Adams-graded

We end this section by describing specific adjunctions between the category of Adams-graded A_∞-categories concentrated in nonnegative Adams degree and the category of (non-Adams-graded) A_∞-categories.

Definition 1.26 If V is an Adams-graded vector space and m is an integer, we denote by V_m the graded vector space whose degree n component is the direct sum of the bidegree (p,k) components of V, where the sum is over the set of couples $(p,k) \in \mathbb{Z} \times \mathbb{Z}$ such that $p - mk = n$.

Definition 1.27 If $\mathcal{C}$ is an Adams-graded A_∞-category, we denote by $\mathcal{C}_m$ the (non-Adams-graded) A_∞-category obtained from $\mathcal{C}$ by changing the grading so that

$$\mathcal{C}_m(X_0, X_1) = \mathcal{C}(X_0, X_1)_m$$

Observe that any A_∞-functor $\Phi\colon \mathcal{C}_1 \to \mathcal{C}_2$ between two Adams-graded A_∞-categories induces an A_∞-functor from $(\mathcal{C}_1)_m$ to $(\mathcal{C}_2)_m$ (that we still denote by Φ) which acts exactly as Φ on objects and morphisms. This defines a functor $\mathcal{C} \mapsto \mathcal{C}_m$ from the category of Adams-graded A_∞-categories to the category of (non-Adams-graded) A_∞-categories.

We denote by $\mathbb{F}[t_m]$ the augmented Adams-graded associative algebra generated by a variable t_m of bidegree $(m,1)$, and by $t_m\mathbb{F}[t_m]$ its augmentation ideal (or equivalently, the ideal generated by t_m).

Definition 1.28 If $\mathcal{D}$ is a (non-Adams-graded) A_∞-category, we denote by $\mathbb{F}[t_m] \otimes \mathcal{D}$ the Adams-graded A_∞-category such that

1. the objects of $\mathbb{F}[t_m] \otimes \mathcal{D}$ are those of $\mathcal{D}$,
2. the space of morphisms from Y_1 to Y_2 is $\mathbb{F}[t_m] \otimes \mathcal{D}(Y_1, Y_2)$, and if $y \in \mathcal{D}(Y_1, Y_2)$ is of degree j, $t_m^k \otimes y$ is of bidegree $(j + mk, k)$,
3. the operations send any sequence $(t_m^{k_0} \otimes y_0, \dots, t_m^{k_{d-1}} \otimes y_{d-1})$ of morphisms to

$$\mu_{\mathbb{F}[t_m]\otimes\mathcal{D}}(t_m^{k_0} \otimes y_0, \dots, t_m^{k_{d-1}} \otimes y_{d-1}) = t_m^{k_0+\cdots+k_{d-1}} \otimes \mu_{\mathcal{D}}(y_0, \dots, y_{d-1}).$$

Observe that any A_∞-functor $\Psi\colon \mathcal{D}_1 \to \mathcal{D}_2$ between (non-Adams-graded) A_∞-categories induces an A_∞-functor $\mathbb{F}[t_m] \otimes \mathcal{D}_1 \to \mathbb{F}[t_m] \otimes \mathcal{D}_2$ which acts as Ψ on objects, and which sends any sequence $(t_m^{k_0} \otimes y_0, \dots, t_m^{k_{d-1}} \otimes y_{d-1})$ of morphisms to $t_m^{k_0+\cdots+k_{d-1}} \otimes \Psi(y_0, \dots, y_{d-1})$. This defines a functor $\mathcal{D} \mapsto \mathbb{F}[t_m] \otimes \mathcal{D}$ from the category of (non-Adams-graded) A_∞-categories to the category of Adams-graded A_∞-categories.

Definition 1.29 Let $\mathcal{C}$ be an Adams-graded A_∞-category concentrated in nonnegative Adams degree, and let $\mathcal{D}$ be a (non-Adams-graded) A_∞-category. To any A_∞-functor $\Psi_m\colon \mathcal{C}_m \to \mathcal{D}$, we associate an A_∞-functor $\Psi\colon \mathcal{C} \to \mathbb{F}[t_m] \otimes \mathcal{D}$ which sends a sequence $(x_0, \dots, x_{d-1})$, where x_j is of bidegree (i_j, k_j), to

$$\Psi(x_0, \dots, x_{d-1}) = t_m^{k_0+\cdots+k_{d-1}} \otimes \Psi_m(x_0, \dots, x_{d-1}).$$

This defines an adjunction between the category of Adams-graded A_∞-categories concentrated in nonnegative Adams degree and the category of (non-Adams-graded) A_∞-categories.

2 Mapping torus of an A_∞-autoequivalence

In this section, we introduce the notion of mapping torus for a quasi-autoequivalence of an A_∞-category, by analogy with the mapping torus associated to an automorphism of a topological space. This terminology was also used in [26], but we do not know if the two notions coincide. The two main theorems of this section allow us to compute this mapping torus under different hypotheses.

Remark In this section, A_∞-categories are always assumed to be *strictly unital* (see [36, Paragraph (2a)]).

2.1 Definitions and main results

2.1.1 Definitions

Definition 2.1 Let τ be a quasi-autoequivalence of an Adams-graded A_∞-category $\mathcal{A}$. The mapping torus of τ is the A_∞-category

$$\mathrm{MT}(\tau) := \operatorname{hocolim}\left(\begin{tikzcd} \mathcal{A} \sqcup \mathcal{A} \arrow[r, "\mathrm{id}\sqcup\tau"] \arrow[d, "\mathrm{id}\sqcup\mathrm{id}"'] & \mathcal{A} \\ \mathcal{A} & \end{tikzcd}\right)$$

(see Definition 1.20).

Remark (1) We use the terminology "mapping torus" by analogy with the analogous situation in the category of topological spaces. Indeed, if f is an automorphism of some topological space X, then the mapping torus of f

$$M_f = (X \times [0,1])/((x,0) \sim (f(x),1))$$

is the homotopy colimit of the diagram

$$\begin{tikzcd} X \sqcup X \arrow[r, "\mathrm{id}\sqcup f"] \arrow[d, "\mathrm{id}\sqcup\mathrm{id}"'] & X \\ X & \end{tikzcd}$$

(2) The terminology "mapping torus of an autoequivalence of A_∞-categories" also appears in [26], where the corresponding DG-category is denoted by M_τ, and it is used in [25] to distinguish open symplectic mapping tori. According to [25, Appendix A], M_τ is equivalent to the homotopy colimit of

$$\mathcal{A} \underset{i_0\otimes\mathrm{id}}{\overset{i_\infty\otimes\tau}{\rightrightarrows}} \mathcal{O}(\mathbb{P}^1)\otimes\mathcal{A}$$

whereas $\mathrm{MT}(\tau)$ should rather be equivalent to the homotopy colimit of

$$\mathcal{A} \underset{\mathrm{id}}{\overset{\tau}{\rightrightarrows}} \mathcal{A}$$

(we did not define $\mathrm{MT}(\tau)$ using the latter diagram because [24] only defines homotopy colimits of diagrams indexed by posets).

(3) The mapping torus of a quasi-autoequivalence is also Adams-graded, because it is the localization of an Adams-graded A_∞-category at morphisms of Adams degree 0.

Definition 2.2 Let $\mathcal{A}$ be an A_∞-category. A $\mathbb{Z}$-splitting of $\mathrm{ob}(\mathcal{A})$ is a bijection

$$\mathbb{Z} \times \mathcal{E} \xrightarrow{\sim} \mathrm{ob}(\mathcal{A}), \quad (n, E) \mapsto X^n(E),$$

where $\mathcal{E}$ is some set. If such a splitting has been chosen, we define the Adams-grading of a homogeneous element $x \in \mathcal{A}(X^i(E), X^j(E))$ to be $j - i$. This turns $\mathcal{A}$ into an Adams-graded A_∞-category.

Let τ be a quasi-autoequivalence of $\mathcal{A}$. We say that a $\mathbb{Z}$-splitting of $\mathrm{ob}(\mathcal{A})$ is compatible with τ if

$$\tau(X^n(E)) = X^{n+1}(E)$$

for every $n \in \mathbb{Z}$ and $E \in \mathcal{E}$.

We say that $\mathcal{A}$ is weakly directed with respect to a $\mathbb{Z}$-splitting of $\mathrm{ob}(\mathcal{A})$ if

$$\mathcal{A}(X^i(E), X^j(E')) = 0$$

for every $i > j$ and $E, E' \in \mathcal{E}$ (we use the term "weakly directed" A_∞-category because the notion is slightly more general than that of directed A_∞-category defined by Seidel in [36, Paragraph (5m)]).

Remark Compatible $\mathbb{Z}$-splittings naturally arise in the context of $\mathbb{Z}$-actions. A strict $\mathbb{Z}$-action on an A_∞-category $\mathcal{A}$ is a family of A_∞-endofunctors $(\tau_n)_{n\in\mathbb{Z}}$ such that $\tau_0 = \mathrm{id}_{\mathcal{A}}$ and $\tau_{i+j} = \tau_i \circ \tau_j$ (see [36, Paragraph (10b)]). If the induced $\mathbb{Z}$-action on $\mathrm{ob}(\mathcal{A})$ is free, then any section σ of the projection $\mathrm{ob}(\mathcal{A}) \to \mathcal{E}$, where $\mathcal{E}$ is the set of equivalence classes of objects in $\mathcal{A}$ under the $\mathbb{Z}$-action, gives a $\mathbb{Z}$-splitting

$$\mathbb{Z} \times \mathcal{E} \xrightarrow{\sim} \mathrm{ob}(\mathcal{A}), \quad (n, E) \mapsto \tau_n(\sigma(E)),$$

which is compatible with the automorphism τ_1.

2.1.2 Main results

First result

Definition 2.3 Let τ be a quasi-autoequivalence of an A_∞-category $\mathcal{A}$ equipped with a compatible $\mathbb{Z}$-splitting of $\mathrm{ob}(\mathcal{A})$. Assume that τ is strict, ie $\tau^d = 0$ for $d \geq 2$, and acts bijectively on hom-sets. In this case, we define an Adams-graded A_∞-category $\mathcal{A}_\tau$ as:

(1) The set of objects of $\mathcal{A}_\tau$ is $\mathcal{E}$.

(2) The space of morphisms $\mathcal{A}_\tau(E, E')$ is the Adams-graded vector space given by

$$\mathcal{A}_\tau(E, E') = \Bigl(\bigoplus_{i,j\in\mathbb{Z}} \mathcal{A}(X^i(E), X^j(E'))\Bigr)\Big/(\tau(x) \sim x).$$

(3) The operations are the unique linear maps such that for every sequence

$$(x_0, \ldots, x_{d-1}) \in \mathcal{A}(X^{i_0}(E_0), X^{i_1}(E_1)) \times \cdots \times \mathcal{A}(X^{i_{d-1}}(E_{d-1}), X^{i_d}(E_d)),$$

we have

$$\mu_{\mathcal{A}_\tau}([x_0], \ldots, [x_{d-1}]) = [\mu_{\mathcal{A}}(x_0, \ldots, x_{d-1})],$$

where $[-]\colon \mathcal{A}(X^i(E), X^j(E')) \to \mathcal{A}_\tau(E, E')$ denotes the projection. (It is not hard to see that such operations exist and satisfy the A_∞-relations.)

Remark When $\mathcal{A}$ is a DG-category, the latter construction is known as the orbit category, see [27; 28, Section 4.9]. In [26, Section 4], it is denoted by $\mathcal{A}\#\mathbb{Z}$ (considering that τ induces a $\mathbb{Z}$-action on $\mathcal{A}$).

Theorem 2.4 *Let τ be a quasi-autoequivalence of an A_∞-category $\mathcal{A}$ equipped with a compatible $\mathbb{Z}$-splitting of* $\mathrm{ob}(\mathcal{A})$. *Assume that τ is strict and acts bijectively on hom-sets. Then there is a quasi-equivalence of Adams-graded A_∞-categories*

$$\mathrm{MT}(\tau) \simeq \mathcal{A}_\tau.$$

Remark (1) The A_∞-category $\mathcal{A}_\tau$ is the (ordinary) colimit of the diagram used to define $\mathrm{MT}(\tau)$. Thus, Theorem 2.4 can be thought of as a "homotopy colimit equals colimit" result.

(2) In [26], given a DG-category $\mathcal{A}$ and an autoequivalence τ acting bijectively on hom-sets, the author defines a DG-category $M_\tau := (\mathcal{O}(\widetilde{\mathcal{T}}_0)_{\mathrm{dg}} \otimes \mathcal{A})\#\mathbb{Z}$ (see [26] for the notation). In the case where τ moreover induces a free $\mathbb{Z}$-action on objects, Theorem 2.4 says that relating $\mathrm{MT}(\tau)$ and M_τ amounts to comparing $\mathcal{A}\#\mathbb{Z}$ and $(\mathcal{O}(\widetilde{\mathcal{T}}_0)_{\mathrm{dg}} \otimes \mathcal{A})\#\mathbb{Z}$.

Second result We denote by $\mathbb{F}[t_m]$ the augmented Adams-graded associative algebra generated by a variable t_m of bidegree $(m, 1)$. Observe that if $\mathcal{C}$ is a subcategory of an A_∞-category $\mathcal{D}$ with $\mathrm{ob}(\mathcal{C}) = \mathrm{ob}(\mathcal{D})$, then $\mathcal{C} \oplus (t\mathbb{F}[t] \otimes \mathcal{D})$ is naturally an Adams-graded A_∞-category, where the Adams degree of $t^k \otimes x$ equals k. Besides, if $\mathcal{C}$ is an A_∞-category equipped with a $\mathbb{Z}$-splitting of $\mathrm{ob}(\mathcal{C})$, we denote by $\mathcal{C}^0$ the full A_∞-subcategory of $\mathcal{C}$ whose set of objects corresponds to $\{0\} \times \mathcal{E}$. Finally, we use the functor $\mathcal{C} \mapsto \mathcal{C}_m$ of Definition 1.27.

Theorem 2.5 (Theorem A in the introduction) *Let τ be a quasi-autoequivalence of an A_∞-category $\mathcal{A}$, weakly directed with respect to some compatible $\mathbb{Z}$-splitting of* $\mathrm{ob}(\mathcal{A})$. *Assume that there exists a closed degree* 0 *bimodule map $f : \mathcal{A}_m(-,-) \to \mathcal{A}_m(-,\tau(-))$ such that*

$$f : \mathcal{A}_m(X^i(E), X^j(E')) \to \mathcal{A}_m(X^i(E), X^{j+1}(E'))$$

is a quasi-isomorphism for every $i < j$ and $E, E' \in \mathcal{E}$. Then there is a quasi-equivalence of Adams-graded A_∞-categories

$$\mathrm{MT}(\tau) \simeq \mathcal{A}_m^0 \oplus (t_m\mathbb{F}[t_m] \otimes \mathcal{A}_m[f(\mathrm{units})^{-1}]^0).$$

Outline of the section In Section 2.2, we consider an A_∞-category $\mathcal{A}$ equipped with a $\mathbb{Z}$-splitting of $\mathrm{ob}(\mathcal{A})$ and a choice of a closed degree 0 morphism $c_n(E) \in \mathcal{A}(X^n(E), X^{n+1}(E))$ for every $n \in \mathbb{Z}$ and every $E \in \mathcal{E}$. We give technical results about specific modules associated to this data. This will be used in the proof of Theorem 2.5 with $c_n(E) = f(e_{X^n(E)})$.

In Section 2.3, we consider the Grothendieck construction $\mathcal{G}$ of a slightly different diagram than the one in Definition 2.1, together with a set $W_{\mathcal{G}}$ of closed degree 0 morphisms. The idea is that the localization

$\mathcal{H} = \mathcal{G}[W_{\mathcal{G}}^{-1}]$ is the homotopy colimit of a diagram obtained from the one in Definition 2.1 by a cofibrant replacement of the diagonal functor $\mathcal{A} \sqcup \mathcal{A} \to \mathcal{A}$. Thus it is not surprising that $\mathcal{H}$ is quasi-equivalent to the mapping torus of τ. Moreover, we prove technical results about specific modules over $\mathcal{G}$ that will be used in the proofs of Theorems 2.4 and 2.5.

In Section 2.4, we prove Theorem 2.4. We first define an A_∞-functor $\Phi\colon \mathcal{G} \to \mathcal{A}_\tau$ which sends $W_{\mathcal{G}}$ to the set of units in $\mathcal{A}_\tau$. Then we prove that the induced A_∞-functor $\widetilde{\Phi}\colon \mathcal{H} \to \mathcal{A}_\tau[\{\text{units}\}^{-1}]$ is a quasi-equivalence. To do that, our strategy is to apply Proposition 1.17 using the results of Section 2.3 about the specific $\mathcal{G}$-modules.

In Section 2.5, we prove Theorem 2.5. We use the fact that $\mathcal{G}$ is "big enough" (there are more objects and morphisms than in the Grothendieck construction of the diagram in Definition 2.1) in order to define an A_∞-functor $\Psi_m\colon \mathcal{G}_m \to \mathcal{A}_m$ (see Definition 1.27). This induces an A_∞-functor

$$\widetilde{\Psi}\colon \mathcal{H} \to \mathbb{F}[t_m] \otimes \mathcal{A}_m[f(\{\text{units}\})^{-1}].$$

Then we prove that for every Adams degree $j \geq 1$, and for every objects X, Y in $\mathcal{H}$, the map

$$\widetilde{\Psi}\colon \mathcal{H}(X,Y)^{*,j} \to (\mathbb{F}[t_m] \otimes \mathcal{A}_m[f(\{\text{units}\})^{-1}])(\Psi X, \Psi Y)^{*,j}$$

is a quasi-isomorphism (if V is an Adams-graded vector space, $V^{*,j}$ denotes the subspace of Adams degree j elements). To do that, we apply once again Proposition 1.17 using the results of Sections 2.2 and 2.3 about the specific modules over $\mathcal{A}_m$ and $\mathcal{G}$ respectively. This allows us to finish the proof of Theorem 2.5.

2.2 Results about specific modules

In this section, we give technical results that will allow us to apply Proposition 1.17 in the proof of Theorem 2.5.

Let $\mathcal{A}$ be an A_∞-category equipped with a $\mathbb{Z}$-splitting of $\operatorname{ob}(\mathcal{A})$. Assume that we have chosen, for every $n \in \mathbb{Z}$ and every $E \in \mathcal{E}$, a closed degree 0 morphism $c_n(E) \in \mathcal{A}(X^n(E), X^{n+1}(E))$. Moreover, assume that we have chosen a set $W_{\mathcal{A}}$ of closed degree 0 morphisms which contains the morphisms $c_n(E)$.

Remark According to Definition 2.2, the $\mathbb{Z}$-splitting of $\operatorname{ob}(\mathcal{A})$ naturally induces an Adams-grading on $\mathcal{A}$. However in this section, we do not consider $\mathcal{A}$ as being Adams-graded.

In the following, we fix some element $E_\diamond \in \mathcal{E}$. When we write an object X^n or a morphism c_n without specifying the element of $\mathcal{E}$, we mean $X^n(E_\diamond)$ or $c_n(E_\diamond)$ respectively. Recall that

$$t_{c_n}\colon \mathcal{A}(-, X^n) \to \mathcal{A}(-, X^{n+1})$$

denotes the $\mathcal{A}$-module map induced by $c_n \in \mathcal{A}(X^n, X^{n+1})$ (see Definition 1.8).

Definition 2.6 We set $\mathcal{M}_{\mathcal{A}}$ to be the $\mathcal{A}$-module

$$\mathcal{M}_{\mathcal{A}} := \left[\begin{tikzcd} \cdots \arrow[dr, "t_{c_{-1}}"] & \mathcal{A}(-,X^0) \arrow[d, "\mathrm{id}"] \arrow[dr, "t_{c_0}"] & \mathcal{A}(-,X^1) \arrow[d, "\mathrm{id}"] \arrow[dr, "t_{c_1}"] & \cdots \\ \cdots & \mathcal{A}(-,X^0) & \mathcal{A}(-,X^1) & \cdots \end{tikzcd}\right] = \left[\begin{tikzcd} \bigoplus_{i\in\mathbb{Z}} \mathcal{A}(-,X^i) \arrow[d, "\bigoplus_{i\in\mathbb{Z}}(\mathrm{id},t_{c_i})"] \\ \bigoplus_{i\in\mathbb{Z}} \mathcal{A}(-,X^i) \end{tikzcd}\right]$$

(see Definition 1.18). Besides, we set $t^n_{\mathcal{A}}\colon \mathcal{A}(-,X^n)\to\mathcal{M}_{\mathcal{A}}$ to be the $\mathcal{A}$-module inclusion for every $n\in\mathbb{Z}$.

The first result highlights a key property of the module $\mathcal{M}_{\mathcal{A}}$.

Lemma 2.7 *For every $n\in\mathbb{Z}$, the closed $\mathcal{A}$-module map $t^{n+1}_{\mathcal{A}}\circ t_{c_n}\colon \mathcal{A}(-,X^n)\to\mathcal{M}_{\mathcal{A}}$ is homotopic to $t^n_{\mathcal{A}}\colon \mathcal{A}(-,X^n)\to\mathcal{M}_{\mathcal{A}}$.*

Proof Consider the degree -1 strict $\mathcal{A}$-module map $s\colon \mathcal{A}(-,X^n)\to\mathcal{M}_{\mathcal{A}}$ which sends a morphism in $\mathcal{A}(X,X^n)$ to the corresponding shifted element in $\mathcal{A}(X,X^n)[1]$. Then an easy computation gives

$$\mu^1_{\mathrm{Mod}_{\mathcal{A}}}(s) = t^{n+1}_{\mathcal{A}}\circ t_{c_n} + t^n_{\mathcal{A}}. \qquad \square$$

In the proof of the two results below, we will use specific $\mathcal{A}$-modules. If p is a fixed nonnegative integer, we set

$$K_p = \left[\begin{tikzcd} \cdots \arrow[dr, "t_{c_{p-2}}"] & \mathcal{A}(-,X^{p-1}) \arrow[d, "\mathrm{id}"] \arrow[dr, "t_{c_{p-1}}"] & \mathcal{A}(-,X^p) \arrow[d, "\mathrm{id}"] \\ \cdots & \mathcal{A}(-,X^{p-1}) & \mathcal{A}(-,X^p) \end{tikzcd}\right]$$

and

$$\widetilde{K}_p = \left[\begin{tikzcd} \mathcal{A}(-,X^p) \arrow[dr, "t_{c_p}"] & \mathcal{A}(-,X^{p+1}) \arrow[d, "\mathrm{id}"] \arrow[dr, "t_{c_{p+1}}"] & \cdots \\ & \mathcal{A}(-,X^{p+1}) & \cdots \end{tikzcd}\right].$$

Moreover, we will consider the sequences of $\mathcal{A}$-modules $(F^q_p)_{q\geq 0}$, $(\widetilde{F}^q_p)_{q\geq 0}$ starting at $F^0_p=\widetilde{F}^0_p=0$ and with

$$F^q_p = \left[\begin{tikzcd} \mathcal{A}(-,X^{p-q+1}) \arrow[d, "\mathrm{id}"] \arrow[dr, "t_{c_{p-q+1}}"] & \cdots \arrow[dr, "t_{c_{p-1}}"] & \mathcal{A}(-,X^p) \arrow[d, "\mathrm{id}"] \\ \mathcal{A}(-,X^{p-q+1}) & \cdots & \mathcal{A}(-,X^p) \end{tikzcd}\right]$$

and

$$\widetilde{F}^q_p = \left[\begin{tikzcd} \mathcal{A}(-,X^p) \arrow[dr, "t_{c_p}"] & \cdots \arrow[dr, "t_{c_{p+q-1}}"] & \\ & \cdots & \mathcal{A}(-,X^{p+q}) \end{tikzcd}\right]$$

for $q\geq 1$.

The following lemma is mostly technical. It will be used in the proofs of Lemmas 2.9 and 2.21.

Lemma 2.8 *Assume that for every $i < j$ and that for every $E \in \mathcal{E}$ the chain map*

$$\mu^2_{\mathcal{A}}(-, c_j)\colon \mathcal{A}(X^i(E), X^j) \to \mathcal{A}(X^i(E), X^{j+1})$$

is a quasi-isomorphism. Then for every $k < n$ and for every $E \in \mathcal{E}$, the inclusion $\mathcal{A}(X^k(E), X^n) \hookrightarrow \mathcal{M}_{\mathcal{A}}(X^k(E))$ is a quasi-isomorphism.

Proof The cone of the inclusion $\mathcal{A}(X^k(E), X^n) \hookrightarrow \mathcal{M}_{\mathcal{A}}(X^k(E))$ is quasi-isomorphic to its cokernel, which is $K_{n-1}(X^k(E)) \oplus \widetilde{K}_n(X^k(E))$.

We have to show that these complexes are acyclic. Observe that

$$\big(F^q_{n-1}(X^k(E))\big)_{q\geq 0} \quad \text{and} \quad \big(\widetilde{F}^q_n(X^k(E))\big)_{q\geq 0}$$

are increasing, exhaustive, and bounded from below filtrations of $K_{n-1}(X^k(E))$ and $\widetilde{K}_n(X^k(E))$, respectively. For every $q \geq 1$, we have

$$F^q_{n-1}(X^k(E))/F^{q-1}_{n-1}(X^k(E)) = \begin{bmatrix} \mathcal{A}(X^k(E), X^{n-q}) \\ \big\downarrow \mathrm{id} \\ \mathcal{A}(X^k(E), X^{n-q}) \end{bmatrix}$$

and

$$\widetilde{F}^q_n(X^k(E))/\widetilde{F}^{q-1}_n(X^k(E)) = \begin{bmatrix} \mathcal{A}(X^k(E), X^{n+q-1}) \xrightarrow{t_{c_{n+q-1}}} \mathcal{A}(X^k(E), X^{n+q}) \end{bmatrix}.$$

The first of the two latter complexes is clearly acyclic, and the second one is acyclic by assumption on the morphisms c_j. Thus the entire complex $K_{n-1}(X^k(E)) \oplus \widetilde{K}_n(X^k(E))$ is acyclic, which is what we needed to prove. □

The following two lemmas will be used later in order to apply Proposition 1.17.

Lemma 2.9 *Assume that for every $i < j < k$ and for every $E \in \mathcal{E}$ that the chain maps*

$$\begin{aligned} \mu^2_{\mathcal{A}}(-, c_j)\colon &\quad \mathcal{A}(X^i(E), X^j) \to \mathcal{A}(X^i(E), X^{j+1}), \\ \mu^2_{\mathcal{A}}(c_j(E), -)\colon &\quad \mathcal{A}(X^{j+1}(E), X^{k+1}) \to \mathcal{A}(X^j(E), X^{k+1}) \end{aligned}$$

are quasi-isomorphisms. Then for every $(n, E) \in \mathbb{Z} \times \mathcal{E}$, the complex $\mathcal{M}_{\mathcal{A}}(\mathrm{Cone}\, c_n(E))$ is acyclic.

Proof We have

$$\mathcal{M}_{\mathcal{A}}(\mathrm{Cone}\, c_n(E)) = \mathrm{Cone}\big(\mathcal{M}_{\mathcal{A}}(X^{n+1}(E)) \xrightarrow{\mu^2_{\mathcal{M}_{\mathcal{A}}}(c_n(E),-)} \mathcal{M}_{\mathcal{A}}(X^n(E))\big),$$

so we have to prove that $\mu^2_{\mathcal{M}_\mathcal{A}}(c_n(E),-)\colon \mathcal{M}_\mathcal{A}(X^{n+1}(E)) \to \mathcal{M}_\mathcal{A}(X^n(E))$ is a quasi-isomorphism. Observe that we have the commutative diagram

$$\begin{tikzcd}
\mathcal{M}_\mathcal{A}(X^{n+1}(E)) \arrow[r, "\mu^2_{\mathcal{M}_\mathcal{A}}(c_n(E),-)"] & \mathcal{M}_\mathcal{A}(X^n(E)) \\
\mathcal{A}(X^{n+1}(E), X^{n+2}) \arrow[u, hook] \arrow[r, "\mu^2_{\mathcal{A}}(c_n(E),-)"] & \mathcal{A}(X^n(E), X^{n+2}) \arrow[u, hook]
\end{tikzcd}$$

The bottom horizontal map is a quasi-isomorphism by assumption on the morphisms $c_j(E)$. Moreover, the vertical maps are quasi-isomorphisms according to Lemma 2.8. This implies that $\mu^2_{\mathcal{M}_\mathcal{A}}(c_n(E),-)$ is indeed a quasi-isomorphism. $\square$

Lemma 2.10 *The $\mathcal{A}$-module map ${}_{W_\mathcal{A}^{-1}}t^n_\mathcal{A}\colon {}_{W_\mathcal{A}^{-1}}\mathcal{A}(-,X^n) \to {}_{W_\mathcal{A}^{-1}}\mathcal{M}_\mathcal{A}$ is a quasi-isomorphism for every $n \in \mathbb{Z}$.*

Proof Let X be some object of $\mathcal{A}$. We want to prove that the chain map

$${}_{W_\mathcal{A}^{-1}}t^n_\mathcal{A}\colon {}_{W_\mathcal{A}^{-1}}\mathcal{A}(X,X^n) \to {}_{W_\mathcal{A}^{-1}}\mathcal{M}_\mathcal{A}(X)$$

is a quasi-isomorphism. Observe that

$${}_{W_\mathcal{A}^{-1}}\mathcal{M}_\mathcal{A}(X) = \left[\begin{tikzcd}[column sep=small]
\cdots \arrow[dr] & \mathcal{A}[W_\mathcal{A}^{-1}](X,X^0) \arrow[d, "\mathrm{id}"] \arrow[dr, "{}_{W_\mathcal{A}^{-1}}t_{c_0}"] & \mathcal{A}[W_\mathcal{A}^{-1}](X,X^1) \arrow[d, "\mathrm{id}"] \arrow[dr, "{}_{W_\mathcal{A}^{-1}}t_{c_1}"] & \cdots \\
\cdots & \mathcal{A}[W_\mathcal{A}^{-1}](X,X^0) & \mathcal{A}[W_\mathcal{A}^{-1}](X,X^1) & \cdots
\end{tikzcd}\right]$$

and the chain map ${}_{W_\mathcal{A}^{-1}}t^n_\mathcal{A}\colon {}_{W_\mathcal{A}^{-1}}\mathcal{A}(X,X^n) \to {}_{W_\mathcal{A}^{-1}}\mathcal{M}_\mathcal{A}(X)$ is the inclusion. The cone of the latter chain map is then quasi-isomorphic to its cokernel, which is ${}_{W_\mathcal{A}^{-1}}K_{n-1}(X) \oplus {}_{W_\mathcal{A}^{-1}}\widetilde{K}_n(X)$. Observe that $({}_{W_\mathcal{A}^{-1}}F^q_{n-1}(X))_{q\geq 0}$, $({}_{W_\mathcal{A}^{-1}}\widetilde{F}^q_n(X))_{q\geq 0}$ are increasing, exhaustive, and bounded from below filtrations of ${}_{W_\mathcal{A}^{-1}}K_{n-1}(X)$, ${}_{W_\mathcal{A}^{-1}}\widetilde{K}_n(X)$, respectively. For every $q \geq 1$, we have

$${}_{W_\mathcal{A}^{-1}}F^q_{n-1}(X)/{}_{W_\mathcal{A}^{-1}}F^{q-1}_{n-1}(X) = \left[\begin{tikzcd}
\mathcal{A}[W_\mathcal{A}^{-1}](X,X^{n-q}) \arrow[d, "\mathrm{id}"] \\
\mathcal{A}[W_\mathcal{A}^{-1}](X,X^{n-q})
\end{tikzcd}\right]$$

and

$${}_{W_\mathcal{A}^{-1}}\widetilde{F}^q_n(X)/{}_{W_\mathcal{A}^{-1}}\widetilde{F}^{q-1}_n(X) = \left[\begin{tikzcd}
\mathcal{A}[W_\mathcal{A}^{-1}](X,X^{n-1+q}) \arrow[dr, "{}_{W_\mathcal{A}^{-1}}t_{c_{n-1+q}}"] & \\
& \mathcal{A}[W_\mathcal{A}^{-1}](X,X^{n+q})
\end{tikzcd}\right].$$

The first of the two latter complexes is clearly acyclic, and the second one is acyclic because c_{n-1+q} belongs to the set $W_\mathcal{A}$ by which we localized (see [23, Lemma 3.12]). Thus the entire complex ${}_{W_\mathcal{A}^{-1}}K_{n-1}(X) \oplus {}_{W_\mathcal{A}^{-1}}\widetilde{K}_n(X)$ is acyclic, which is what we needed to prove. $\square$

2.3 The A_∞-category and modules for the mapping torus

In this section, we consider an A_∞-category $\mathcal{G}$, together with a set $W_\mathcal{G}$ of closed degree 0 morphisms. We prove that $\mathcal{H} = \mathcal{G}[W_\mathcal{G}^{-1}]$ is quasi-equivalent to the mapping torus of τ, and we prove technical results about specific $\mathcal{G}$-modules that will allow us to apply Proposition 1.17 in the proofs of Theorems 2.4 and 2.5.

Let τ be a quasi-autoequivalence of an A_∞-category $\mathcal{A}$ equipped with a compatible $\mathbb{Z}$-splitting of $\mathrm{ob}(\mathcal{A})$. If $\mathcal{A}_\triangle$ is a copy of $\mathcal{A}$, we denote by $X^n_\triangle(E)$ the object of $\mathcal{A}_\triangle$ corresponding to $(n, E) \in \mathbb{Z} \times \mathcal{E}$.

2.3.1 The Grothendieck construction $\mathcal{G}$ The A_∞-category $\mathcal{G}$ will be the Grothendieck construction of a slightly different diagram than the one in Definition 2.1. The idea is to introduce an A_∞-category $\mathcal{C}$ together with a set of closed degree 0 morphisms $W_\mathcal{C}$ such that the localization $\mathcal{C}[W_\mathcal{C}^{-1}]$ is a cylinder object for $\mathcal{A}$. Observe that this kind of cofibrant replacement is common in homotopy colimits computation, and indeed we need it to prove Theorem 2.5.

Definition 2.11 Let $\mathcal{A}_\perp$, $\mathcal{A}_I$ and $\mathcal{A}_\top$ be three copies of $\mathcal{A}$. We denote by $\mathcal{C}$ the Grothendieck construction (see Definition 1.20) of the diagram

$$\begin{tikzcd} \mathcal{A}_I \arrow[r, "\mathrm{id}"] \arrow[d, "\mathrm{id}"'] & \mathcal{A}_\top \\ \mathcal{A}_\perp & \end{tikzcd}$$

and we let $\iota_\perp, \iota_I, \iota_\top : \mathcal{A} \to \mathcal{C}$ be the strict inclusions with images $\mathcal{A}_\perp$, $\mathcal{A}_I$, $\mathcal{A}_\top$, respectively. Finally, we denote by $W_\mathcal{C}$ the set of adjacent units in $\mathcal{C}$, and we let $\mathcal{C}yl_\mathcal{A} = \mathcal{C}[W_\mathcal{C}^{-1}]$ be the homotopy colimit of the diagram above.

Definition 2.12 Let $\mathcal{A}_-$, $\mathcal{A}_+$, $\mathcal{A}_\bullet$ be three copies of $\mathcal{A}$. We denote by $\mathcal{G}$ the Grothendieck construction of the diagram

$$\begin{tikzcd} \mathcal{A}_- \sqcup \mathcal{A}_+ \arrow[r, "\mathrm{id} \sqcup \tau"] \arrow[d, "\iota_\perp \sqcup \iota_\top"'] & \mathcal{A}_\bullet \\ \mathcal{C} & \end{tikzcd}$$

Also, we denote by $W_\mathcal{G}$ the union of $W_\mathcal{C}$ and the set of adjacent units in $\mathcal{G}$, and we set

$$\mathcal{H} := \mathcal{G}[W_\mathcal{G}^{-1}].$$

According to Proposition 1.24, $\mathcal{C}yl_\mathcal{A}$ can be thought as a cylinder object for $\mathcal{A}$. Therefore, the following result should not be surprising.

Lemma 2.13 *The mapping torus of τ is quasi-equivalent to $\mathcal{H}$.*

Proof Let $\pi : \mathcal{C} \to \mathcal{A}$ be the A_∞-functor induced by the commutative diagram

$$\begin{tikzcd} \mathcal{A}_I \arrow[r, "\mathrm{id}"] \arrow[d, "\mathrm{id}"] & \mathcal{A}_\top \arrow[d, "\mathrm{id}"] \\ \mathcal{A}_\perp \arrow[r, "\mathrm{id}"] & \mathcal{A} \end{tikzcd}$$

(see Proposition 1.21). We get a commutative diagram

$$\begin{tikzcd} \mathcal{C} \arrow[d, "\pi"] & \mathcal{A}_- \sqcup \mathcal{A}_+ \arrow[l, "\iota_\perp \sqcup \iota_\top"'] \arrow[r, "\mathrm{id} \sqcup \tau"] \arrow[d, "\mathrm{id}"] & \mathcal{A}_\bullet \arrow[d, "\mathrm{id}"] \\ \mathcal{A} & \mathcal{A}_- \sqcup \mathcal{A}_+ \arrow[l, "\mathrm{id} \sqcup \mathrm{id}"] \arrow[r, "\mathrm{id} \sqcup \tau"'] & \mathcal{A}_\bullet \end{tikzcd}$$

which induces an A_∞-functor χ from $\mathcal{G}$ to the Grothendieck construction of the bottom line (see Proposition 1.22). Observe that χ sends $W_\mathcal{C}$ to the set U of units in $\mathcal{A}$. Now, according to Proposition 1.24, the A_∞-functor $\tilde{\pi} : \mathcal{C}yl_\mathcal{A} = \mathcal{C}[W_\mathcal{C}^{-1}] \to \mathcal{A}[U^{-1}]$ is a quasi-equivalence. According to Lemma A.6 in [24] (called "localization and homotopy colimits commute"), this implies that the A_∞-functor induced by χ,

$$\operatorname{hocolim} \left(\begin{tikzcd} \mathcal{A}_- \sqcup \mathcal{A}_+ \arrow[r, "\mathrm{id} \sqcup \tau"] \arrow[d, "\iota_\perp \sqcup \iota_\top"] & \mathcal{A}_\bullet \\ \mathcal{C} & \end{tikzcd} \right) [W_\mathcal{C}^{-1}] \xrightarrow{\tilde{\chi}} \operatorname{hocolim} \left(\begin{tikzcd} \mathcal{A}_- \sqcup \mathcal{A}_+ \arrow[r, "\mathrm{id} \sqcup \tau"] \arrow[d, "\mathrm{id} \sqcup \mathrm{id}"] & \mathcal{A}_\bullet \\ \mathcal{A} & \end{tikzcd} \right) [U^{-1}],$$

is a quasi-equivalence. This completes the proof because the source of $\tilde{\chi}$ is exactly $\mathcal{H}$, and its target is quasi-equivalent to the mapping torus of τ. $\square$

2.3.2 Modules over $\mathcal{G}$ In the following, we fix some element $E_\diamond \in \mathcal{E}$. When we write an object $X^n_\triangle$ without specifying the element of $\mathcal{E}$, we mean $X^n_\triangle(E_\diamond)$. Moreover, we denote by

$$t^n_{\triangle\square} : \mathcal{G}(-, X^n_\triangle) \to \mathcal{G}(-, X^{n+\delta_{\triangle\square}}_\square)$$

the $\mathcal{G}$-module map induced by the adjacent unit in $\mathcal{G}(X^n_\triangle, X^{n+\delta_{\triangle\square}}_\square)$ (see Definition 1.8), where

$$\delta_{\triangle\square} = \begin{cases} 1 & \text{if } (\triangle, \square) = (+, \bullet), \\ 0 & \text{otherwise.} \end{cases}$$

Definition 2.14 We denote by $\mathcal{M}_\mathcal{G}$ the $\mathcal{G}$-module defined by

$$\mathcal{M}_\mathcal{G} = \left[\begin{tikzcd}[column sep=small] \cdots \arrow[dr, "t^{-1}_{+\bullet}"] & \mathcal{G}(-, X^0_-) \arrow[d, "t^0_{-\bullet}"] \arrow[dr, "t^0_{-\perp}"] & \mathcal{G}(-, X^0_I) \arrow[d, "t^0_{I\perp}"] \arrow[dr, "t^0_{I\top}"] & \mathcal{G}(-, X^0_+) \arrow[d, "t^0_{+\top}"] \arrow[dr, "t^0_{+\bullet}"] & \mathcal{G}(-, X^1_-) \arrow[d, "t^1_{-\bullet}"] \arrow[dr, "t^1_{-\perp}"] & \cdots \\ \cdots & \mathcal{G}(-, X^0_\bullet) & \mathcal{G}(-, X^0_\perp) & \mathcal{G}(-, X^0_\top) & \mathcal{G}(-, X^1_\bullet) & \cdots \end{tikzcd} \right]$$

(see Definition 1.18). For practical reasons, we also consider the $\mathcal{G}$-modules

$$\mathcal{M}^n_\star := \left[\begin{tikzcd} & \mathcal{G}(-, X^n_I) \arrow[dl, "t^n_{I\perp}"'] \arrow[dr, "t^n_{I\top}"] & \\ \mathcal{G}(-, X^n_\perp) & & \mathcal{G}(-, X^n_\top) \end{tikzcd} \right], \quad n \in \mathbb{Z}.$$

We denote by $t_\mathcal{G} : \mathcal{G}(-, X^0_\bullet) \to \mathcal{M}_\mathcal{G}$ the $\mathcal{G}$-module inclusion.

Remark We can write

$$\mathcal{M}_{\mathcal{G}} = \left[\begin{tikzcd} \bigoplus_{n\in\mathbb{Z}} \mathcal{G}(-, X_-^n) \oplus \mathcal{G}(-, X_I^n) \oplus \mathcal{G}(-, X_+^n) \arrow[d, "\bigoplus_{n\in\mathbb{Z}}(t_{-\bullet}^n, t_{-\perp}^n, t_{I\perp}^n, t_{I\top}^n, t_{+\top}^n, t_{+\bullet}^n)"] \\ \bigoplus_{n\in\mathbb{Z}} \mathcal{G}(-, X_\bullet^n) \oplus \mathcal{G}(-, X_\perp^n) \oplus \mathcal{G}(-, X_\top^n) \end{tikzcd}\right]$$

and also

$$\mathcal{M}_{\mathcal{G}} = \left[\begin{tikzcd} & \bigoplus_{n\in\mathbb{Z}}(\mathcal{G}(-, X_-^n) \oplus \mathcal{G}(-, X_+^n)) \arrow[dl, "\bigoplus_{n\in\mathbb{Z}}(t_{-\perp}^n, t_{+\top}^n)"'] \arrow[dr, "\bigoplus_{n\in\mathbb{Z}}(t_{-\bullet}^n, t_{+\bullet}^n)"] & \\ \bigoplus_{n\in\mathbb{Z}} \mathcal{M}_\star^n & & \bigoplus_{n\in\mathbb{Z}} \mathcal{G}(-, X_\bullet^n) \end{tikzcd}\right].$$

The following two lemmas are analogs of Lemmas 2.9 and 2.10, respectively. They will be used later in order to apply Proposition 1.17.

Lemma 2.15 *For every w in $W_{\mathcal{G}}$, the complex $\mathcal{M}_{\mathcal{G}}(\mathrm{Cone}\, w)$ is acyclic.*

Proof Let w be the morphism in $W_{\mathcal{G}} \cap \mathcal{G}(X_I^k(E), X_\top^k(E))$ (the proof is analogous for the morphism in $W_{\mathcal{G}} \cap \mathcal{G}(X_I^k(E), X_\perp^k(E))$). Then

$$\begin{aligned} \mathcal{M}_{\mathcal{G}}(\mathrm{Cone}\, w) &= \mathrm{Cone}\big(\mathcal{M}_{\mathcal{G}}(X_\top^k(E)) \xrightarrow{\mu^2_{\mathcal{M}_{\mathcal{G}}}(w,-)} \mathcal{M}_{\mathcal{G}}(X_I^k(E))\big) \\ &= \bigoplus_n \mathrm{Cone}\big(\mathcal{G}(X_\top^k(E), X_\top^n) \xrightarrow{\mu^2_{\mathcal{M}_{\mathcal{G}}}(w,-)} \mathcal{M}_\star^n(X_I^k(E))\big). \end{aligned}$$

We want to prove that $\mu^2_{\mathcal{M}_{\mathcal{G}}}(w,-)\colon \mathcal{G}(X_\top^k(E), X_\top^n) \to \mathcal{M}_\star^n(X_I^k(E))$ is a quasi-isomorphism for every n. Observe that the following diagram of chain complexes is commutative:

$$\begin{tikzcd} \mathcal{G}(X_\top^k(E), X_\top^n) \arrow[r, "\mu^2_{\mathcal{M}_{\mathcal{G}}}(w,-)"] \arrow[d, equal] & \mathcal{M}_\star^n(X_I^k(E)) \\ \mathcal{G}(X_\top^k(E), X_\top^n) \arrow[r, "\mu^2_{\mathcal{G}}(w,-)"] & \mathcal{G}(X_I^k(E), X_\top^n) \arrow[u, hook] \end{tikzcd}$$

The rightmost vertical arrow is injective, so its cone is quasi-isomorphic to its cokernel, which is the cone of $t_{I\perp}^n\colon \mathcal{G}(X_I^k(E), X_I^n) \to \mathcal{G}(X_I^k(E), X_\perp^n)$. Since the latter map is a quasi-isomorphism, the cone of $\mu^2_{\mathcal{M}_{\mathcal{G}}}(w,-)\colon \mathcal{G}(X_\top^k(E), X_\top^n) \to \mathcal{M}_\star^n(X_I^k(E))$ is quasi-isomorphic to the cone of

$$\mu^2_{\mathcal{G}}(w,-)\colon \mathcal{G}(X_\top^k(E), X_\top^n) \to \mathcal{G}(X_I^k(E), X_\top^n).$$

The latter map is a quasi-isomorphism, so we conclude that $\mu^2_{\mathcal{M}_{\mathcal{G}}}(w,-)\colon \mathcal{G}(X_\top^k(E), X_\top^n) \to \mathcal{M}_\star^n(X_I^k(E))$ is a quasi-isomorphism for every n, and thus $\mathcal{M}_{\mathcal{G}}(\mathrm{Cone}\, w)$ is acyclic.

Now let w be the morphism in $W_{\mathcal{G}} \cap \mathcal{G}(X_+^k(E), X_\top^k(E))$ (the proof is analogous for the morphism in $W_{\mathcal{G}} \cap \mathcal{G}(X_-^k(E), X_\perp^k(E))$). Then

$$\begin{aligned} \mathcal{M}_{\mathcal{G}}(\mathrm{Cone}\, w) &= \mathrm{Cone}\big(\mathcal{M}_{\mathcal{G}}(X_\top^k(E)) \xrightarrow{\mu^2_{\mathcal{M}_{\mathcal{G}}}(w,-)} \mathcal{M}_{\mathcal{G}}(X_+^k(E))\big) \\ &= \bigoplus_n \mathrm{Cone}\big(\mathcal{G}(X_\top^k(E), X_\top^n) \xrightarrow{\mu^2_{\mathcal{M}_{\mathcal{G}}}(w,-)} K^n\big), \end{aligned}$$

where

$$K^n = \left[\begin{tikzcd}[column sep=small] & \mathcal{G}(X_+^k(E), X_+^n) \arrow[dl, "t^n_{+\top}"'] \arrow[dr, "t^n_{+\bullet}"] & \\ \mathcal{G}(X_+^k(E), X_\top^n) & & \mathcal{G}(X_+^k(E), X_\bullet^{n+1}) \end{tikzcd}\right].$$

Observe that $\mu^2_{\mathcal{M}_\mathcal{G}}(w,-)\colon \mathcal{G}(X_\top^k(E), X_\top^n) \to K^n$ is injective (it is basically an inclusion once we unravel the definitions), so its cone is quasi-isomorphic to its cokernel, which is the cone of

$$t^n_{+\bullet}\colon \mathcal{G}(X_+^k(E), X_+^n) \to \mathcal{G}(X_+^k(E), X_\bullet^{n+1}).$$

The latter map is a quasi-isomorphism because τ is a quasi-equivalence. This implies that the cone of $\mu^2_{\mathcal{M}_\mathcal{G}}(w,-)\colon \mathcal{G}(X_\top^k(E), X_\top^n) \to K^n$ is acyclic for every n, and thus $\mathcal{M}_\mathcal{G}(\operatorname{Cone} w)$ is acyclic.

It remains to consider a morphism w in $W_\mathcal{G} \cap \mathcal{G}(X_+^k(E), X_\bullet^{k+1}(E))$ (the proof is analogous for the morphism in $W_\mathcal{G} \cap \mathcal{G}(X_-^k(E), X_\bullet^k(E))$). Then

$$\begin{aligned} \mathcal{M}_\mathcal{G}(\operatorname{Cone} w) &= \operatorname{Cone}\bigl(\mathcal{M}_\mathcal{G}(X_\bullet^{k+1}(E)) \xrightarrow{\mu^2_{\mathcal{M}_\mathcal{G}}(w,-)} \mathcal{M}_\mathcal{G}(X_+^k(E))\bigr) \\ &= \bigoplus_n \operatorname{Cone}\bigl(\mathcal{G}(X_\bullet^{k+1}(E), X_\bullet^n) \xrightarrow{\mu^2_{\mathcal{M}_\mathcal{G}}(w,-)} K^n\bigr), \end{aligned}$$

where

$$K^n = \left[\begin{tikzcd}[column sep=small] & \mathcal{G}(X_+^k(E), X_+^{n-1}) \arrow[dl, "t^{n-1}_{+\top}"'] \arrow[dr, "t^{n-1}_{+\bullet}"] & \\ \mathcal{G}(X_+^k(E), X_\top^{n-1}) & & \mathcal{G}(X_+^k(E), X_\bullet^n) \end{tikzcd}\right].$$

Observe that $\mu^2_{\mathcal{M}_\mathcal{G}}(w,-)\colon \mathcal{G}(X_\bullet^{k+1}(E), X_\bullet^n) \to K^n$ is injective, so its cone is quasi-isomorphic to its cokernel, which is the cone of $t^{n-1}_{+\top}\colon \mathcal{G}(X_+^k(E), X_+^{n-1}) \to \mathcal{G}(X_+^k(E), X_\top^{n-1})$. The latter map is a quasi-isomorphism, so we conclude that the cone of $\mu^2_{\mathcal{M}_\mathcal{G}}(w,-)\colon \mathcal{G}(X_\bullet^{k+1}(E), X_\bullet^n) \to K^n$ is acyclic for every n, and thus $\mathcal{M}_\mathcal{G}(\operatorname{Cone} w)$ is acyclic. □

Lemma 2.16 *The $\mathcal{H}$-module map ${}_{W_\mathcal{G}^{-1}}t_\mathcal{G}\colon {}_{W_\mathcal{G}^{-1}}\mathcal{G}(-, X_\bullet^0) \to {}_{W_\mathcal{G}^{-1}}\mathcal{M}_\mathcal{G}$ is a quasi-isomorphism.*

Proof We fix an object X in $\mathcal{G}$, and we want to prove that ${}_{W_\mathcal{G}^{-1}}t_\mathcal{G}\colon {}_{W_\mathcal{G}^{-1}}\mathcal{G}(X, X_\bullet^0) \to {}_{W_\mathcal{G}^{-1}}\mathcal{M}_\mathcal{G}(X)$ is a quasi-isomorphism. Observe that

$${}_{W_\mathcal{G}^{-1}}\mathcal{M}_\mathcal{G} := \left[\begin{tikzcd}[column sep=small] \cdots \arrow[dr] & \mathcal{G}[W_\mathcal{G}^{-1}](-, X_+^{-1}) \arrow[d, "{}_{W_\mathcal{G}^{-1}}t^{-1}_{+\top}"] \arrow[dr, "{}_{W_\mathcal{G}^{-1}}t^{-1}_{+\bullet}"'] & \mathcal{G}[W_\mathcal{G}^{-1}](-, X_-^0) \arrow[d, "{}_{W_\mathcal{G}^{-1}}t^0_{-\bullet}"] \arrow[dr, "{}_{W_\mathcal{G}^{-1}}t^0_{-\perp}"] & \cdots \\ \cdots & \mathcal{G}[W_\mathcal{G}^{-1}](-, X_\top^{-1}) & \mathcal{G}[W_\mathcal{G}^{-1}](-, X_\bullet^0) & \cdots \end{tikzcd}\right]$$

and that the chain map ${}_{W_{\mathcal{G}}^{-1}}t_{\mathcal{G}} \colon {}_{W_{\mathcal{G}}^{-1}}\mathcal{G}(X, X^0_\bullet) \to {}_{W_{\mathcal{G}}^{-1}}\mathcal{M}_{\mathcal{G}}(X)$ is the inclusion. The cone of the latter chain map is then quasi-isomorphic to its cokernel, which can be written $K' \oplus K''$, where

$$K' = \left[\begin{tikzcd} \cdots \arrow[dr] & \mathcal{G}[W_{\mathcal{G}}^{-1}](X, X^{-1}_I) \arrow[d, "{}_{W_{\mathcal{G}}^{-1}}t^{-1}_{I\perp}"] \arrow[dr, "{}_{W_{\mathcal{G}}^{-1}}t^{-1}_{I\top}"'] & \mathcal{G}[W_{\mathcal{G}}^{-1}](X, X^{-1}_+) \arrow[d, "{}_{W_{\mathcal{G}}^{-1}}t^{-1}_{+\top}"] \\ \cdots & \mathcal{G}[W_{\mathcal{G}}^{-1}](X, X^{-1}_\perp) & \mathcal{G}[W_{\mathcal{G}}^{-1}](X, X^{-1}_\top) \end{tikzcd}\right]$$

and

$$K'' = \left[\begin{tikzcd} \mathcal{G}[W_{\mathcal{G}}^{-1}](X, X^0_-) \arrow[dr, "{}_{W_{\mathcal{G}}^{-1}}t^0_{-\perp}"'] & \mathcal{G}[W_{\mathcal{G}}^{-1}](X, X^0_I) \arrow[d, "{}_{W_{\mathcal{G}}^{-1}}t^0_{I\perp}"'] \arrow[dr, "{}_{W_{\mathcal{G}}^{-1}}t^0_{I\top}"] & \cdots \\ & \mathcal{G}[W_{\mathcal{G}}^{-1}](X, X^0_\perp) & \cdots \end{tikzcd}\right].$$

Observe that the maps defining the chain complex structures in K' and K'' are all quasi-isomorphisms (see [23, Lemma 3.12]). Thus it is not difficult to show using an increasing exhaustive and bounded from below filtration of K' and K'' that these complexes are acyclic (compare the proof of Lemma 2.10). This implies that the map ${}_{W_{\mathcal{G}}^{-1}}t_{\mathcal{G}} \colon {}_{W_{\mathcal{G}}^{-1}}\mathcal{G}(X, X^0_\bullet) \to {}_{W_{\mathcal{G}}^{-1}}\mathcal{M}_{\mathcal{G}}(X)$ is a quasi-isomorphism. □

2.4 Proof of the first result

Let τ be a quasi-autoequivalence of an A_∞-category $\mathcal{A}$ equipped with a compatible $\mathbb{Z}$-splitting of $\mathrm{ob}(\mathcal{A})$. Assume that τ is strict and acts bijectively on hom-sets.

Observe that there is a strict A_∞-functor $\sigma \colon \mathcal{A} \to \mathcal{A}_\tau$ which sends $X^n(E)$ to E, and which sends $x \in \mathcal{A}(X^i(E_1), X^j(E_2))$ to $[x] \in \mathcal{A}_\tau(E_1, E_2)$. Besides, let $\pi \colon \mathcal{C} \to \mathcal{A}$ be the A_∞-functor induced by the commutative diagram

$$\begin{tikzcd} \mathcal{A} \arrow[r, "\mathrm{id}"] \arrow[d, "\mathrm{id}"] & \mathcal{A} \arrow[d, "\mathrm{id}"] \\ \mathcal{A} \arrow[r, "\mathrm{id}"] & \mathcal{A} \end{tikzcd}$$

(see Proposition 1.21). Then the diagram of Adams-graded A_∞-categories

$$\begin{tikzcd} \mathcal{A}_- \sqcup \mathcal{A}_+ \arrow[r, "\mathrm{id} \sqcup \tau"] \arrow[d, "\iota_\perp \sqcup \iota_\top"'] & \mathcal{A}_\bullet \arrow[d, "\sigma"] \\ \mathcal{C} \arrow[r, "\sigma \circ \pi"] & \mathcal{A}_\tau \end{tikzcd}$$

is commutative because $\sigma \circ \tau = \sigma$. Moreover, the induced A_∞-functor $\Phi \colon \mathcal{G} \to \mathcal{A}_\tau$ is strict, and it sends $W_{\mathcal{G}}$ to the set of units in $\mathcal{A}_\tau$. Let

$$\widetilde{\Phi} \colon \mathcal{H} \to \mathcal{A}_\tau[\{\text{units}\}^{-1}]$$

be the A_∞-functor induced by Φ.

According to Lemma 2.13, $\mathcal{H}$ is quasi-equivalent to the mapping torus of τ. Moreover, $\mathcal{A}_\tau[\{\text{units}\}^{-1}]$ is quasi-equivalent to $\mathcal{A}_\tau$. Thus, Theorem 2.4 will follow if we prove that $\widetilde{\Phi}$ is a quasi-equivalence. Our strategy is to apply Proposition 1.17. Observe that it suffices to prove that

$$\widetilde{\Phi}\colon \mathcal{H}(X,Y) \to \mathcal{A}_\tau[\{\text{units}\}^{-1}](\Phi X, \Phi Y)$$

is a quasi-isomorphism for every object X, Y in $\mathcal{A}^0_\bullet$ (recall that $\mathcal{A}^0_\bullet$ denotes the subcategory of $\mathcal{A}_\bullet$ generated by the objects $X^0_\bullet(E)$, $E \in \mathcal{E}$) because every object of $\mathcal{G}$ can be related to one of $\mathcal{A}^0_\bullet$ by a zigzag of morphisms in $W_\mathcal{G}$, which are quasi-isomorphisms in $\mathcal{H}$ (see [23, Lemma 3.12]).

In the following, we fix some element $E_\diamond \in \mathcal{E}$. When we write an object $X^n_\triangle$ without specifying the element of $\mathcal{E}$, we mean $X^n_\triangle(E_\diamond)$. We consider the corresponding $\mathcal{G}$-module $\mathcal{M}_\mathcal{G}$ and the $\mathcal{G}$-module map $t_\mathcal{G}\colon \mathcal{G}(-, X^0_\bullet) \to \mathcal{G}$ of Definition 2.14. Moreover, we set

$$\mathcal{M}_{\mathcal{A}_\tau} := \mathcal{A}_\tau(-, E_\diamond) \quad \text{and} \quad t_{\mathcal{A}_\tau} := \mathrm{id}\colon \mathcal{A}_\tau(-, E_\diamond) \to \mathcal{M}_{\mathcal{A}_\tau}.$$

Lemma 2.17 *There exists a $\mathcal{G}$-module map $t_0\colon \mathcal{M}_\mathcal{G} \to \Phi^*\mathcal{M}_{\mathcal{A}_\tau}$ (see Definition 1.11 for the pullback functor) such that:*

(1) *The following diagram of $\mathcal{G}$-modules commutes:*

$$\begin{tikzcd} \mathcal{G}(-, X^0_\bullet) \arrow[r, "t_\Phi"] \arrow[d, "t_\mathcal{G}"] & \Phi^*\mathcal{A}_\tau(-, E_\diamond) \arrow[d, "\Phi^* t_{\mathcal{A}_\tau} = \mathrm{id}"] \\ \mathcal{M}_\mathcal{G} \arrow[r, "t_0"] & \Phi^*\mathcal{M}_{\mathcal{A}_\tau} \end{tikzcd}$$

(see Definition 1.13 for the map t_Φ).

(2) *For every $E \in \mathcal{E}$, the map $t_0\colon \mathcal{M}_\mathcal{G}(X^0_\bullet(E)) \to \Phi^*\mathcal{M}_{\mathcal{A}_\tau}(X^0_\bullet(E))$ is a quasi-isomorphism.*

Proof Observe that the diagram of $\mathcal{G}$-modules

$$\begin{tikzcd} \mathcal{G}(-, X^n_I) \arrow[r, "t^n_{I\top}"] \arrow[d, "t^n_{I\perp}"] & \mathcal{G}(-, X^n_\top) \arrow[d, "t_\Phi"] \\ \mathcal{G}(-, X^n_\perp) \arrow[r, "t_\Phi"] & \Phi^*\mathcal{M}_{\mathcal{A}_\tau} \end{tikzcd}$$

is commutative, so that it induces a $\mathcal{G}$-module map $\mathcal{M}^n_\star \to \Phi^*\mathcal{M}_{\mathcal{A}_\tau}$ (see Proposition 1.19). Now observe that the following diagram of $\mathcal{G}$-modules commutes:

$$\begin{tikzcd} \bigoplus_{n\in\mathbb{Z}}(\mathcal{G}(-, X^n_-) \oplus \mathcal{G}(-, X^n_+)) \arrow[r, "\bigoplus_{n\in\mathbb{Z}}(t^n_{-\bullet} \oplus t^n_{+\bullet})"] \arrow[d, "\bigoplus_{n\in\mathbb{Z}}(t^n_{-\perp} \oplus t^n_{+\top})"] & \bigoplus_{n\in\mathbb{Z}} \mathcal{G}(-, X^n_\bullet) \arrow[d, "t_\Phi"] \\ \bigoplus_{n\in\mathbb{Z}} \mathcal{M}^n_\star \arrow[r] & \Phi^*\mathcal{M}_{\mathcal{A}_\tau} \end{tikzcd}$$

We let $t_0\colon \mathcal{M}_{\mathcal{G}} \to \Phi^*\mathcal{M}_{\mathcal{A}_\tau}$ be the induced $\mathcal{G}$-module map. It is then easy to verify that the following diagram of $\mathcal{G}$-modules is commutative:

$$\begin{tikzcd} \mathcal{G}(-, X^0_\bullet) \arrow[r, "t_\Phi"] \arrow[d, "t_{\mathcal{G}}"] & \Phi^*\mathcal{A}_\tau(-, E_\diamond) \arrow[d, "\Phi^* t_{\mathcal{A}_\tau} = \mathrm{id}"] \\ \mathcal{M}_{\mathcal{G}} \arrow[r, "t_0"] & \Phi^*\mathcal{M}_{\mathcal{A}_\tau} \end{tikzcd}$$

We now prove the second part of the lemma. We have

$$\mathcal{M}_{\mathcal{G}}(X^0_\bullet(E)) = \bigoplus_n \mathcal{G}(X^0_\bullet(E), X^n_\bullet) = \bigoplus_n \mathcal{A}(X^k(E), X^n)$$

and

$$t_0\colon \bigoplus_n \mathcal{A}(X^k(E), X^n) = \mathcal{M}_{\mathcal{G}}(X^0_\bullet(E)) \to \Phi^*\mathcal{M}_{\mathcal{A}_\tau}(X^0_\bullet(E)) = \mathcal{A}_\tau(E, E_\diamond)$$

is the sum of the projections, which is an isomorphism. □

Lemma 2.18 *For every $E \in \mathcal{E}$, the chain map*

$$\widetilde{\Phi}\colon \mathcal{H}(X^0_\bullet(E), X^0_\bullet) \to \mathcal{A}_\tau[\{\text{units}\}^{-1}](E, E_\diamond)$$

is a quasi-isomorphism.

Proof According to Lemmas 2.15, 2.16 and 2.17, the assumptions of Proposition 1.17 are satisfied. □

As explained above, Theorem 2.4 follows from Lemma 2.18 since $\mathcal{H}$ is quasi-equivalent to the mapping torus of τ (see Lemma 2.13) and $\mathcal{A}_\tau[\{\text{units}\}^{-1}]$ is quasi-equivalent to $\mathcal{A}_\tau$.

2.5 Proof of the second result

Let τ be a quasi-autoequivalence of an A_∞-category $\mathcal{A}$ equipped with a compatible $\mathbb{Z}$-splitting of $\mathrm{ob}(\mathcal{A})$. Assume that the following holds:

(1) $\mathcal{A}$ is weakly directed with respect to the $\mathbb{Z}$-splitting of $\mathrm{ob}(\mathcal{A})$ (see Definition 2.2).

(2) There exists a closed degree 0 bimodule map $f\colon \mathcal{A}_m(-,-) \to \mathcal{A}_m(-, \tau(-))$ (see Definitions 1.4 and 1.5) such that $f\colon \mathcal{A}_m(X^i(E), X^j(E')) \to \mathcal{A}_m(X^i(E), X^{j+1}(E'))$ is a quasi-isomorphism for every $i < j$ and $E, E' \in \mathcal{E}$.

Remark It follows from Corollary 1.10 and τ being a quasi-equivalence that the chain maps

$$\begin{aligned} \mu^2_{\mathcal{A}_m}(-, f(e_{X^j(E)}))\colon &\quad \mathcal{A}_m(X^i(E'), X^j(E)) \to \mathcal{A}_m(X^i(E'), X^{j+1}(E)), \\ \mu^2_{\mathcal{A}_m}(f(e_{X^j(E)}), -)\colon &\quad \mathcal{A}_m(X^{j+1}(E), X^{k+1}(E')) \to \mathcal{A}_m(X^j(E), X^{k+1}(E')), \end{aligned}$$

are quasi-isomorphisms for every $i < j < k$ and $E, E' \in \mathcal{E}$.

In the following, we set

$$c_n(E) := f(e_{X^n(E)})$$

for every $n \in \mathbb{Z}$, $E \in \mathcal{E}$, and

$$W_{\mathcal{A}_m} := \{c_n(E) \mid n \in \mathbb{Z}, E \in \mathcal{E}\} \cup \{\text{units of } \mathcal{A}_m\}.$$

2.5.1 Generalized homotopy Recall that we introduced a functor $\mathcal{B} \mapsto \mathcal{B}_m$ from the category of Adams-graded A_∞-categories to the category of (non-Adams-graded) A_∞-categories. Also, recall that we introduced Adams-graded A_∞-categories $\mathcal{C}$ and $\mathcal{G}$ in Definitions 2.11 and 2.12, respectively. Observe that $\mathcal{C}_m$ and $\mathcal{G}_m$ are the Grothendieck constructions of the diagrams

$$\begin{tikzcd} (\mathcal{A}_I)_m \arrow[r, "\mathrm{id}"] \arrow[d, "\mathrm{id}"'] & (\mathcal{A}_\top)_m \\ (\mathcal{A}_\perp)_m & \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} (\mathcal{A}_-)_m \sqcup (\mathcal{A}_+)_m \arrow[r, "\mathrm{id}\sqcup\tau"] \arrow[d, "\iota_\perp \sqcup \iota_\top"'] & (\mathcal{A}_\bullet)_m \\ \mathcal{C}_m & \end{tikzcd}$$

respectively. We denote by $W_{\mathcal{C}_m}$ the set of adjacent units in $\mathcal{C}_m$, and by $W_{\mathcal{G}_m}$ the union of $W_{\mathcal{C}_m}$ and the set of adjacent units in $\mathcal{G}_m$.

We would like to define an A_∞-functor $\Psi_m\colon \mathcal{G}_m \to \mathcal{A}_m$ which sends $W_{\mathcal{G}_m}$ to $W_{\mathcal{A}_m}$. According to Proposition 1.21, it is enough to prove the following result.

Lemma 2.19 *There exists an A_∞-functor $\eta\colon \mathcal{C}_m \to \mathcal{A}_m$ which sends $W_{\mathcal{C}_m}$ to $W_{\mathcal{A}_m}$, and such that*

$$\eta \circ \iota_I = \eta \circ \iota_\perp = \mathrm{id}, \quad \eta \circ \iota_\top = \tau.$$

Proof We first define η to be id on $(\mathcal{A}_\perp)_m$, $(\mathcal{A}_I)_m$. and to be τ on $(\mathcal{A}_\top)_m$. Observe that this completely defines η on the objects. Also, we ask for η to act as the identity on the sequences involving an adjacent morphism from $(\mathcal{A}_I)_m$ to $(\mathcal{A}_\perp)_m$.

It remains to define η on the sequences involving an adjacent morphism from $(\mathcal{A}_I)_m$ to $(\mathcal{A}_\top)_m$. Consider a sequence of morphisms

$$\begin{aligned}(x_0, \dots, x_{p+q}) &\\ \in \mathcal{C}_m(X_I^{i_0}(E_0), X_I^{i_1}(E_1)) &\times \dots \times \mathcal{C}_m(X_I^{i_{p-1}}(E_{p-1}), X_I^{i_p}(E_p)) \times \mathcal{C}_m(X_I^{i_p}(E_p), X_\top^{i_{p+1}}(E_{p+1})) \\ &\times \mathcal{C}_m(X_\top^{i_{p+1}}(E_{p+1}), X_\top^{i_{p+2}}(E_{p+2})) \times \dots \times \mathcal{C}_m(X_\top^{i_{p+q}}(E_{p+q}), X_\top^{i_{p+q+1}}(E_{p+q+1})).\end{aligned}$$

Observe that

$$\mathcal{C}_m(X_I^i(E), X_I^j(E')) = \mathcal{C}_m(X_I^i(E), X_\top^j(E')) = \mathcal{C}_m(X_\top^i(E), X_\top^j(E')) = \mathcal{A}_m(X^i(E), X^j(E)).$$

Then we set

$$\eta(x_0, \dots, x_{p+q}) := f(x_0, \dots, x_{p-1}, x_p, x_{p+1}, \dots, x_{p+q+1}) \in \mathcal{A}_m(X^{i_0}(E_0), \tau X^{i_{p+q+1}}(E_{p+q+1})).$$

The functor η we defined satisfies the A_∞-relations because $f\colon \mathcal{A}_m(-,-) \to \mathcal{A}_m(-,\tau(-))$ is a closed degree 0 bimodule map. Moreover, η sends $W_{\mathcal{C}_m}$ to $W_{\mathcal{A}_m}$ by construction. □

Remark First observe that

$$\mathcal{C}yl_{\mathcal{A}_m} = \mathcal{C}_m[W_{\mathcal{C}_m}^{-1}] = (\mathcal{C}[W_{\mathcal{C}}^{-1}])_m = (\mathcal{C}yl_{\mathcal{A}})_m.$$

According to Lemma 2.19, the functor η induces an A_∞-functor $\tilde{\eta}\colon \mathcal{C}yl_{\mathcal{A}_m} \to \mathcal{A}_m[W_{\mathcal{A}_m}^{-1}]$. Moreover, Lemma 2.19 implies that the following diagram commutes:

$$\begin{tikzcd}
(\mathcal{A}_+)_m \arrow[d, "\lambda_{\mathcal{C}_m}\circ\iota_\top"'] \arrow[dr, bend left, "\lambda_{\mathcal{A}_m}\circ\tau"] & \\
\mathcal{C}yl_{\mathcal{A}_m} \arrow[r, "\tilde{\eta}"] & \mathcal{A}_m[W_{\mathcal{A}_m}^{-1}] \\
(\mathcal{A}_-)_m \arrow[u, "\lambda_{\mathcal{C}_m}\circ\iota_\perp"] \arrow[ur, bend right, "\lambda_{\mathcal{A}_m}"'] &
\end{tikzcd}$$

($\lambda_{\mathcal{A}_m}\colon \mathcal{A}_m \to \mathcal{A}_m[W_{\mathcal{A}_m}^{-1}]$ and $\lambda_{\mathcal{C}_m}\colon \mathcal{C}_m \to \mathcal{C}_m[W_{\mathcal{C}_m}^{-1}]$ denote the localization functors). Since $\mathcal{C}yl_{\mathcal{A}_m}$ should be thought of as a cylinder object for $\mathcal{A}_m$ (see Proposition 1.24), we should think that the functors $\lambda_{\mathcal{A}_m}$ and $\lambda_{\mathcal{A}_m}\circ\tau$ are homotopic (even if they do not act the same way on objects) and that $\tilde{\eta}$ is a generalized homotopy between them (see Proposition 1.25 for a justification of this terminology).

2.5.2 Relation between $\mathcal{G}$ and $\mathcal{A}_m$ Using the A_∞-functor $\eta\colon \mathcal{C}_m \to \mathcal{A}_m$ of Lemma 2.19, we get a commutative diagram of (non-Adams-graded) A_∞-categories

$$\begin{tikzcd}
(\mathcal{A}_-)_m \sqcup (\mathcal{A}_+)_m \arrow[r, "\mathrm{id}\sqcup\tau"] \arrow[d, "\iota_\perp\sqcup\iota_\top"'] & (\mathcal{A}_\bullet)_m \arrow[d, "\mathrm{id}"] \\
\mathcal{C}_m \arrow[r, "\eta"] & \mathcal{A}_m
\end{tikzcd}$$

and the induced A_∞-functor $\Psi_m\colon \mathcal{G}_m \to \mathcal{A}_m$ (see Proposition 1.21) sends $W_{\mathcal{G}_m}$ to $W_{\mathcal{A}_m}$ (see Lemma 2.19). Let

$$\tilde{\Psi}_m\colon \mathcal{H}_m = \mathcal{G}_m[W_{\mathcal{G}_m}^{-1}] \to \mathcal{A}_m[W_{\mathcal{A}_m}^{-1}]$$

be the A_∞-functor induced by Ψ_m. Observe that, since $\mathcal{A}$ is assumed to be weakly directed and since the Adams degree of $\mathcal{A}$ comes from the $\mathbb{Z}$-splitting of $\mathrm{ob}(\mathcal{A})$ (see Definition 2.2), $\mathcal{H}$ is concentrated in nonnegative Adams degree. In particular, we can apply the adjunction of Definition 1.29 to $\tilde{\Psi}_m$, which gives an A_∞-functor

$$\tilde{\Psi}\colon \mathcal{H} = \mathcal{G}[W_{\mathcal{G}}^{-1}] \to \mathbb{F}[t_m] \otimes \mathcal{A}_m[W_{\mathcal{A}_m}^{-1}].$$

We would like to prove that for every Adams degree $j \geq 1$, and for every objects X, Y in $\mathcal{A}_\bullet^0$ (recall that $\mathcal{A}_\bullet^0$ denotes the subcategory of $\mathcal{A}_\bullet$ generated by the objects $X_\bullet^0(E)$, $E \in \mathcal{E}$), the map

$$\tilde{\Psi}\colon \mathcal{H}(X,Y)^{*,j} \to (\mathbb{F}[t_m] \otimes \mathcal{A}_m[W_{\mathcal{A}_m}^{-1}])(\Psi X, \Psi Y)^{*,j} = \mathbb{F}_m^j \otimes \mathcal{A}_m[W_{\mathcal{A}_m}^{-1}](\Psi X, \Psi Y)$$

is a quasi-isomorphism, ($\mathbb{F}_m^j$ is the vector space generated by t_m^j; also recall that if V is an Adams-graded vector space, $V^{*,j}$ denotes the subspace of Adams degree j elements). Our strategy is once again to apply Proposition 1.17.

In the following we fix some element $E_\diamond \in \mathcal{E}$. When we write $X^n_\triangle$ or c_n without specifying the element of $\mathcal{E}$, we mean $X^n_\triangle(E_\diamond)$ or $c_n(E_\diamond)$, respectively. We consider the corresponding $\mathcal{G}$-module $\mathcal{M}_\mathcal{G}$, and the $\mathcal{G}$-module map $t_\mathcal{G}\colon \mathcal{G}(-, X^n_\triangle) \to \mathcal{G}$ of Definition 2.14. Moreover, we consider the $\mathcal{A}_m$-module $\mathcal{M}_{\mathcal{A}_m}$ and the $\mathcal{A}_m$-module maps

$$t^n_{\mathcal{A}_m}\colon \mathcal{A}_m(-, X^n) \to \mathcal{M}_{\mathcal{A}_m}, \quad n \in \mathbb{Z},$$

associated to the morphisms $(c_n)_{n\in\mathbb{Z}}$ as in Definition 2.6.

The following result is a first step in order to define a $\mathcal{G}_m$-module map $(t_0)_m\colon (\mathcal{M}_\mathcal{G})_m \to \Psi^*_m\mathcal{M}_{\mathcal{A}_m}$ as in Proposition 1.17.

Lemma 2.20 *For every $n \in \mathbb{Z}$, the diagram of $\mathcal{G}_m$-modules*

$$\begin{tikzcd}
\mathcal{G}_m(-, X^n_I) \arrow[r, "t^n_{I\top}"] \arrow[d, "t^n_{I\bot}"] & \mathcal{G}_m(-, X^n_\top) \arrow[d, "\Psi^*_m t^{n+1}_{\mathcal{A}_m}\circ t_{\Psi_m}"] \\
\mathcal{G}_m(-, X^n_\bot) \arrow[r, "\Psi^*_m t^n_{\mathcal{A}_m}\circ t_{\Psi_m}"] & \Psi^*_m\mathcal{M}_{\mathcal{A}_m}
\end{tikzcd}$$

commutes up to homotopy.

Proof First observe that $\iota_I^*\mathcal{G}_m(-, X^n_I) = \mathcal{A}_m(-, X^n)$, and $\iota_I^*\Psi^*_m\mathcal{M}_{\mathcal{A}_m} = \mathcal{M}_{\mathcal{A}_m}$ because $\Psi_m \circ \iota_I = \mathrm{id}$ (see Remark 1.12). Moreover, it suffices to show that the $\mathcal{A}_m$-module maps

$$\iota_I^*(\Psi^*_m t^n_{\mathcal{A}_m} \circ t_{\Psi_m} \circ t^n_{I\bot}) = t^n_{\mathcal{A}_m} \circ \iota_I^*(t_{\Psi_m} \circ t^n_{I\bot})\colon \mathcal{A}_m(-, X^n) \to \mathcal{M}_{\mathcal{A}_m}$$

and

$$\iota_I^*(\Psi^*_m t^{n+1}_{\mathcal{A}_m} \circ t_{\Psi_m} \circ t^n_{I\top}) = t^{n+1}_{\mathcal{A}_m} \circ \iota_I^*(t_{\Psi_m} \circ t^n_{I\top})\colon \mathcal{A}_m(-, X^n) \to \mathcal{M}_{\mathcal{A}_m}$$

are homotopic because

$$\mathcal{G}_m(X^k_\triangle, X^n_I) = 0 \quad \text{if } \triangle \neq I.$$

On the one hand,

$$t^n_{\mathcal{A}_m} \circ \iota_I^*(t_{\Psi_m} \circ t^n_{I\bot}) = t^n_{\mathcal{A}_m}.$$

On the other hand, $\iota_I^*(t_{\Psi_m} \circ t^n_{I\top})\colon \mathcal{A}_m(-, X^n) \to \mathcal{A}_m(-, X^{n+1})$ is closed (as composition and pullback of closed module maps), and

$$\iota_I^*(t_{\Psi_m} \circ t^n_{I\top})(e_{X^n}) = \eta(e_{X^n}) = c_n$$

according to Lemma 2.19. Therefore, $\iota_I^*(t_{\Psi_m} \circ t^n_{I\top})$ is homotopic to t_{c_n} according to Corollary 1.10, and thus $t^{n+1}_{\mathcal{A}_m} \circ \iota_I^*(t_{\Psi_m} \circ t^n_{I\top})$ is homotopic to $t^{n+1}_{\mathcal{A}_m} \circ t_{c_n}$. Now according to Lemma 2.7, $t^{n+1}_{\mathcal{A}_m} \circ t_{c_n}$ is homotopic to $t^n_{\mathcal{A}_m}$. □

We can now state the result establishing the existence of a $\mathcal{G}_m$-module map $(t_0)_m\colon (\mathcal{M}_\mathcal{G})_m \to \Psi^*_m\mathcal{M}_{\mathcal{A}_m}$ as in Proposition 1.17.

Lemma 2.21 *There exists a $\mathcal{G}_m$-module map $(t_0)_m\colon (\mathcal{M}_\mathcal{G})_m \to \Psi^*_m\mathcal{M}_{\mathcal{A}_m}$ such that the following holds:*

(1) *The following diagram of $\mathcal{G}_m$-modules commutes*:

$$\begin{tikzcd} \mathcal{G}_m(-, X^0_\bullet) \arrow[r, "t_{\Psi_m}"] \arrow[d, "t_{\mathcal{G}_m}"] & \Psi_m^* \mathcal{A}_m(-, X^0) \arrow[d, "\Psi_m^* t^0_{\mathcal{A}_m}"] \\ (\mathcal{M}_{\mathcal{G}})_m \arrow[r, "(t_0)_m"] & \Psi_m^* \mathcal{M}_{\mathcal{A}_m} \end{tikzcd}$$

(2) *For every $E \in \mathcal{E}$ and $j \geq 1$, the induced map $t_0\colon \mathcal{M}_{\mathcal{G}}(X^0_\bullet(E)) \to \mathbb{F}[t_m] \otimes \Psi_m^* \mathcal{M}_{\mathcal{A}_m}(X^0_\bullet(E))$ (see Definition 1.29) is a quasi-isomorphism in each positive Adams degree.*

Proof Using Lemma 2.20 and Proposition 1.19, we get a $\mathcal{G}_m$-module map $t^n_\star\colon (\mathcal{M}^n_\star)_m \to \Psi_m^* \mathcal{M}_{\mathcal{A}_m}$ for every $n \in \mathbb{Z}$ (see Definition 2.14 for the $\mathcal{G}$-modules $\mathcal{M}^n_\star$). Observe that the diagram of $\mathcal{G}_m$-modules

$$\begin{tikzcd} \bigoplus_{n\in\mathbb{Z}} (\mathcal{G}_m(-, X^n_-) \oplus \mathcal{G}_m(-, X^n_+)) \arrow[r, "\bigoplus_{n\in\mathbb{Z}}(t^n_{-\bullet} \oplus t^n_{+\bullet})"] \arrow[d, "\bigoplus_{n\in\mathbb{Z}}(t^n_{-\perp} \oplus t^n_{+\top})"] & \bigoplus_{n\in\mathbb{Z}} \mathcal{G}_m(-, X^n_\bullet) \arrow[d, "\bigoplus_{n\in\mathbb{Z}} \Psi_m^* t^n_{\mathcal{A}_m} \circ t_{\Psi_m}"] \\ \bigoplus_{n\in\mathbb{Z}} (\mathcal{M}^n_\star)_m \arrow[r, "\bigoplus_{n\in\mathbb{Z}} t^n_\star"] & \Psi_m^* \mathcal{M}_{\mathcal{A}_m} \end{tikzcd}$$

is commutative (the composition is id for $-$-terms and τ for $+$-terms), so that it induces a $\mathcal{G}_m$-module map $(t_0)_m\colon (\mathcal{M}_{\mathcal{G}})_m \to \Psi_m^* \mathcal{M}_{\mathcal{A}_m}$. It is then easy to verify that the following diagram of $\mathcal{G}_m$-modules is commutative:

$$\begin{tikzcd} \mathcal{G}_m(-, X^0_\bullet) \arrow[r, "t_{\Psi_m}"] \arrow[d, "t_{\mathcal{G}_m}"] & \Psi_m^* \mathcal{A}_m(-, X^0) \arrow[d, "\Psi_m^* t^0_{\mathcal{A}_m}"] \\ (\mathcal{M}_{\mathcal{G}})_m \arrow[r, "(t_0)_m"] & \Psi_m^* \mathcal{M}_{\mathcal{A}_m} \end{tikzcd}$$

It remains to show that the map $t_0\colon \mathcal{M}_{\mathcal{G}}(X^0_\bullet(E))^{*,j} \to \mathbb{F}^j_m \otimes \mathcal{M}_{\mathcal{A}_m}(X^0(E))$ is a quasi-isomorphism for every $E \in \mathcal{E}$ and $j \geq 1$. Note that

$$\mathcal{M}_{\mathcal{G}}(X^0_\bullet(E))^{*,j} = \mathcal{G}(X^0_\bullet(E), X^j_\bullet) = \mathcal{A}(X^0, X^j)$$

and the map

$$\mathcal{A}(X^0(E), X^j) = \mathcal{M}_{\mathcal{G}}(X^0_\bullet(E))^{*,j} \xrightarrow{t_0} \mathbb{F}^j_m \otimes \mathcal{M}_{\mathcal{A}_m}(X^0(E))$$

is the inclusion. Now observe that the following diagram of chain complexes commutes:

$$\begin{tikzcd} \mathcal{A}(X^0(E), X^j) \arrow[r, "t_0"] \arrow[d, equal] & \mathbb{F}^j_m \otimes \mathcal{M}_{\mathcal{A}_m}(X^0(E)) \\ \mathbb{F}^j_m \otimes \mathcal{A}_m(X^0(E), X^j) \arrow[r, equal] & \mathbb{F}^j_m \otimes \mathcal{A}_m(X^0(E), X^j) \arrow[u, hook] \end{tikzcd}$$

The inclusion $\mathcal{A}_m(X^0(E), X^j) \hookrightarrow \mathcal{M}_{\mathcal{A}_m}(X^0(E))$ is a quasi-isomorphism according to Lemma 2.8 (observe that it is important here that j is *strictly* greater than 0). Therefore the map

$$t_0\colon \mathcal{A}(X^0(E), X^j) \to \mathbb{F}^j_m \otimes \mathcal{M}_{\mathcal{A}_m}(X^0(E))$$

is a quasi-isomorphism, which is what we needed to prove. □

Lemma 2.22 *For every $E \in \mathcal{E}$ and $j \geq 1$, the map*

$$\widetilde{\Psi}\colon \mathcal{H}(X^0_\bullet(E), X^0_\bullet)^{*,j} \to (\mathbb{F}[t_m] \otimes \mathcal{A}_m[W^{-1}_{\mathcal{A}_m}])(X^0(E), X^0)^{*,j} = \mathbb{F}^j_m \otimes \mathcal{A}_m[W^{-1}_{\mathcal{A}_m}](X^0(E), X^0)$$

is a quasi-isomorphism.

Proof Using the first part of Lemma 2.21 and Proposition 1.17, we know that there exists a chain map $u_m\colon {}_{W^{-1}_{\mathcal{G}_m}}(\mathcal{M}_\mathcal{G})_m(X^0_\bullet(E)) \to {}_{W^{-1}_{\mathcal{A}_m}}\mathcal{M}_{\mathcal{A}_m}(X^0)$ such that the following diagram of chain complexes commutes:

$$\begin{tikzcd}
\mathcal{H}_m(X^0_\bullet(E), X^0_\bullet) \arrow[r, "\widetilde{\Psi}_m"] \arrow[d, "W^{-1}_{\mathcal{G}_m} t_{\mathcal{G}_m}"] & \mathcal{A}_m[W^{-1}_{\mathcal{A}_m}](X^0(E), X^0) \arrow[d, "W^{-1}_{\mathcal{A}_m} t^0_{\mathcal{A}_m}"] \\
{}_{W^{-1}_{\mathcal{G}_m}}(\mathcal{M}_\mathcal{G})_m(X^0_\bullet(E)) \arrow[r, "u_m"] & {}_{W^{-1}_{\mathcal{A}_m}}\mathcal{M}_{\mathcal{A}_m}(\Psi_m X) \\
(\mathcal{M}_\mathcal{G})_m(X^0_\bullet(E)) \arrow[r, "(t_0)_m"] \arrow[u, hook] & \mathcal{M}_{\mathcal{A}_m}(\Psi_m X) \arrow[u, hook]
\end{tikzcd}$$

Observe that

$$\begin{aligned}
\mathcal{H}_m(X^0_\bullet(E), X^0_\bullet) &= \mathcal{H}(X^0_\bullet(E), X^0_\bullet)_m, \\
{}_{W^{-1}_{\mathcal{G}_m}}(\mathcal{M}_\mathcal{G})_m(X^0_\bullet(E)) &= {}_{W^{-1}_{\mathcal{G}}}\mathcal{M}_\mathcal{G}(X^0_\bullet(E))_m, \\
(\mathcal{M}_\mathcal{G})_m(X^0_\bullet(E)) &= \mathcal{M}_\mathcal{G}(X^0_\bullet(E))_m.
\end{aligned}$$

Applying the adjunction of Definition 1.29 to the last diagram, we get the commutative diagram of Adams-graded chain complexes

$$\begin{tikzcd}
\mathcal{H}(X^0_\bullet(E), X^0_\bullet) \arrow[r, "\widetilde{\Psi}"] \arrow[d, "W^{-1}_{\mathcal{G}} t_{\mathcal{G}}"] & \mathbb{F}[t_m] \otimes \mathcal{A}_m[W^{-1}_{\mathcal{A}_m}](X^0(E), X^0) \arrow[d, "\mathrm{id} \otimes {}_{W^{-1}_{\mathcal{A}_m}} t^0_{\mathcal{A}_m}"] \\
{}_{W^{-1}_{\mathcal{G}}}\mathcal{M}_\mathcal{G}(X^0_\bullet(E)) \arrow[r, "u"] & \mathbb{F}[t_m] \otimes {}_{W^{-1}_{\mathcal{A}_m}}\mathcal{M}_{\mathcal{A}_m}(\Psi_m X) \\
\mathcal{M}_\mathcal{G}(X^0_\bullet(E)) \arrow[r, "t_0"] \arrow[u, hook] & \mathbb{F}[t_m] \otimes \mathcal{M}_{\mathcal{A}_m}(\Psi_m X) \arrow[u, hook]
\end{tikzcd}$$

Specializing to the components of fixed Adams degree $j \geq 1$, we get the commutative diagram of chain complexes

$$\begin{tikzcd}
\mathcal{H}(X^0_\bullet(E), X^0_\bullet)^{*,j} \arrow[r, "\widetilde{\Psi}"] \arrow[d, "W^{-1}_{\mathcal{G}} t_{\mathcal{G}}"] & \mathbb{F}^j_m \otimes \mathcal{A}_m[W^{-1}_{\mathcal{A}_m}](X^0(E), X^0) \arrow[d, "\mathrm{id} \otimes {}_{W^{-1}_{\mathcal{A}_m}} t^0_{\mathcal{A}_m}"] \\
{}_{W^{-1}_{\mathcal{G}}}\mathcal{M}_\mathcal{G}(X^0_\bullet(E))^{*,j} \arrow[r, "u"] & \mathbb{F}^j_m \otimes {}_{W^{-1}_{\mathcal{A}_m}}\mathcal{M}_{\mathcal{A}_m}(\Psi_m X) \\
\mathcal{M}_\mathcal{G}(X^0_\bullet(E))^{*,j} \arrow[r, "t_0"] \arrow[u, hook] & \mathbb{F}^j_m \otimes \mathcal{M}_{\mathcal{A}_m}(\Psi_m X) \arrow[u, hook]
\end{tikzcd}$$

Using Lemmas 2.15, 2.16, and [23, Lemma 3.13], we know that all the vertical maps on the left are quasi-isomorphisms. Similarly, using Lemmas 2.9, 2.10, and [23, Lemma 3.13], we know that all the vertical maps on the right are quasi-isomorphisms. Moreover, the second part of Lemma 2.21 states that the bottom horizontal map is a quasi-isomorphism. Thus, the chain map

$$\widetilde{\Psi}\colon \mathcal{H}(X^0_\bullet(E), X^0_\bullet)^{*,j} \to \mathbb{F}^j_m \otimes \mathcal{A}_m[W^{-1}_{\mathcal{A}_m}](X^0(E), X^0)$$

is a quasi-isomorphism. □

2.5.3 End of the proof We end the section with the proof of Theorem 2.5. Now that we have proved Lemma 2.22 which takes care of the *positive* Adams degrees, we have to treat the zero Adams degree part (recall that $\mathcal{H}$ is concentrated in nonnegative Adams degree because $\mathcal{A}$ is assumed to be weakly directed).

Let $\mathcal{I}$ be the (nonfull) A_∞-subcategory of $\mathcal{H}$ with

$$\operatorname{ob}(\mathcal{I}) = \{X^0_\bullet(E) \mid E \in \mathcal{E}\} \quad \text{and} \quad \mathcal{I}(X,Y) = \mathcal{G}(X,Y) \oplus \Bigl(\bigoplus_{j\geq 1} \mathcal{H}(X,Y)^{*,j}\Bigr)$$

(recall that if V is an Adams-graded vector space, we denote by $V^{*,j}$ its component of Adams degree j).

Lemma 2.23 *The inclusion $\mathcal{I} \hookrightarrow \mathcal{H}$ is a quasi-equivalence.*

Proof Observe that the inclusion $\mathcal{I} \hookrightarrow \mathcal{H}$ is cohomologically essentially surjective because every object of $\mathcal{H}$ can be related to one of $\mathcal{I}$ by a zigzag of morphisms in $W_\mathcal{G}$, which are quasi-isomorphisms in $\mathcal{H}$ (see [23, Lemma 3.12]). Therefore, it suffices to show that the inclusion

$$\mathcal{G}(X^0_\bullet(E), X^0_\bullet(E_\diamond)) \hookrightarrow \mathcal{H}(X^0_\bullet(E), X^0_\bullet(E_\diamond))^{*,0}$$

is a quasi-isomorphism for every $E, E_\diamond \in \mathcal{E}$.

Let $E_\diamond$ be an element of $\mathcal{E}$. When we write an object $X^n_\bullet$ without specifying the element of $\mathcal{E}$, we mean $X^n_\bullet(E_\diamond)$. Recall that we introduced a pair $(\mathcal{M}_\mathcal{G}, t_\mathcal{G})$ in Definition 2.14. According to Lemmas 2.15, 2.16 and [23, Lemma 3.13], the inclusion $\mathcal{M}_\mathcal{G}(X^0_\bullet(E)) \hookrightarrow {}_{W_\mathcal{G}^{-1}}\mathcal{M}_\mathcal{G}(X^0_\bullet(E))$ and the map ${}_{W_\mathcal{G}^{-1}}t_\mathcal{G}\colon \mathcal{H}(X^0_\bullet(E), X^0_\bullet) \to {}_{W_\mathcal{G}^{-1}}\mathcal{M}_\mathcal{G}(X^0_\bullet(E))$ are quasi-isomorphisms for every $E \in \mathcal{E}$. Also, observe that

$$\mathcal{M}_\mathcal{G}(X^0_\bullet(E))^{*,0} = \mathcal{G}(X^0_\bullet(E), X^0_\bullet).$$

The result then follows from the commutativity of the diagram

$$\begin{tikzcd}
\mathcal{G}(X^0_\bullet(E), X^0_\bullet) \arrow[r, equal] \arrow[d, equal] & \mathcal{G}(X^0_\bullet(E), X^0_\bullet) \arrow[r, equal] \arrow[d, hook] & \mathcal{G}(X^0_\bullet(E), X^0_\bullet) \arrow[d, hook] \\
\mathcal{M}_\mathcal{G}(X^0_\bullet(E))^{*,0} \arrow[r, hook, "\sim"] & {}_{W_\mathcal{G}^{-1}}\mathcal{M}_\mathcal{G}(X^0_\bullet(E))^{*,0} & \mathcal{H}(X^0_\bullet(E), X^0_\bullet)^{*,0} \arrow[l, "\sim", "{}_{W_\mathcal{G}^{-1}}t_\mathcal{G}"']
\end{tikzcd}$$

□

The following diagram of Adams-graded A_∞-categories is commutative:

$$\begin{tikzcd} \mathcal{H} \arrow[r, "\tilde{\Psi}"] & \mathbb{F}[t_m] \otimes \mathcal{A}_m[W_{\mathcal{A}_m}^{-1}] \\ \mathcal{I} \arrow[u, hook, "\sim"] \arrow[r, "\tilde{\Psi}"] & \mathcal{A}_m^0 \oplus (t\mathbb{F}[t_m] \otimes \mathcal{A}_m[W_{\mathcal{A}_m}^{-1}]^0) \arrow[u, hook] \end{tikzcd}$$

(recall that if $\mathcal{C}$ is an A_∞-category equipped with a splitting $\mathrm{ob}(\mathcal{C}) \simeq \mathbb{Z} \times \mathcal{E}$, then we denote by $\mathcal{C}^0$ the full A_∞-subcategory of $\mathcal{C}$ whose set of objects corresponds to $\{0\} \times \mathcal{E}$). Moreover, since $\mathcal{A}$ is assumed to be weakly directed with respect to the $\mathbb{Z}$-splitting of $\mathrm{ob}(\mathcal{A})$, Lemma 2.22 implies that the bottom horizontal A_∞-functor is a quasi-equivalence. Therefore we have

$$\mathcal{H} \simeq \mathcal{A}_m^0 \oplus (t\mathbb{F}[t_m] \otimes \mathcal{A}_m[W_{\mathcal{A}_m}^{-1}]^0).$$

Recall that $W_{\mathcal{A}_m} = f(\{\text{units}\}) \cup \{\text{units}\}$, so that

$$\mathcal{A}_m[W_{\mathcal{A}_m}^{-1}] \simeq \mathcal{A}_m[f(\{\text{units}\})^{-1}].$$

This concludes the proof of Theorem 2.5, since $\mathcal{H}$ is quasi-equivalent to the mapping torus of τ (see Lemma 2.13).

3 Chekanov–Eliashberg DG-category

Recall the following terminology.

Definition 3.1 A contact form is said to be *hypertight* if its Reeb vector field has no contractible periodic orbits.

In this section, we recall the definition of the Chekanov–Eliashberg DG-category associated to a family of Legendrians in a contact manifold equipped with a hypertight contact form α. We also describe the behavior of the Chekanov–Eliashberg DG-category under change of data.

In the following, (V, ξ) is a contact manifold of dimension $2n+1$. In order to have well defined gradings in $\mathbb{Z}$, we assume that $H_1(V)$ is free and that the first Chern class of ξ (equipped with any compatible almost complex structure) is 2-torsion. We will need the following definition.

Definition 3.2 We say that a Legendrian submanifold Λ in (V, ξ) is *chord generic* with respect to a contact form α if

(1) for every Reeb chord $c\colon [0, T] \to V$ of Λ, the space $D\varphi_{R_\alpha}^T(T_{c(0)}\Lambda)$ is transverse to $T_{c(T)}\Lambda$ in ξ,

(2) different Reeb chords belong to different Reeb trajectories.

3.1 Conley–Zehnder index

Let α be a hypertight contact form on (V, ξ) and let Λ be a chord generic Legendrian submanifold of (V, α). In the following, we define the Conley–Zehnder index of a Reeb chord of Λ starting and ending on the same connected component (such chords are called *pure*).

We briefly recall what is the Maslov index of a loop in the Grassmannian of Lagrangian subspaces in $\mathbb{C}^n$. We refer to [33] for a precise exposition. Fix a Lagrangian subspace K, and denote by $\Sigma_k(K)$ the set of Lagrangian subspaces in $\mathbb{C}^n$ whose intersection with K is k dimensional. Consider the *Maslov cycle*

$$\Sigma = \Sigma_1(K) \cup \cdots \cup \Sigma_n(K).$$

This is an algebraic variety of codimension one in the Lagrangian Grassmannian. Now if Γ is a loop in the Lagrangian Grassmannian, its Maslov index $\mu(\Gamma) \in \mathbb{Z}$ is the intersection number of Γ with Σ. The contribution of an intersection instant t_0 is computed as follows. Choose a Lagrangian complement W of K in $\mathbb{C}^n$. Then for each v in $\Gamma(t_0) \cap K$, there exists a vector $w(t)$ in W such that $v + w(t)$ is in $\Gamma(t)$ for every t near t_0. Consider the quadratic form

$$Q(v) = \frac{d}{dt}\omega(v, w(t))\Big|_{t=t_0}$$

on $\Gamma(t_0) \cap K$. Without loss of generality, Q can be assumed to be nonsingular and the contribution of t_0 to $\mu(\Gamma)$ is the signature of Q.

Recall that $H_1(V)$ is assumed to be free. We choose a family $(h_1, \dots, h_r)$ of embedded circles in V which represent a basis of $H_1(V)$, and a symplectic trivialization of ξ over each h_i. If γ is some loop in Λ, there is a unique family $(a_1, \dots, a_r)$ of integers such that $\left[\gamma_c - \sum_i a_i h_i\right]$ is zero in $H_1(V)$. Choose a surface Σ_γ in V such that

$$\partial \Sigma_\gamma = \gamma - \sum_i a_i h_i.$$

There is a unique trivialization of ξ over Σ_γ which extends the chosen trivializations over h_i. Thus we get a trivialization $\gamma^{-1}\xi \simeq S^1 \times \mathbb{C}^n$ (where n is the dimension of Λ). We denote by Γ the loop of Lagrangian planes in $\mathbb{C}^n$ corresponding, via the latter trivialization, to the loop $t \mapsto T_{\gamma(t)}\Lambda$. The Maslov index of Γ does not depend on the choice of the surface Σ_γ because we assumed $2c_1(\xi) = 0$. This construction defines a morphism $H_1(\Lambda, \mathbb{Z}) \to \mathbb{Z}$, and the *Maslov number* $m(\Lambda)$ of Λ is the generator of its image. In the following, we assume that the Maslov number of Λ is zero.

Now, let c be a pure Reeb chord of Λ (a Reeb chord is called pure if it starts and ends on the same connected component of the Legendrian). We choose a path $\gamma_c : [0, 1] \to \Lambda$ which starts at the endpoint of c, and ends at its starting point (γ_c is called a *capping path* of c). We denote by $\bar{\gamma}_c$ the loop obtained by concatenating γ and c. Let $(a_1, \dots, a_r)$ be the unique family of integers such that $\left[\bar{\gamma}_c - \sum_i a_i h_i\right]$ is zero in $H_1(V)$, and choose a surface Σ_c in V such that

$$\partial \Sigma_c = \bar{\gamma}_c - \sum_i a_i h_i.$$

There is a unique trivialization of ξ over Σ_c which extends the chosen trivializations over h_i. Thus we get a trivialization $\overline{\gamma}_c^{-1}\xi \simeq S^1 \times \mathbb{C}^n$ (where n is the dimension of Λ). We denote by Γ_c the path of Lagrangian planes in $\mathbb{C}^n$ corresponding, via the latter trivialization, to the concatenation of $t \mapsto T_{\gamma(t)}\Lambda$ and $t \mapsto D\varphi^t_{R_\alpha}(T_{c(0)}\Lambda)$. Since Λ is chord generic, Γ_c is not a loop: we close it in the following way. Let I be a complex structure on $\mathbb{C}^n$ which is compatible with the standard symplectic form on $\mathbb{C}^n$ and such that $I(\Gamma_c(1)) = \Gamma_c(0)$. Then we let $\overline{\Gamma}_c$ be the loop of Lagrangian subspaces obtained by concatenating Γ_c and the path $t \in \left[0, \frac{\pi}{2}\right] \mapsto e^{tI}\Gamma_c(1)$. The Conley–Zehnder index of c is the Maslov index of $\overline{\Gamma}_c$:

$$\mathrm{CZ}(c) := \mu(\overline{\Gamma}_c).$$

The Conley–Zehnder index of a Reeb chord does not depend on the choice of Σ_c because the first Chern class of ξ is 2-torsion, and it does not depend on the choice of γ_c because the Maslov number of Λ vanishes.

Remark In the case where $c(V,\xi)$ (where $c(V,\xi)$ is the positive generator of $\langle 2c_1(\xi), H_1(V)\rangle$) or $m(\Lambda)$ is nonzero, the Conley–Zehnder index is well defined in $\mathbb{Z}/d\mathbb{Z}$, where

$$d = \gcd(c(V,\xi), m(\Lambda)).$$

3.2 Moduli spaces

Recall that (V,ξ) is a contact manifold such that $H_1(V)$ is free and the first Chern class of ξ (equipped with any compatible almost complex structure) is 2-torsion. Let α be a hypertight contact form on (V,ξ) and let Λ be a chord generic Legendrian submanifold of (V,α) with vanishing Maslov number. In the following, we introduce the moduli spaces needed to define the Chekanov–Eliashberg category of Λ.

Definition 3.3 A Riemann $(d+1)$-pointed disk is a triple $(D, \boldsymbol{\zeta}, j)$ such that

(1) D is a smooth oriented manifold-with-boundary diffeomorphic to the closed unit disk in $\mathbb{C}$,

(2) $\boldsymbol{\zeta} = (\zeta_d, \dots, \zeta_1, \zeta_0)$ is a cyclically ordered family of distinct points on ∂D,

(3) j is an integrable almost complex structure on D which induces the given orientation on D.

If $(D, \boldsymbol{\zeta}, j)$ is a Riemann pointed disk, we denote by $\Delta := D \setminus \{\zeta_d, \dots, \zeta_1, \zeta_0\}$ the corresponding punctured disk.

Definition 3.4 A family of Riemann $(d+1)$-pointed discs is a bundle $\mathcal{S} \to \mathcal{R}$ with

(1) a family $\boldsymbol{\zeta} = (\zeta_d, \dots, \zeta_1, \zeta_0)$ of nonintersecting sections $\zeta_k : \mathcal{R} \to \mathcal{S}$ and

(2) a section $j : \mathcal{R} \to \mathrm{End}(T\mathcal{S})$

such that $(\mathcal{S}_r, \boldsymbol{\zeta}(r), j(r))$ is a Riemann $(d+1)$-pointed disk for every $r \in \mathcal{R}$.

Definition 3.5 Let $\mathcal{S} \to \mathcal{R}$ be a family of Riemann $(d+1)$-pointed discs. A choice of strip-like ends for $\mathcal{S} \to \mathcal{R}$ is a family of sections

$$\epsilon_d, \dots, \epsilon_1 \colon \mathcal{R} \times \mathbb{R}_{\le 0} \times [0,1] \to \Delta_r, \quad \epsilon_0 \colon \mathcal{R} \times \mathbb{R}_{\ge 0} \times [0,1] \to \Delta_r,$$

such that

(1) $\epsilon_d(r), \dots, \epsilon_1(r), \epsilon_0(r)$ are proper embeddings with

$$\epsilon_k(r)(\mathbb{R}_{\le 0} \times \{0,1\}) \subset \partial\Delta_r \quad \text{and} \quad \epsilon_0(r)(\mathbb{R}_{\ge 0} \times \{0,1\}) \subset \partial\Delta_r,$$

(2) $\epsilon_d(r), \dots, \epsilon_1(r), \epsilon_0(r)$ satisfy the asymptotic conditions

$$\epsilon_k(r)(s,t) \xrightarrow[s \to -\infty]{} \zeta_k(r) \quad \text{and} \quad \epsilon_0(r)(s,t) \xrightarrow[s \to +\infty]{} \zeta_0(r),$$

(3) $\epsilon_d(r), \dots, \epsilon_1(r), \epsilon_0(r)$ are $(i, j(r))$-holomorphic, where i is the standard complex structure on $\mathbb{C}$.

As explained in [36, Section (9c)], there is a universal family $\mathcal{S}^{d+1} \to \mathcal{R}^{d+1}$ of Riemann $(d+1)$-pointed discs when $d \ge 2$, which means that any other family $\mathcal{S} \to \mathcal{R}$ is isomorphic to the pullback of $\mathcal{S}^{d+1} \to \mathcal{R}^{d+1}$ by a map $\mathcal{R} \to \mathcal{R}^{d+1}$. In the following, we fix a choice of strip-like ends for the universal family $\mathcal{S}^{d+1} \to \mathcal{R}^{d+1}$.

Definition 3.6 Let J be an almost complex structure on ξ compatible with $(d\alpha)|_\xi$. We denote by J^α the unique almost complex structure on $\mathbb{R}_\sigma \times V$ which sends ∂_σ to R_α and which restricts to J on ξ. Let $c_d, \dots, c_1, c_0$ be Reeb chords of Λ, where $c_k \colon [0, T_k] \to V$.

(1) If $d = 1$, we denote by $\widetilde{\mathcal{M}}_{c_1,c_0}(\mathbb{R}\times\Lambda, J, \alpha)$ the set of equivalence classes of maps $u \colon \mathbb{R}\times[0,1] \to \mathbb{R}\times V$ such that

- u maps the boundary of $\mathbb{R} \times [0,1]$ to $\mathbb{R} \times \Lambda$,
- u satisfies the asymptotic conditions

$$u(s,t) \xrightarrow[s \to -\infty]{} (-\infty, c_1(T_1 t)) \text{ and } u(s,t) \xrightarrow[s \to +\infty]{} (+\infty, c_0(T_0 t)),$$

- u is (i, J^α)-holomorphic,

where two maps u and u' are identified if there exists $s_0 \in \mathbb{R}$ such that $u'(\cdot,\cdot) = u(\cdot + s_0, \cdot)$.

(2) If $d \ge 2$, we denote by $\widetilde{\mathcal{M}}_{c_d,\dots,c_1,c_0}(\mathbb{R} \times \Lambda, J, \alpha)$ the set of pairs (r, u) such that

- $r \in \mathcal{R}^{d+1}$ and $u \colon \Delta_r \to \mathbb{R} \times V$ maps the boundary of Δ_r to $\mathbb{R} \times \Lambda$,
- u satisfies the asymptotic conditions

$$(u \circ \epsilon_k(r))(s,t) \xrightarrow[s \to -\infty]{} (-\infty, c_k(T_k t)) \quad \text{and} \quad (u \circ \epsilon_0(r))(s,t) \xrightarrow[s \to +\infty]{} (+\infty, c_0(T_0 t)),$$

- u is (i, J^α)-holomorphic.

Observe that $\mathbb{R}$ acts on $\widetilde{\mathcal{M}}_{c_d,\dots,c_1,c_0}(\mathbb{R} \times \Lambda, J, \alpha)$ by translation in the $\mathbb{R}_\sigma$-coordinate. We set

$$\mathcal{M}_{c_d,\dots,c_1,c_0}(\mathbb{R} \times \Lambda, J, \alpha) := \widetilde{\mathcal{M}}_{c_d,\dots,c_1,c_0}(\mathbb{R} \times \Lambda, J, \alpha)/\mathbb{R}.$$

The moduli space $\widetilde{\mathcal{M}}_{c_d,\dots,c_1,c_0}(\mathbb{R}\times\Lambda, J, \alpha)$ can be realized as the zero-set of a section $\bar\partial\colon \mathcal{B}\to\mathcal{E}$ of a Banach bundle $\mathcal{E}\to\mathcal{B}$ (see for example [16]). We say that $\widetilde{\mathcal{M}}_{c_d,\dots,c_1,c_0}(\mathbb{R}\times\Lambda, J)$ is transversely cut out if $\bar\partial$ is transverse to the 0-section.

Definition 3.7 We say that J is *regular* (with respect to α and Λ) if the moduli spaces

$$\widetilde{\mathcal{M}}_{c_d,\dots,c_1,c_0}(\mathbb{R}\times\Lambda, J, \alpha)$$

are all transversely cut out.

Proposition 3.8 [9, Proposition 3.13] *The set of regular almost complex structures on ξ is Baire. Moreover, the dimension of a transversely cut out moduli space is*

$$\dim \widetilde{\mathcal{M}}_{c_d,\dots,c_1,c_0}(\mathbb{R}\times\Lambda, J, \alpha) = \mathrm{CZ}(a) - \left(\sum_{k=1}^{d} \mathrm{CZ}(b_k)\right) + d - 1.$$

3.3 Chekanov–Eliashberg DG-category

Recall that (V,ξ) is a contact manifold such that $H_1(V)$ is free and the first Chern class of ξ (equipped with any compatible almost complex structure) is 2-torsion. Let α be a hypertight contact form on (V,ξ) and let $\mathbf{\Lambda} = (\Lambda(E))_{E\in\mathcal{E}}$ be a family of Legendrian submanifolds of (V,ξ). We set $\Lambda := \bigcup_{E\in\mathcal{E}} \Lambda(E)$ and we assume that Λ is chord generic with vanishing Maslov number. Moreover, we denote by $\mathcal{C}(\Lambda(E),\Lambda(E'))$ the graded vector space generated by the words of Reeb chords $c_1\cdots c_d$, $d\geq 1$, where c_1 starts on $\Lambda(E)$, c_d ends on $\Lambda(E')$, and the ending component of c_i is the starting component of c_{i+1} for every $1\leq i\leq d-1$, with grading

$$|c_1\cdots c_d| := \sum_{i=1}^{d} (\mathrm{CZ}(c_i) - 1).$$

Finally, let J be a regular almost complex structure on ξ.

Definition 3.9 We denote by $\mathrm{CE}_*(\mathbf{\Lambda}) = \mathrm{CE}_*(\mathbf{\Lambda}, J, \alpha)$ the graded category defined as follows:

(1) The objects are the Legendrians $\Lambda(E)$, $E\in\mathcal{E}$.

(2) The space of morphisms from $\Lambda(E)$ to $\Lambda(E')$ is

$$\mathcal{C}(\Lambda(E),\Lambda(E')) \quad \text{if } E\neq E', \qquad \mathbb{F}\oplus\mathcal{C}(\Lambda(E),\Lambda(E')) \quad \text{if } E=E'$$

(the summand $\mathbb{F}$ corresponds to the "empty word").

(3) The composition is given by concatenation of words.

If c_0 is a Reeb chord in $\mathrm{CE}_*(\mathbf{\Lambda})$, we set

$$\partial(c_0) := \sum_{c_d,\dots,c_1} \#\mathcal{M}_{c_d,\dots,c_1,c_0}(\mathbb{R}\times\Lambda, J, \alpha)c_d\cdots c_1,$$

where $\#\mathcal{M} \in \mathbb{F}$ denotes the number of elements modulo 2 in $\mathcal{M}$ if $\mathcal{M}$ is finite, and 0 otherwise. Finally, we extend ∂ to $\mathrm{CE}_*(\mathbf{\Lambda})$ so that it is linear and satisfies the Leibniz rule with respect to the concatenation product.

Theorem 3.10 *$\partial\colon \mathrm{CE}_*(\mathbf{\Lambda}) \to \mathrm{CE}_*(\mathbf{\Lambda})$ decreases the grading by 1 and satisfies $\partial \circ \partial = 0$. As a result, $(\mathrm{CE}_{-*}(\mathbf{\Lambda}), \partial)$ is a DG-category.*

Proof This follows from Proposition 3.8, SFT compactness (see [1; 6], in particular [1, Theorem 3.20]) and pseudoholomorphic gluing. See [12; 14; 16] for details. □

Augmentations and Legendrian A_∞-(co)category Let $\mathbb{F}_\mathcal{E}$ be the category with $\mathcal{E}$ as set of objects, and morphism space from E to E' equal to $\mathbb{F}$ if $E = E'$, or 0 if $E \neq E'$. Assume that we have an augmentation of $\mathrm{CE}_{-*}(\mathbf{\Lambda})$, ie a DG-functor $\varepsilon\colon \mathrm{CE}_{-*}(\mathbf{\Lambda}) \to \mathbb{F}_\mathcal{E}$. Denote by ϕ_ε the automorphism of $\mathrm{CE}_{-*}(\mathbf{\Lambda})$ defined by

$$\phi_\varepsilon(c) = c + \varepsilon(c)$$

for every Reeb chord c of Λ. We denote by $\mathrm{CE}^\varepsilon_{-*}(\mathbf{\Lambda})$ the DG-category whose underlying graded category is the same as for $\mathrm{CE}_{-*}(\mathbf{\Lambda})$, but the differential is $\partial_\varepsilon = \phi_\varepsilon \circ \partial \circ \phi_\varepsilon^{-1}$. Now let $\overline{\mathrm{LC}^\varepsilon_*(\mathbf{\Lambda})}$ be the graded precategory (no composition) with

(1) objects the set of Legendrians $\{\Lambda(E) \mid E \in \mathcal{E}\}$,
(2) morphisms from $\Lambda(E)$ to $\Lambda(E')$ the vector space generated by (individual, not words of) Reeb chords c which start on $\Lambda(E)$ and end on $\Lambda(E')$, with grading

$$|c| := -\mathrm{CZ}(c).$$

Observe that, as a graded precategory, we have

$$\mathrm{CE}^\varepsilon_{-*}(\mathbf{\Lambda}) = \mathbb{F}_\mathcal{E} \oplus \Big(\bigoplus_{d\geq 1} \overline{\mathrm{LC}^\varepsilon_*(\mathbf{\Lambda})}[-1]^{\otimes d}\Big).$$

If we write

$$(\partial_\varepsilon)|_{\overline{\mathrm{LC}^\varepsilon_*(\mathbf{\Lambda})}} = \sum_{d\geq 0} \partial^d_\varepsilon \quad \text{with } \partial^d_\varepsilon\colon \overline{\mathrm{LC}^\varepsilon_*(\mathbf{\Lambda})} \to \overline{\mathrm{LC}^\varepsilon_*(\mathbf{\Lambda})}^{\otimes d},$$

then $\partial^0_\varepsilon = \varepsilon \circ \partial = 0$. Moreover, the operations $(\partial^d_\varepsilon)_{d\geq 1}$ make $\overline{\mathrm{LC}^\varepsilon_*(\mathbf{\Lambda})}$ a (noncounital) A_∞-cocategory (see Definition 1.2). We define the coaugmented A_∞-cocategory of $(\mathbf{\Lambda}, \varepsilon)$ to be

$$\mathrm{LC}^\varepsilon_*(\mathbf{\Lambda}) := \mathbb{F}_\mathcal{E} \oplus \overline{\mathrm{LC}^\varepsilon_*(\mathbf{\Lambda})}$$

(the A_∞-cooperations are naturally extended so that $1 \in \mathbb{F}_\mathcal{E}(E, E)$, $E \in \mathcal{E}$ are counits). Now observe that, as a DG-category,

$$\mathrm{CE}^\varepsilon_{-*}(\mathbf{\Lambda}) = \Omega(\mathrm{LC}^\varepsilon_*(\mathbf{\Lambda}))$$

(see [17, Section 2.2] for the cobar construction). Finally, we define the augmented A_∞-category of $(\mathbf{\Lambda}, \varepsilon)$ to be the graded dual (see [17, Section 2.1.3]) of $\mathrm{LC}^\varepsilon_*(\mathbf{\Lambda})$:

$$\mathrm{LA}^*_\varepsilon(\mathbf{\Lambda}) = \mathrm{LC}^\varepsilon_*(\mathbf{\Lambda})^\#.$$

3.4 Functoriality

Recall that (V,ξ) is a contact manifold such that $H_1(V)$ is free and the first Chern class of ξ (equipped with any compatible almost complex structure) is 2-torsion. Let $\boldsymbol{M} = (M(E))_{E\in\mathcal{E}}$ be a family of n-dimensional manifolds. When we write a map $\boldsymbol{\Lambda}\colon \boldsymbol{M} \to V$, we mean that $\boldsymbol{\Lambda}$ is a family of maps $\Lambda(E)\colon M(E) \to V$ indexed by $\mathcal{E}$, and we set

$$\Lambda = \bigsqcup_{E\in\mathcal{E}} \Lambda(E)\colon \bigsqcup_{E\in\mathcal{E}} M(E) \to V.$$

Definition 3.11 Let α be a hypertight contact form on (V,ξ). We denote by $\mathcal{L}_{\boldsymbol{M}}(\alpha)$ the bicategory where:

(1) Objects are the pairs $(\boldsymbol{\Lambda}, J)$, where $\boldsymbol{\Lambda}\colon \boldsymbol{M} \to V$ is a family of Legendrian embedding such that Λ is chord generic with vanishing Maslov number, and J is a regular almost complex structure on ξ.

(2) Morphisms from $(\boldsymbol{\Lambda}_0, J_0)$ to $(\boldsymbol{\Lambda}_1, J_1)$ are the smooth paths $\Phi = (\boldsymbol{\Lambda}_t, J_t)_{0\le t\le 1}$ going from $(\boldsymbol{\Lambda}_0, J_0)$ to $(\boldsymbol{\Lambda}_1, J_1)$, where $\boldsymbol{\Lambda}_t\colon \boldsymbol{M} \to V$ is a family of Legendrian embeddings and J_t is an almost complex structure on ξ.

(3) Homotopies from a morphism $\Phi = (\boldsymbol{\Lambda}_t, J_t)_{0\le t\le 1}\colon (\boldsymbol{\Lambda}_0, J_0) \to (\boldsymbol{\Lambda}_1, J_1)$ to another morphism $\Phi' = (\boldsymbol{\Lambda}'_t, J'_t)_{0\le t\le 1}\colon (\boldsymbol{\Lambda}_0, J_0) \to (\boldsymbol{\Lambda}_1, J_1)$ are the smooth families $(\boldsymbol{\Lambda}_{s,t}, J_{s,t})_{0\le s\le S,\,0\le t\le 1}$, where $\boldsymbol{\Lambda}_{s,t}\colon \boldsymbol{M} \to V$ is a family of Legendrian embeddings, $J_{s,t}$ is an almost complex structure on ξ, and

$$(\boldsymbol{\Lambda}_{s,0}, J_{s,0}) = (\boldsymbol{\Lambda}_0, J_0), \quad (\boldsymbol{\Lambda}_{s,1}, J_{s,1}) = (\boldsymbol{\Lambda}_1, J_1), \quad (\boldsymbol{\Lambda}_{0,t}, J_{0,t}) = (\boldsymbol{\Lambda}_t, J_t), \quad (\boldsymbol{\Lambda}_{S,t}, J_{S,t}) = (\boldsymbol{\Lambda}'_t, J'_t).$$

Definition 3.12 Let α, α' be hypertight contact forms on (V,ξ), and let φ be a contactomorphism of (V,ξ) such that $\varphi^*\alpha = \alpha'$. If $\Phi = (\boldsymbol{\Lambda}_t, J_t)_{0\le t\le 1}$ is a morphism in $\mathcal{L}_{\boldsymbol{M}}(\alpha)$, we denote by

$$\varphi^*\Phi = (\varphi^{-1}(\boldsymbol{\Lambda}_t), \varphi^* J_t)_{0\le t\le 1}$$

the corresponding morphism in $\mathcal{L}_{\boldsymbol{M}}(\alpha')$, and by

$$f^{\varphi}_{(\boldsymbol{\Lambda}_t, J_t)}\colon \mathrm{CE}_{-*}(\boldsymbol{\Lambda}_t, J_t, \alpha) \to \mathrm{CE}_{-*}(\varphi^{-1}(\boldsymbol{\Lambda}_t), \varphi^* J_t, \alpha')$$

the DG-functor which sends a Reeb chord c to $\varphi^{-1}(c)$.

Definition 3.13 Let α be a hypertight contact form on (V,ξ), and let $\Phi = (\boldsymbol{\Lambda}_t, J_t)_{0\le t\le 1}$ be a morphism in $\mathcal{L}_{\boldsymbol{M}}(\alpha)$. A *handle slide instant* in Φ is a time t_0 where $\boldsymbol{\Lambda}_{t_0}$ is chord generic and has Reeb chords $c_d, \dots, c_1, c_0$ such that the moduli space $\widetilde{\mathcal{M}}_{c_d,\dots,c_1,c_0}(\mathbb{R}\times\Lambda_{t_0}, J_{t_0}, \alpha)$ is not transversely cut out.

Theorem 3.14 *There exist functors $\mathcal{F}_\alpha$ from $\mathcal{L}_{\boldsymbol{M}}(\alpha)$ to the bicategory*[3] *of DG-categories such that*:

(1) *$\mathcal{F}_\alpha$ sends an object $(\boldsymbol{\Lambda}, J)$ to* $\mathrm{CE}_{-*}(\boldsymbol{\Lambda}, J, \alpha)$.

[3] Homotopies between DG-maps are DG-homotopies, see for example [31, Section 2.1].

(2) *$\mathcal{F}_\alpha$ sends a morphism to a homotopy equivalence.*

(3) *If φ is a contactomorphism of (V, ξ) such that $\varphi^*\alpha = \alpha'$ and if $\Phi = (\mathbf{\Lambda}_t, J_t)_{0\le t\le 1}$ is a morphism in $\mathcal{L}_{\boldsymbol{M}}(\alpha)$, then*

$$\mathcal{F}_{\alpha'}(\varphi^*\Phi) = f^{\varphi}_{(\mathbf{\Lambda}_1,J_1)} \circ \mathcal{F}_\alpha(\Phi) \circ (f^{\varphi}_{(\mathbf{\Lambda}_0,J_0)})^{-1}.$$

(4) *If $(\varphi_t)_{0\le t\le 1}$ is a contact isotopy of (V, ξ) satisfying $\varphi_t^*\alpha = \alpha'$ for every t, and if $(\mathbf{\Lambda}, J)$ is an object of $\mathcal{L}_{\boldsymbol{M}}(\alpha)$ such that there is neither birth/death of Reeb chords nor handle slide instants in the path $\Phi' = (\varphi_t^{-1}(\mathbf{\Lambda}), \varphi_t^* J)_t$, then*

$$\mathcal{F}_{\alpha'}(\Phi') = f^{\varphi_1}_{(\mathbf{\Lambda},J)} \circ (f^{\varphi_0}_{(\mathbf{\Lambda},J)})^{-1}.$$

Proof The existence of such functors at the category level (without homotopies) has been established in [14; 16] for the case $(V, \alpha) = (\mathbb{R} \times P, dz - \lambda)$. Statements in the general case can be found in [12, Section 4; 19, Section 5]. □

Note that I proved a weaker version of this result in my thesis by generalizing methods of [14; 16; 31]. The following is the only particular case of Theorem 3.14 that we will use in this paper.

Theorem 3.15 [32, Theorem 3.8] *Theorem 3.14 holds if we replace the categories $\mathcal{L}_{\boldsymbol{M}}(\alpha)$ by the subcategories $\mathcal{L}^0_{\boldsymbol{M}}(\alpha)$ where*

(1) *objects are the pairs $(\mathbf{\Lambda}, J)$ such that Λ has finitely many Reeb chords,*

(2) *morphisms from $(\mathbf{\Lambda}_0, J_0)$ to $(\mathbf{\Lambda}_1, J_1)$ are the families $\Phi = (\mathbf{\Lambda}_t, J_t)_{0\le t\le 1}$ such that Λ_t is chord generic and has finitely many Reeb chords for every t,*

(3) *homotopies from a morphism $\Phi = (\mathbf{\Lambda}_t, J_t)_{0\le t\le 1} \colon (\mathbf{\Lambda}_0, J_0) \to (\mathbf{\Lambda}_1, J_1)$ to another morphism $\Phi' = (\mathbf{\Lambda}'_t, J'_t)_{0\le t\le 1} \colon (\mathbf{\Lambda}_0, J_0) \to (\mathbf{\Lambda}_1, J_1)$ are the families $(\mathbf{\Lambda}_{s,t}, J_{s,t})_{0\le s\le S,\, 0\le t\le 1}$ such that $\Lambda_{s,t}$ is chord generic and has finitely many Reeb chords for every s, t.*

Remark We expect that the finiteness of Reeb chords condition in Theorem 3.15 (which is very restrictive) can be easily dropped using (homotopy) colimits of DG-categories diagrams. On the other hand, studying birth/death of Reeb chords phenomena is a more serious issue that we will address in future work.

4 Legendrian lifts of exact Lagrangians in the circular contactization

In this section, we start with a family $\boldsymbol{L}$ of mutually transverse compact connected exact Lagrangian submanifolds in a Liouville manifold, and we study a Legendrian lift $\mathbf{\Lambda}^\circ$ of $\boldsymbol{L}$ in the circular contactization. For the standard contact form, each point on a Legendrian gives rise to a (countable) infinite set of Reeb chords, and thus $\mathbf{\Lambda}^\circ$ is not chord generic. In Section 4.1, we explain how we perturb the contact form and we state our main result, which relates the Chekanov–Eliashberg DG-category of $\mathbf{\Lambda}^\circ$ and the Fukaya A_∞-category of $\boldsymbol{L}$.

4.1 Setting

Let (P,λ) be a Liouville manifold, and let

$$\boldsymbol{L} = (L(E))_{E\in\mathcal{E}}, \quad \mathcal{E} = \{1,\dots,N\},$$

be a family of mutually transverse compact connected exact Lagrangian submanifolds in (P,λ) such that there are primitives $f_E\colon L(E)\to\mathbb{R}$ of $\lambda|_{L(E)}$ satisfying $0 \le f_1 < \dots < f_N \le \frac{1}{2}$. We consider the contact manifold

$$(V^\circ,\xi^\circ) = (S^1\times P, \ker\alpha^\circ), \quad \text{where } S^1 = \mathbb{R}_\theta/\mathbb{Z},\ \alpha^\circ = d\theta - \lambda,$$

and the family of Legendrian submanifolds

$$\boldsymbol{\Lambda^\circ} := (\Lambda^\circ(E))_{E\in\mathcal{E}}, \quad \text{where } \Lambda^\circ(E) = \{(f_E(x),x)\in(\mathbb{R}/\mathbb{Z})\times P \mid x\in L(E)\}.$$

In order for the Chekanov–Eliashberg category of $\boldsymbol{\Lambda^\circ}$ and the Fukaya category of $\boldsymbol{L}$ to be $\mathbb{Z}$-graded, we assume that $H_1(P)$ is free, that the first Chern class of P (equipped with any almost complex structure compatible with $(-d\lambda)$) is 2-torsion, and that the Maslov classes of the Lagrangians $L(E)$ vanish.

4.1.1 Reeb chords Observe that $\Lambda^\circ = \bigcup_{E\in\mathcal{E}}\Lambda(E)$ is not chord generic for α° (see Definition 3.2). We will choose a compactly supported function $H\colon P\to\mathbb{R}$, and consider the perturbed contact form

$$\alpha_H^\circ = e^H\alpha^\circ.$$

The Reeb vector field of α_H° is then

$$R_{\alpha_H^\circ} = e^{-H}\begin{pmatrix}1+\lambda(X_H)\\ X_H\end{pmatrix},$$

where X_H is the unique vector field on P satisfying $\iota_{X_H}\,d\lambda = -dH$.

We fix a compact neighborhood K of L which is contained in a Weinstein neighborhood of L in P. It is not hard to see that for every positive integer N, the space of smooth functions H on P supported in K, such that the $R_{\alpha_H^\circ}$-chords of Λ° with action less than N are generic, is open and dense in $C_K^\infty(P)$. Therefore, the space of functions $H\in C_K^\infty(P)$ such that Λ° is chord generic with respect to α_H° is a Baire subset of $C_K^\infty(P)$. In particular, the latter is dense in $C_K^\infty(P)$. In the following, we choose $H\in C_K^\infty(P)$ such that

(1) Λ° is chord generic with respect to α_H°,

(2) H is sufficiently close to 0 so that

$$d\theta(R_{\alpha_H^\circ}) = e^{-H}(1+\lambda(X_H)) \ge \tfrac{1}{2}.$$

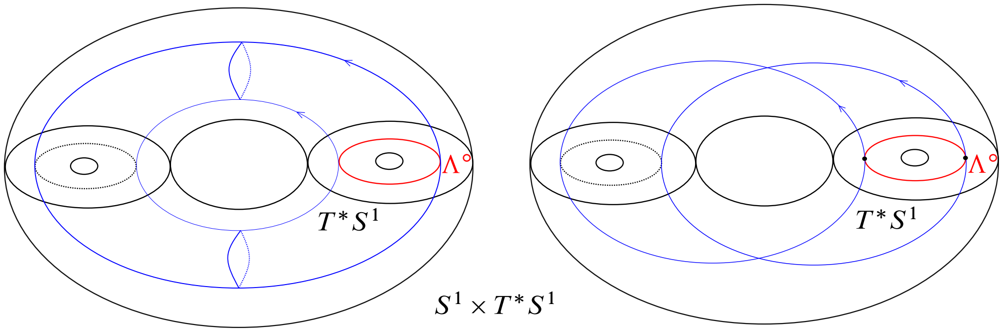


Figure 1: Reeb chords (in blue) of $\Lambda^\circ = \{0\} \times 0_{S^1}$. Left: for α°. Right: for α°_H.

Example 4.1 Assume that we are in the case

$$(P, \lambda) = (T^*M, p\,dq), \quad L = 0_M, \quad \text{and} \quad H(q, p) = h(q),$$

where $h\colon M \to \mathbb{R}$ is a Morse function (we present this example in order to see what happens, even if H is not compactly supported in T^*M). The Reeb vector field of α°_H is

$$R_{\alpha^\circ_H} = e^{-h}\begin{pmatrix} 1 \\ 0 \\ -dh \end{pmatrix},$$

and therefore the Reeb flow satisfies

$$\varphi^t_{R_{\alpha^\circ_H}}(\theta, (q, p)) = \big(\theta + te^{-h(q)}, (q, p - te^{-h(q)}\,dh(q))\big).$$

Thus, the $R_{\alpha^\circ_H}$-chords of Λ° are the paths $c\colon [0, T] \to S^1 \times T^*M$ of the form

$$c(t) = (te^{-h(q_0)}, (q_0, 0)), \quad \text{with } Te^{-h(q_0)} \in \mathbb{Z}_{\geq 1} \text{ and } q_0 \in \operatorname{Crit} h.$$

Observe that these Reeb chords are transverse but lie on top of each other. See Figure 1, where we illustrate this perturbation when $M = S^1$.

Conley–Zehnder index In order to define the Conley–Zehnder index (see Section 3.1), we need to choose a family $(h_0, h_1, \dots, h_s)$ of embedded circles in $V^\circ = S^1 \times P$ which represent a basis of $H_1(V^\circ)$, and a symplectic trivialization of ξ° over each h_i. We let $h_0 = S^1 \times \{a_0\}$ be some fiber of $S^1 \times P \to P$, and we fix $(h_1, \dots, h_s)$ to be any family of embedded circles in P which represent a basis of $H_1(P)$. We choose a symplectic isomorphism $\psi\colon (T_{a_0}P, -d\lambda_{a_0}) \xrightarrow{\sim} (\mathbb{C}^n, dx \wedge dy)$, and then we choose the symplectic trivialization

$$(\xi^\circ|_{h_0}, d\alpha^\circ) \xrightarrow{\sim} (h_0 \times \mathbb{C}^n, dx \wedge dy), \quad ((\theta, a_0), (\lambda_{a_0}(v), v)) \mapsto ((\theta, a_0), e^{2i\pi r\theta}\psi(v)),$$

where r is some integer, that we call *r-trivialization of ξ° over the fiber*. Finally, we choose some trivialization of ξ° over each h_i, $1 \leq i \leq s$.

Example 4.2 We compute the Conley–Zehnder index of a Reeb chord in the case of Example 4.1, ie when

$$(P,\lambda)=(T^*M,p\,dq),\quad L=0_M,\quad \text{and}\quad H(q,p)=h(q),$$

where $h\colon M\to\mathbb{R}$ is a Morse function. In this case, the Reeb flow is given by

$$\varphi^t_{R_{\alpha^\circ_H}}(\theta,(q,p))=\big(\theta+te^{-h(q)},(q,p-te^{-h(q)}\,dh(q))\big).$$

Let $c\colon[0,T]\to V^\circ$ be a Reeb chord of Λ°. Then there exists a positive integer k and a critical point q_0 of h such that

$$c(t)=(te^{-h(q_0)},(q_0,0))\quad\text{and}\quad Te^{-h(q_0)}=k.$$

Observe that $c(0)=c(T)$, and thus there is no need to choose a capping path for c. Besides, for every u in $T_{q_0}M$, we have

$$D\varphi^t_{R_{\alpha^\circ_H}}(c(0))(0,u,0)=(0,u,-te^{-h(q_0)}D^2h(q_0)u).$$

In order to compute the index of c, we first choose coordinates $(x_1,\dots,x_n)$ around $q_0\in M$ in which

$$h=h(q_0)+\frac{1}{2}\sum_{j=1}^{\dim(M)}\sigma_jx_j^2,\quad\text{where } \sigma_j=\pm1,$$

and we extend it to symplectic coordinates $(x_1,\dots,x_n,y_1,\dots,y_n)$ around $(q_0,0)\in T^*M$ by setting

$$y_j(q,p)=\Big\langle p,\frac{\partial}{\partial x_j}(q)\Big\rangle.$$

Our choice of trivialization for a fiber of $S^1\times P\to P$ induces the trivialization

$$e^{2i\pi rkt/T}(dx+i\,dy)\colon c^{-1}\xi^\circ\xrightarrow{\sim}(\mathbb{R}/T\mathbb{Z})\times\mathbb{C}^n$$

(observe that $\xi^\circ_{c(t)}=\{0\}\times T_{(q_0,0)}(T^*M)$). Accordingly, the path $t\mapsto D\varphi^t_{R_{\alpha^\circ_H}}(T_{c(0)}\Lambda^\circ)$ induces a path of Lagrangians

$$\Gamma_c\colon t\in[0,T]\mapsto\big\{\big(e^{2i\pi mkt/T}(u_j-ite^{-h(q_0)}\sigma_ju_j)\big)_{1\le j\le n}\mid u\in\mathbb{R}^n\big\}\subset\mathbb{C}^n.$$

We close this path using a counterclockwise rotation Γ, and call the resulting loop $\overline{\Gamma}_c$. In order to compute the Conley–Zehnder index of c, we have to look at how $\overline{\Gamma}_c$ intersects the Lagrangian $i\mathbb{R}^n$ (as explained in [14, Section 2.2]). Observe that Γ_c intersects $i\mathbb{R}^n$ positively $2rk$ times, so that Γ_c contributes $2rk$ to the Conley–Zehnder index of c. Moreover, since Γ is a counterclockwise rotation bringing

$$\big\{(u_j-iTe^{-h(q_0)}\sigma_ju_j)_{1\le j\le n}\mid u\in\mathbb{R}^n\big\}\quad\text{to }\mathbb{R}^n,$$

the contributions to the intersection between Γ and $i\mathbb{R}^n$ come from the negative eigenvalues σ_j. The computation done in [15, Lemma 3.4] implies that Γ contributes $\operatorname{ind}(q_0)$ to the Conley–Zehnder index of c. We conclude that the Conley–Zehnder index of c is

$$\mathrm{CZ}(c)=\mu(\overline{\Gamma}_c)=2rk+\operatorname{ind}(q_0).$$

4.1.2 Main result Let j be an almost complex structure on P compatible with $(-d\lambda)$, and let J° be its lift to a complex structure on ξ°. Recall from Section 3.3 the definition of the Chekanov–Eliashberg DG-category of a family of Legendrians. In our situation, $\mathrm{CE}^r_{-*}(\mathbf{\Lambda}^\circ) = \mathrm{CE}_{-*}(\mathbf{\Lambda}^\circ, J^\circ, \alpha^\circ_H)$ (with grading induced by the r-trivialization of ξ° over the fiber) is an Adams-graded DG-algebra, where the Adams degree of a Reeb chord c is the number of times c winds around the fiber. Besides, the map $\mathrm{CE}^r_{-*}(\mathbf{\Lambda}^\circ) \to \mathbb{F}$ which sends every Reeb chord to zero (and preserves units) defines an augmentation of $\mathrm{CE}^r_{-*}(\mathbf{\Lambda}^\circ)$.

Remark In the case of Example 4.1, the cohomological degree of a Reeb chord c in $\mathrm{CE}^r_{-*}(\mathbf{\Lambda}^\circ)$ corresponding to a positive integer k and a critical point q_0 is

$$1 - \mathrm{CZ}(c) = 1 - 2rk - \mathrm{ind}(q_0)$$

(see Example 4.2).

Besides, we denote by $\mathcal{F}uk(\boldsymbol{L})$ the Fukaya category with objects being the set of Lagrangians $\{L(E) \mid E \in \mathcal{E}\}$ (see for example [36, Chapter 2]), and by $\overrightarrow{\mathcal{F}uk}(\boldsymbol{L})$ its directed subcategory:

$$\hom_{\overrightarrow{\mathcal{F}uk}(\boldsymbol{L})}(L(E), L(E')) = \begin{cases} \langle L(E) \cap L(E') \rangle & \text{if } E < E', \\ \mathbb{F} & \text{if } E = E', \\ 0 & \text{if } E > E'; \end{cases}$$

see [36, Paragraph (5n)].

Let $\mathbb{F}[t_m]$ be the augmented Adams-graded associative algebra generated by a variable t_m of bidegree $(m, 1)$. Observe that if $\mathcal{C}$ is a subcategory of an A_∞-category $\mathcal{D}$ with $\mathrm{ob}(\mathcal{C}) = \mathrm{ob}(\mathcal{D})$, then $\mathcal{C} \oplus (t_m \mathbb{F}[t_m] \otimes \mathcal{D})$ is naturally an Adams-graded A_∞-category, where the Adams degree of $t_m^k \otimes x$ equals k. Moreover, we denote by $E(-) = B(-)^\#$ (graded dual of bar construction) the Koszul dual functor (see [17, Section 2.3] or [29, Section 2]). We say that Koszul duality holds for an augmented Adams-graded A_∞-category A if the natural map $A \to E(E(A))$ is a quasi-isomorphism (see [29, Theorem 2.4] or [17, Definition 17]).

Theorem 4.3 (Theorem B in the introduction) *Koszul duality holds for* $\mathrm{CE}^r_{-*}(\mathbf{\Lambda}^\circ)$, *and there is a quasi-equivalence of augmented Adams-graded* A_∞*-categories*

$$E(\mathrm{CE}^r_{-*}(\mathbf{\Lambda}^\circ)) \simeq \overrightarrow{\mathcal{F}uk}(\boldsymbol{L}) \oplus (t_{2r} \mathbb{F}[t_{2r}] \otimes \mathcal{F}uk(\boldsymbol{L})).$$

Corollary 4.4 *If L is a connected compact exact Lagrangian and Λ° is a Legendrian lift of L in the circular contactization, then there is a quasi-equivalence of augmented DG-algebras,*

$$\mathrm{CE}^1_{-*}(\Lambda^\circ) \simeq C_{-*}(\Omega(\mathbb{CP}^\infty \rtimes L)).$$

Proof Let x_0 be the basepoint of $\mathbb{CP}^\infty$, and set $P := \mathbb{CP}^\infty \setminus \{x_0\}$. Observe that

$$(P \times L)^* = P^* \wedge L^* = \mathbb{CP}^\infty \wedge L^* = \mathbb{CP}^\infty \rtimes L.$$

We have

$$\mathbb{F} \oplus (t_2 \mathbb{F}[t_2] \otimes CF^*(L)) \simeq \mathbb{F} \oplus (t_2 \mathbb{F}[t_2] \otimes C^*(L)) \simeq C^*((P \times L)^*) \simeq C^*(\mathbb{CP}^\infty \rtimes L).$$

Thus, it follows from Theorem 4.3 that

$$E(\mathrm{CE}^1_{-*}(\Lambda^\circ)) \simeq C^*(\mathbb{CP}^\infty \rtimes L).$$

Since Koszul duality holds for $\mathrm{CE}^1_{-*}(\Lambda^\circ)$,

$$\mathrm{CE}^1_{-*}(\Lambda^\circ) \simeq E(C^*(\mathbb{CP}^\infty \rtimes L)).$$

Observe that the graded algebra $H^*(\mathbb{CP}^\infty \rtimes L)$ is locally finite (ie each degree component is finitely generated) and simply connected (ie its augmentation ideal is concentrated in components of degree strictly greater than 1). Thus, according to the homological perturbation lemma (see [36, Proposition 1.12]), we can assume that $C^*(\mathbb{CP}^\infty \rtimes L)$ is a locally finite and simply connected A_∞ model for the DG-algebra of cochains on $\mathbb{CP}^\infty \rtimes L$. Therefore, [17, Lemma 10] implies that

$$\mathrm{CE}^1_{-*}(\Lambda^\circ) \simeq \Omega(C_{-*}(\mathbb{CP}^\infty \rtimes L)).$$

Now, since $\mathbb{CP}^\infty \rtimes L$ is simply connected, Adams' result (see [2; 3; 17]) yields

$$\Omega(C_{-*}(\mathbb{CP}^\infty \rtimes L)) \simeq C_{-*}(\Omega(\mathbb{CP}^\infty \rtimes L)). \qquad \square$$

4.1.3 Strategy of proof We explain the strategy to compute $E(\mathrm{CE}^r_{-*}(\mathbf{\Lambda}^\circ))$. Recall from the last paragraph of Section 3.3 that there is a coaugmented A_∞-cocategory $\mathrm{LC}_*(\mathbf{\Lambda}^\circ)$ such that

$$\mathrm{CE}^r_{-*}(\mathbf{\Lambda}^\circ) = \Omega(\mathrm{LC}_*(\mathbf{\Lambda}^\circ)).$$

$\mathrm{LC}_*(\mathbf{\Lambda}^\circ)$ inherits an Adams-grading from $\mathrm{CE}^r_{-*}(\mathbf{\Lambda}^\circ)$ (the same), and we denote by $\mathrm{LA}^*(\mathbf{\Lambda}^\circ)$ its graded dual (see [17, Section 2.1.3]). In our situation, $\mathrm{LA}^*(\mathbf{\Lambda}^\circ)$ is an augmented Adams-graded A_∞-category whose augmentation ideal is generated by the Reeb chords of $\mathbf{\Lambda}^\circ$ (and the Adams degree of a Reeb chord c is the number of times c winds around the fiber). Since there is a quasi-isomorphism $B(\Omega C) \simeq C$ for every A_∞-cocategory C (see [17, Section 2.2.2]), it follows that

$$E(\mathrm{CE}^r_{-*}(\mathbf{\Lambda}^\circ)) = B(\mathrm{CE}^r_{-*}(\mathbf{\Lambda}^\circ))^\# \simeq \mathrm{LC}_*(\mathbf{\Lambda}^\circ)^\# = \mathrm{LA}^*(\mathbf{\Lambda}^\circ)$$

(graded dual preserves quasi-isomorphisms).

Remark In the case of Example 4.1, the cohomological degree of a Reeb chord c in $\mathrm{LA}^*(\mathbf{\Lambda}^\circ)$ corresponding to a positive integer k and a critical point q_0 is

$$\mathrm{CZ}(c) = 2rk + \mathrm{ind}(q_0)$$

(see Example 4.2).

In order to compute $\mathrm{LA}^*(\mathbf{\Lambda}^\circ)$, we lift the problem to the contact manifold

$$(V, \xi) = (\mathbb{R}_\theta \times P, \ker(d\theta - \lambda)),$$

and introduce the following objects.

Definition 4.5 Let $\boldsymbol{M} = (M^n(E))_{(n,E)\in\mathbb{Z}\times\mathcal{E}}$ be a family of Legendrian submanifolds in (V,ξ), K an almost complex structure on ξ, and β a hypertight contact form on (V,ξ) for which $\boldsymbol{M}$ is chord-generic. We denote by $\mathcal{A}(\boldsymbol{M},K,\beta)$ the A_∞-category defined as follows:

(1) The objects of $\mathcal{A}(\boldsymbol{M},K,\beta)$ are the Legendrians $M^n(E)$, $(n,E)\in\mathbb{Z}\times\mathcal{E}$.

(2) The space of morphisms from $M^i(E)$ to $M^j(E')$ is either generated by the R_β-chords from $M^i(E)$ to $M^j(E')$ if $(i,E)<(j,E')$, or $\mathbb{F}$ if $(i,E)=(j,E')$, or 0 otherwise.

(3) The operations are such that $1\in\mathcal{A}(\boldsymbol{M},K,\beta)(M^n(E),M^n(E))$ is a strict unit, and for every sequence $(i_0,E_0)<\cdots<(i_d,E_d)$, for every sequence of Reeb chords

$$(c_1,\ldots,c_d)\in\mathcal{R}(M^{i_0}(E_0),M^{i_1}(E_1))\times\cdots\times\mathcal{R}(M^{i_{d-1}}(E_{d-1}),M^{i_d}(E_d)),$$

we have

$$\mu_{\mathcal{A}(\boldsymbol{M},K,\beta)}(c_1,\ldots,c_d)=\sum_{c_0\in\mathcal{R}(M^{i_0}(E_0),M^{i_d}(E_d))}\#\mathcal{M}_{c_d,\ldots,c_1,c_0}(\mathbb{R}\times M,K,\beta)c_0$$

(see Definition 3.6 for the moduli spaces).

Definition 4.6 Consider a path $(\boldsymbol{M}_t)_{0\le t\le 1}$, where $\boldsymbol{M}_t=(M_t^n(E))_{(n,E)\in\mathbb{Z}\times\mathcal{E}}$ is a family of Legendrian submanifolds in (V,ξ), such that $M_1^{n-1}(E)=M_0^n(E)=:M^n(E)$. Let K be an almost complex structure on ξ, and β a hypertight contact form on (V,ξ) for which $\boldsymbol{M}=(M^n(E))_{(n,E)\in\mathbb{Z}\times\mathcal{E}}$ is chord-generic. We denote by $\tau_{(\boldsymbol{M}_t)_t,K,\beta}\colon\mathcal{A}(\boldsymbol{M},K,\beta)\to\mathcal{A}(\boldsymbol{M},K,\beta)$ the A_∞-functor defined as follows:

(1) On objects, $\tau_{(\boldsymbol{M}_t)_t,K,\beta}$ sends $M^n(E)=M_0^n(E)$ to $M^{n+1}(E)=M_1^n(E)$.

(2) On morphisms, the map

$$\begin{aligned}\tau_{(\boldsymbol{M}_t)_t,(K_t)_t,\beta}\colon\mathcal{A}(\boldsymbol{M},K,\beta)(M^{i_0}(E_0),M^{i_1}(E_1))\otimes\cdots\otimes\mathcal{A}(\boldsymbol{M},K,\beta)(M^{i_{d-1}}(E_{d-1}),M^{i_d}(E_d))\\ \to\mathcal{A}(\boldsymbol{M},K,\beta)(M^{i_0+1}(E_0),M^{i_d+1}(E_d))\end{aligned}$$

is obtained by dualizing the components of the DG-isomorphism

$$\mathrm{CE}_{-*}((M_1^n)_{i_0\le n\le i_d},K_1,\beta)\to\mathrm{CE}_{-*}((M_0^n)_{i_0\le n\le i_d},K_0,\beta)$$

induced by the path $((M_{1-t}^n)_{i_0\le n\le i_d},K_{1-t})_{0\le t\le 1}$ (see Theorem 3.14).

Remark (1) The A_∞-functor $\tau_{(\boldsymbol{M}_t)_t,K,\beta}$ is a quasi-equivalence because it is defined by dualizing the components of a DG-isomorphism.

(2) The $\mathbb{Z}$-splitting

$$\mathbb{Z}\times\mathcal{E}\xrightarrow{\sim}\mathrm{ob}(\mathcal{A}(\boldsymbol{M},K,\beta)),\quad (n,E)\mapsto M^n(E),$$

is compatible with the quasi-autoequivalence $\tau_{(\boldsymbol{M}_t)_t,K,\beta}$ in the sense of Definition 2.2. As explained there, this turns $\mathcal{A}(\boldsymbol{M},K,\beta)$ into an Adams-graded A_∞-category: the Adams degree of a morphism c from $M^i(E)$ to $M^j(E')$ is $j-i$.

In Section 4.2, we lift the data used to define $\mathrm{LA}^*(\boldsymbol{\Lambda}^\circ)$ (Legendrian $\boldsymbol{\Lambda}^\circ$, almost complex structure J°, contact form α_H°) to $\mathbb{R}\times P$. This gives us a path $(\boldsymbol{\Lambda}_t)_t$, an almost complex structure J and a contact form α_H for which we can prove, using Theorem 2.4, that

$$\mathrm{LA}^*(\boldsymbol{\Lambda}^\circ) \simeq \mathrm{MT}(\tau_{(\boldsymbol{\Lambda}_t)_t,J,\alpha_H}).$$

In Section 4.3, we use a contactomorphism ϕ_H satisfying $\phi_H^*\alpha_H = (d\theta-\lambda) =: \alpha$ to change our data into $(\boldsymbol{\Lambda}_{H,t})_t$, J_H and α. We then prove that

$$\mathrm{MT}(\tau_{(\boldsymbol{\Lambda}_t)_t,J,\alpha_H}) \simeq \mathrm{MT}(\tau_{(\boldsymbol{\Lambda}_{H,t})_t,J_H,\alpha}).$$

In Section 4.4, we change the almost complex structure J_H to the original one J, and use Theorem 3.15 to prove that

$$\mathrm{MT}(\tau_{(\boldsymbol{\Lambda}_{H,t})_t,J_H,\alpha}) \simeq \mathrm{MT}(\tau_{(\boldsymbol{\Lambda}_{H,t})_t,J,\alpha}).$$

In Section 4.5 we project our data to P, so that we get a path $(\boldsymbol{L}_{H,t})_t$ of Lagrangians in P and the almost complex structure j. We use these new data to define an A_∞-category $\mathcal{O}$ and a quasi-autoequivalence $\gamma\colon \mathcal{O}\to\mathcal{O}$. Then we use [10, Theorem 2.1] to prove that

$$\mathrm{MT}(\tau_{(\boldsymbol{\Lambda}_{H,t})_t,J,\alpha}) \simeq \mathrm{MT}(\gamma).$$

Finally in Section 4.6, we use Theorem 2.5 (Theorem A in the introduction) to conclude.

4.2 Lift to $\mathbb{R}\times P$

In the following we consider the contact manifold

$$(V,\xi) = (\mathbb{R}_\theta\times P, \ker(\alpha)), \quad \text{where } \alpha = d\theta-\lambda,$$

and the family of Legendrian submanifolds

$$\boldsymbol{\Lambda} := (\Lambda^n(E))_{(n,E)\in\mathbb{Z}\times\mathcal{E}}, \quad \text{where } \Lambda^\theta(E) = \{(f_E(x)+\theta, x)\in\mathbb{R}\times P \mid x\in L(E)\}.$$

Recall from Section 4.1.1 that we chose a compactly supported function $H\colon P\to\mathbb{R}$ such that

(1) $\boldsymbol{\Lambda}^\circ$ is chord generic with respect to α_H°,

(2) H is sufficiently close to 0 so that

$$d\theta(R_{\alpha_H^\circ}) = e^{-H}(1+\lambda(X_H)) \ge \tfrac{1}{2}.$$

We consider the contact form

$$\alpha_H := e^H\alpha,$$

with Reeb vector field

$$R_{\alpha_H} = e^{-H}\begin{pmatrix} 1+\lambda(X_H) \\ X_H \end{pmatrix}.$$

Moreover, we denote by J the lift of J° to an almost complex structure on ξ.

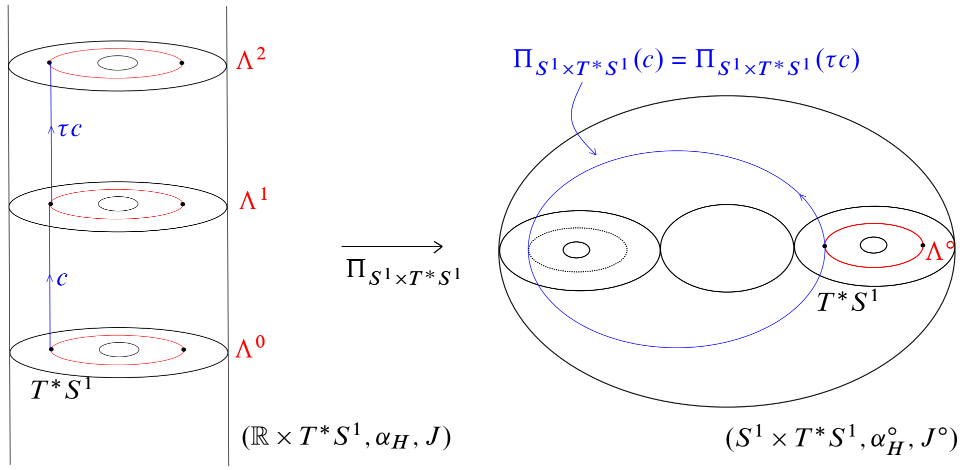


Figure 2: Action of the projection $\Pi_{S^1\times T^*S^1}$.

Definition 4.7 Consider the path of Legendrians $(\boldsymbol{\Lambda}_t)_{0\le t\le 1}$, where $\Lambda^n_t(E) = \Lambda^{n+t}(E)$. We set

$$\mathcal{A} := \mathcal{A}(\boldsymbol{\Lambda}, J, \alpha_H) \quad \text{and} \quad \tau := \tau_{(\boldsymbol{\Lambda}_t)_t, J, \alpha_H}$$

(see Definitions 4.5 and 4.6).

Relation between $\mathrm{LA}^*(\boldsymbol{\Lambda}^\circ)$ and $(\mathcal{A}, \tau)$ We now explain how $\mathrm{LA}^*(\boldsymbol{\Lambda}^\circ)$ and $(\mathcal{A}, \tau)$ are related. See Figure 2, where we illustrate the action of the projection $\Pi_{S^1\times P}$ in the case

$$(P, \lambda) = (T^*S^1, p\,dq), \quad L = 0_{S^1}, \quad \text{and} \quad H(q,p) = h(q),$$

where $h\colon S^1 \to \mathbb{R}$ is a Morse function.

Lemma 4.8 *The A_∞-functor τ is strict, and it sends a Reeb chord $t \mapsto (\theta(t), x(t))$ in $\mathcal{A}(\Lambda^i(E), \Lambda^j(E'))$ to the Reeb chord $t \mapsto (\theta(t)+1, x(t))$ in $\mathcal{A}(\Lambda^{i+1}(E), \Lambda^{j+1}(E'))$. In particular, τ acts bijectively on hom-sets.*

Proof Recall that $\alpha_H = e^H\alpha$, with H a function defined on the base manifold P. In particular, the flow $\varphi^t_{\partial_\theta}$ of ∂_θ is a strict contactomorphism of (V, α_H). Moreover, since J is the lift of an almost complex structure j on P, we have

$$((\Lambda^{n+1-t})_{i_0\le n\le i_d}, J) = \big(((\varphi^t_{\partial_\theta})^{-1}\Lambda^{n+1})_{i_0\le n\le i_d}, (\varphi^t_{\partial_\theta})^*J\big).$$

The result follows from Theorem 3.15. $\square$

We denote by $\mathcal{A}_\tau$ the Adams-graded A_∞-category associated to τ as in Definition 2.3.

Lemma 4.9 *There is a quasi-isomorphism of Adams-graded A_∞-categories*

$$\mathrm{LA}^*(\boldsymbol{\Lambda}^\circ) \simeq \mathcal{A}_\tau.$$

Proof Consider the map which sends a Reeb chord $c \in \mathcal{R}(\Lambda^i(E), \Lambda^j(E'))$ to the corresponding chord $\Pi_{S^1\times P}(c) \in \mathcal{R}(\Lambda^\circ(E), \Lambda^\circ(E'))$ (where $\Pi_{S^1\times P}\colon \mathbb{R}\times P \to S^1\times P$ is the projection). According to Lemma 4.8, $\Pi_{S^1\times P}(\tau c) = \Pi_{S^1\times P}(c)$, and thus the map $c \mapsto \Pi_{S^1\times P}(c)$ induces a map $\psi\colon \mathcal{A}_\tau \to \mathrm{LA}^*(\Lambda^\circ)$. Moreover, observe that ψ is a bijection on hom-spaces. It remains to prove that ψ is an A_∞-map. This follows from the fact that the map

$$u = (\sigma, v) \mapsto (\sigma, \Pi_{S^1\times P} \circ v)$$

induces a bijection

$$\mathcal{M}_{c_d,\dots,c_1,c_0}(\mathbb{R}\times\Lambda, J, \alpha_H) \xrightarrow{\sim} \mathcal{M}_{\psi(c_d),\dots,\psi(c_1),\psi(c_0)}(\mathbb{R}\times\Lambda^\circ, J^\circ, \alpha_H^\circ). \qquad \square$$

Lemma 4.10 *The Adams-graded A_∞-category $\mathrm{LA}^*(\boldsymbol{\Lambda}^\circ)$ is quasi-equivalent to the mapping torus of $\tau\colon \mathcal{A}\to\mathcal{A}$ (see Definition 2.1).*

Proof This follows directly from Theorem 2.4 using Lemmas 4.8 and 4.9. $\square$

4.3 Rectification of the contact form

Now that we are in the usual contactization, we have the following result.

Lemma 4.11 *There exists a contactomorphism ϕ_H of (V,ξ) such that*

$$\phi_H^*\alpha_H = \alpha.$$

Proof Recall that $\alpha_H = e^H\alpha$, with H a compactly supported function on the base manifold P such that $e^{-H}(1+\lambda(X_H)) \geq \frac{1}{2}$.

Assume that there is a contact isotopy $(\phi_t)_{0\leq t\leq 1}$ such that $\phi_0 = \mathrm{id}$ and

$$\phi_t^*\alpha_{tH} = \alpha \tag{1}$$

for every t. Let $(F_t)_t$ be the family of functions on V such that

$$\frac{d}{dt}\phi_t = Y_{F_t}\circ\phi_t,$$

where, for each fixed t, Y_{F_t} is the vector field on V satisfying

$$\alpha(Y_{F_t}) = F_t, \quad \iota_{Y_{F_t}}d\alpha = dF_t(R_\alpha)\alpha - dF_t.$$

Let us prove that Y_{F_t} satisfies

$$\alpha_{tH}(Y_{F_t}) = e^{tH}F_t, \quad \iota_{Y_{F_t}}d\alpha_{tH} = d(e^{tH}F_t)(R_{\alpha_{tH}})\alpha_{tH} - d(e^{tH}F_t).$$

Observe that the first equality is clear, and that it is enough to show that the second equality holds on $\xi = \ker(\alpha)$. Now for every $Z \in \xi$, we have

$$\begin{aligned}
\iota_{Y_{F_t}} d\alpha_{tH}(Z) &= (d(e^{tH}) \wedge \alpha)(Y_{F_t}, Z) + e^{tH} d\alpha(Y_{F_t}, Z) \\
&= -\alpha(Y_{F_t}) d(e^{tH})(Z) - e^{tH} dF_t(Z) && \text{(because } Z \in \xi\text{)} \\
&= -F_t d(e^{tH})(Z) - e^{tH} dF_t(Z) && \text{(because } \alpha(Y_{F_t}) = F_t\text{)} \\
&= -d(e^{tH} F_t)(Z).
\end{aligned}$$

Taking the derivative of (1) with respect to t, and using what we just proved, we get

$$H + d(e^{tH} F_t)(R_{\alpha_{tH}}) = 0. \tag{2}$$

Besides, we deduce from

$$R_{\alpha_{tH}} = e^{-tH} \begin{pmatrix} 1 + t\lambda(X_H) \\ tX_H \end{pmatrix}, \quad \iota_{X_H} d\lambda = -dH,$$

that

$$dH(R_{\alpha_{tH}}) = 0.$$

Then (2) gives

$$dF_t(R_{\alpha_{tH}}) = -He^{-tH}. \tag{3}$$

Conversely, if $(F_t)_t$ is a family of functions on V satisfying (3), then the contact isotopy $(\phi_t)_t$ defined by

$$\phi_0 = \mathrm{id} \quad \text{and} \quad \frac{d}{dt}\phi_t = Y_{F_t} \circ \phi_t$$

satisfies

$$\frac{d}{dt}(\phi_t^* \alpha_{tH}) = 0,$$

and thus $\phi_H := \phi_1$ gives the desired result.

Therefore, it remains to find a family $(F_t)_t$ satisfying (3). First recall that

$$R_{\alpha_{tH}} = e^{-tH} \begin{pmatrix} 1 + t\lambda(X_H) \\ tX_H \end{pmatrix}.$$

By assumption on H, the function $d\theta(R_{\alpha_{tH}})$ is greater than $\frac{1}{2}$ for every $t \in [0, 1]$. Thus, for every $t \in [0, 1]$ and every (θ, x) in V, there exists a unique real number $\rho_t(\theta, x)$ such that

$$\varphi_{R_{\alpha_{tH}}}^{-\rho_t(\theta,x)}(\theta, x) \in \{0\} \times P.$$

Then we let

$$F_t := -\rho_t He^{-tH}.$$

For every real number t, we have

$$F_t \circ \varphi_{R_{\alpha_{tH}}}^t = -(\rho_t \circ \varphi_{R_{\alpha_{tH}}}^t) He^{-tH} \quad \text{because } dH(R_{\alpha_{tH}}) = 0.$$

But the map $\varphi_{R_{\alpha_{tH}}}^{-\rho_t \circ \varphi_{R_{\alpha_{tH}}}^t + t}$ takes its values in $\{0\} \times P$ by definition of ρ_t, so by uniqueness we have

$$\rho_t \circ \varphi_{R_{\alpha_{tH}}}^t = \rho_t + t.$$

Then we have

$$F_t \circ \varphi^t_{R_{\alpha_{tH}}} = -(\rho_t + t) H e^{-tH},$$

and thus

$$dF_t(R_{\alpha_{tH}}) = -He^{-tH}. \qquad \square$$

Example 4.12 Assume that we are in the case

$$(P, \lambda) = (T^*M, p\,dq), \quad L = 0_M, \quad \text{and} \quad H(q,p) = h(q),$$

where $h\colon M \to \mathbb{R}$ is a Morse function. Then the diffeomorphism ϕ_H defined by

$$\phi_H^{-1}(\theta, (q,p)) = \big(\theta e^{h(q)}, (q, e^{h(q)} p + \theta e^{h(q)}\, dh(q))\big)$$

satisfies $\phi_H^* \alpha_H = \alpha$. With this choice of ϕ_H, we have in particular

$$\phi_H^{-1}(\{\theta\} \times 0_M) = j^1(\theta e^h) \subset \mathbb{R} \times T^*M.$$

In the following, we fix a contactomorphism ϕ_H as in Lemma 4.11. We define a pair $(\mathcal{A}_1, \tau_1)$, which is roughly obtained by pulling back the data of $(\mathcal{A}, \tau)$ by ϕ_H.

Definition 4.13 Let

$$\mathbf{\Lambda}_H := (\Lambda^n_H(E))_{(n,E)\in\mathbb{Z}\times\mathcal{E}}, \quad \text{where } \Lambda^\theta_H(E) := \phi_H^{-1}(\Lambda^\theta(E)), \text{ and } J_H := \phi_H^* J.$$

Consider the path of Legendrians $(\mathbf{\Lambda}_{H,t})_{0\le t\le 1}$, where $\Lambda^n_{H,t}(E) = \Lambda^{n+t}_H(E)$. We set

$$\mathcal{A}_1 := \mathcal{A}(\mathbf{\Lambda}_H, J_H, \alpha) \quad \text{and} \quad \tau_1 := \tau_{(\mathbf{\Lambda}_{H,t})_t, J_H, \alpha}$$

(see Definitions 4.5 and 4.6).

Relation between $(\mathcal{A}, \tau)$ and $(\mathcal{A}_1, \tau_1)$ We now explain how the pairs $(\mathcal{A}, \tau)$ and $(\mathcal{A}_1, \tau_1)$ (Definitions 4.7 and 4.13) are related. See Figure 3, where we illustrate the action of the contactomorphism ϕ_H^{-1} in the case

$$(P, \lambda) = (T^*S^1, p\,dq), \quad L = 0_{S^1}, \quad \text{and} \quad H(q,p) = h(q),$$

where $h\colon S^1 \to \mathbb{R}$ is a Morse function.

Lemma 4.14 *There is a strict A_∞-isomorphism $\zeta_1\colon \mathcal{A} \to \mathcal{A}_1$ defined as follows:*

1. *On objects, $\zeta_1(\Lambda^n(E)) = \Lambda^n_H(E)$.*
2. *On morphisms, ζ_1 sends a Reeb chord c in $\mathcal{A}(\Lambda^i(E), \Lambda^j(E'))$ to the Reeb chord*

   $$\zeta_1(c) = \phi_H^{-1} \circ c$$

   in $\mathcal{A}_1(\Lambda^i_H(E), \Lambda^j_H(E'))$.

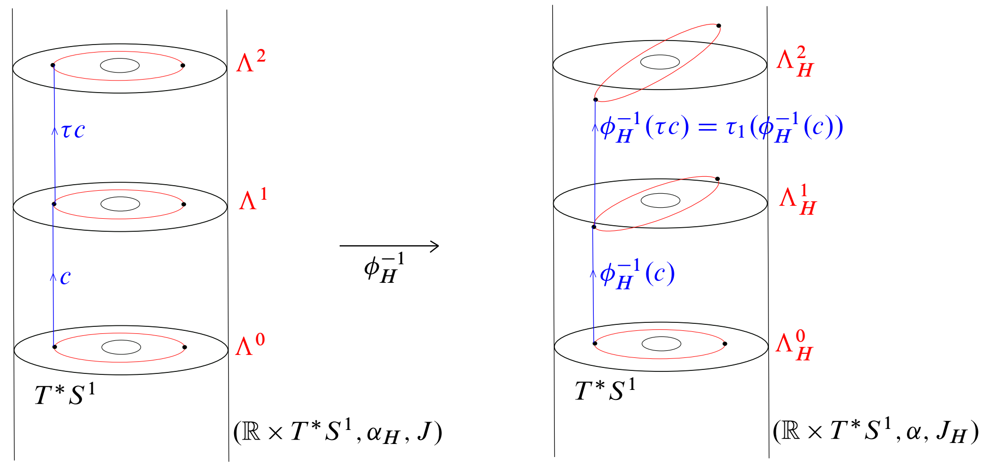


Figure 3: Action of the contactomorphism ϕ_H^{-1}.

Proof We have to show that ζ_1 is an A_∞-map. This follows from the fact that the map

$$u=(\sigma,v)\mapsto(\sigma,\phi_H^{-1}\circ v)$$

induces a bijection

$$\mathcal{M}_{c_d,\dots,c_1,c_0}(\mathbb{R}\times\Lambda,J,\alpha_H)\xrightarrow{\sim}\mathcal{M}_{\phi_H^{-1}(c_d)\dots\phi_H^{-1}(c_1),\phi_H^{-1}(c_0)}(\mathbb{R}\times\Lambda_H,J_H,\alpha).\qquad\square$$

Lemma 4.15
$$\tau_1=\zeta_1\circ\tau\circ\zeta_1^{-1}.$$

Proof This follows from Theorem 3.15 using that $\phi_H^*\alpha_H=\alpha$ and

$$((\Lambda_H^{n+1-t})_{i_0\le n\le i_d},J_H)=((\phi_H^{-1}\Lambda^{n+1-t})_{i_0\le n\le i_d},\phi_H^*J).\qquad\square$$

Lemma 4.16 *The mapping torus of $\tau\colon\mathcal{A}\to\mathcal{A}$ is quasi-equivalent to the mapping torus of $\tau_1\colon\mathcal{A}_1\to\mathcal{A}_1$ (see Definition 2.1).*

Proof According to Lemma 4.15 the following diagram of Adams-graded A_∞-categories is commutative:

$$\begin{tikzcd}
\mathcal{A} \arrow[d, "\zeta_1"] & \mathcal{A}\sqcup\mathcal{A} \arrow[l, "\mathrm{id}\sqcup\mathrm{id}"'] \arrow[r, "\mathrm{id}\sqcup\tau"] \arrow[d, "\zeta_1\sqcup\zeta_1"] & \mathcal{A} \arrow[d, "\zeta_1"] \\
\mathcal{A}_1 & \mathcal{A}_1\sqcup\mathcal{A}_1 \arrow[l, "\mathrm{id}\sqcup\mathrm{id}"'] \arrow[r, "\mathrm{id}\sqcup\tau_1"] & \mathcal{A}_1
\end{tikzcd}$$

Moreover, each vertical arrow is a quasi-equivalence according to Lemma 4.14. Thus the result follows from Proposition 1.22. $\square$

4.4 Back to the original almost complex structure

In this section, we introduce a pair $(\mathcal{A}_2,\tau_2)$ defined using the same data as $(\mathcal{A}_1,\tau_1)$ (Definition 4.13), except we are using the almost complex structure J instead of J_H.

Definition 4.17 We set

$$\mathcal{A}_2 := \mathcal{A}(\mathbf{\Lambda}_H, J, \alpha) \quad \text{and} \quad \tau_2 := \tau_{(\mathbf{\Lambda}_{H,t})_t, J, \alpha}$$

(see Definitions 4.5 and 4.6).

Relation between $(\mathcal{A}_1, \tau_1)$ and $(\mathcal{A}_2, \tau_2)$

Lemma 4.18 *Choose a generic path $(J_t^{12})_{0\le t\le 1}$ such that $J_0^{12} = J$ and $J_1^{12} = J_H$. There is an A_∞-isomorphism $\zeta_{12}\colon \mathcal{A}_1 \to \mathcal{A}_2$ defined as follows*:

(1) *On objects*, $\zeta_{12}(\Lambda_H^n(E)) = \Lambda_H^n(E)$.

(2) *On morphisms*, *the map*

$$\zeta_{12}\colon \mathcal{A}_1(\Lambda_H^{i_0}(E_0), \Lambda_H^{i_1}(E_1)) \otimes \cdots \otimes \mathcal{A}_1(\Lambda_H^{i_{d-1}}(E_{d-1}), \Lambda_H^{i_d}(E_d)) \to \mathcal{A}_2(\Lambda_H^{i_0}(E_0), \Lambda_H^{i_d}(E_d))$$

is obtained by dualizing the components of the DG-isomorphism

$$\mathrm{CE}_{-*}((\Lambda_H^n)_{i_0\le n\le i_d}, J, \alpha) \to \mathrm{CE}_{-*}((\Lambda_H^n)_{i_0\le n\le i_d}, J_H, \alpha)$$

induced by the path $((\Lambda_H^n)_{i_0\le n\le i_d}, J_t^{12})_{0\le t\le 1}$ (see Theorem 3.15).

Proof We have to prove that ζ_{12} is an isomorphism. This follows from the fact that it is defined by dualizing the components of a DG-isomorphism. □

Lemma 4.19 *The A_∞-functor τ_2 is homotopic to $\zeta_{12} \circ \tau_1 \circ \zeta_{12}^{-1}$ (see* [36, Paragraph (1h)]).

Proof First recall that τ_1 is obtained by dualizing the components of the DG-map

$$\mathrm{CE}_{-*}((\Lambda_H^{n+1})_{i_0\le n\le i_d}, J_H, \alpha) \to \mathrm{CE}_{-*}((\Lambda_H^n)_{i_0\le n\le i_d}, J_H, \alpha)$$

induced by the path $((\Lambda_H^{n+1-t})_{i_0\le n\le i_d}, J_H)_{0\le t\le 1}$. Thus, $\zeta_{12}\circ\tau_1$ is obtained by dualizing the components of the composition

$$\mathrm{CE}_{-*}((\Lambda_H^{n+1})_{i_0\le n\le i_d}, J, \alpha) \to \mathrm{CE}_{-*}((\Lambda_H^{n+1})_{i_0\le n\le i_d}, J_H, \alpha) \to \mathrm{CE}_{-*}((\Lambda_H^n)_{i_0\le n\le i_d}, J_H, \alpha).$$

On the other hand, τ_2 is obtained by dualizing the components of the DG-map

$$\mathrm{CE}_{-*}((\Lambda_H^{n+1})_{i_0\le n\le i_d}, J, \alpha) \to \mathrm{CE}_{-*}((\Lambda_H^n)_{i_0\le n\le i_d}, J, \alpha)$$

induced by the path $((\Lambda_H^{n+1-t})_{i_0\le n\le i_d}, J)_{0\le t\le 1}$. Thus, $\tau_2 \circ \zeta_{12}$ is obtained by dualizing the components of the composition

$$\mathrm{CE}_{-*}((\Lambda_H^{n+1})_{i_0\le n\le i_d}, J, \alpha) \to \mathrm{CE}_{-*}((\Lambda_H^n)_{i_0\le n\le i_d}, J, \alpha) \to \mathrm{CE}_{-*}((\Lambda_H^n)_{i_0\le n\le i_d}, J_H, \alpha).$$

According to Theorem 3.15, the DG-maps used to define $\zeta_{12}\circ\tau_1$ and $\tau_2\circ\zeta_{12}$ are DG-homotopic. Therefore the A_∞-functors $\zeta_{12} \circ \tau_1$ and $\tau_2 \circ \zeta_{12}$ are homotopic. □

Lemma 4.20 *The mapping torus of $\tau_1 : \mathcal{A}_1 \to \mathcal{A}_1$ is quasi-equivalent to the mapping torus of $\tau_2 : \mathcal{A}_2 \to \mathcal{A}_2$ (see Definition 2.1).*

Proof Let $\tau_{12} := \zeta_{12} \circ \tau_1 \circ \zeta_{12}^{-1}$. Consider the commutative diagram of Adams-graded A_∞-categories

$$\begin{tikzcd} \mathcal{A}_1 \arrow[d, "\zeta_{12}"] & \mathcal{A}_1 \sqcup \mathcal{A}_1 \arrow[l, "\mathrm{id} \sqcup \mathrm{id}"'] \arrow[r, "\mathrm{id} \sqcup \tau_1"] \arrow[d, "\zeta_{12} \sqcup \zeta_{12}"] & \mathcal{A}_1 \arrow[d, "\zeta_{12}"] \\ \mathcal{A}_2 & \mathcal{A}_2 \sqcup \mathcal{A}_2 \arrow[l, "\mathrm{id} \sqcup \mathrm{id}"'] \arrow[r, "\mathrm{id} \sqcup \tau_{12}"] & \mathcal{A}_2 \end{tikzcd}$$

Each vertical arrow is a quasi-equivalence according to Lemma 4.18, so it follows from Proposition 1.22 that the mapping torus of τ_1 is quasi-equivalent to the mapping torus of τ_{12}. Now according to Lemma 4.19, τ_{12} is homotopic to τ_2. Thus the result follows from Proposition 1.23. $\square$

4.5 Projection to P

4.5.1 The A_∞-category $\mathcal{O}$ In order to define the A_∞-category $\mathcal{O}$, we need to introduce moduli spaces of pseudoholomorphic discs in P.

Definition 4.21 Let $\boldsymbol{L} = (L^n(E))_{(n,E) \in \mathbb{Z} \times \mathcal{E}}$ be a family of mutually transverse connected compact exact Lagrangians in (P, λ). Consider a sequence of integers $i_0 < \dots < i_d$, and a family of intersection points $(x_0, x_1, \dots, x_d)$, where

$$x_0 \in L^{i_0}(E_0) \cap L^{i_d}(E_d) \qquad \text{and} \qquad x_k \in L^{i_{k-1}}(E_{k-1}) \cap L^{i_k}(E_k), \quad 1 \le k \le d.$$

(1) If $d = 1$, we denote by $\mathcal{M}_{x_1, x_0}(\boldsymbol{L}, j)$ the set of equivalence classes of maps $u : \mathbb{R} \times [0,1] \to P$ such that

- u maps $\mathbb{R} \times \{0\}$ to $L^{i_0}(E_0)$ and $\mathbb{R} \times \{1\}$ to $L^{i_1(E_1)}$,
- u satisfies the asymptotic conditions
$$u(s,t) \xrightarrow[s \to -\infty]{} x_1 \quad \text{and} \quad u(s,t) \xrightarrow[s \to +\infty]{} x_0,$$
- u is (i, j)-holomorphic,

where two maps u and u' are identified if there exists $s_0 \in \mathbb{R}$ such that $u'(\cdot, \cdot) = u(\cdot + s_0, \cdot)$.

(2) If $d \ge 2$, we denote by $\mathcal{M}_{x_d, \dots, x_1, x_0}(\boldsymbol{L}, j)$ the set of pairs (r, u) such that

- $r \in \mathcal{R}^{d+1}$ and $u : \Delta_r \to P$ maps the boundary arc (ζ_{k+1}, ζ_k) of Δ_r to $L^{i_k}(E_k)$,
- u satisfies the asymptotic conditions
$$(u \circ \epsilon_k(r))(s,t) \xrightarrow[s \to -\infty]{} x_k \quad \text{and} \quad (u \circ \epsilon_0(r))(s,t) \xrightarrow[s \to +\infty]{} x_0,$$
- u is (i, j)-holomorphic.

Recall that we have chosen a contactomorphism ϕ_H as in Lemma 4.11. We set

$$L_H^n := \Pi_P(\Lambda_H^n(E)) \subset P \quad \text{and} \quad \boldsymbol{L}_H := (L_H^n(E))_{(n,E) \in \mathbb{Z} \times \mathcal{E}}.$$

Definition 4.22 We denote by $\mathcal{O}$ the A_∞-category defined as follows:

(1) The objects of $\mathcal{O}$ are the Lagrangians $L_H^n(E)$, $(n, E) \in \mathbb{Z} \times \mathcal{E}$.

(2) The space of morphisms from $L_H^i(E)$ to $L_H^j(E')$ is either generated by $L_H^i(E) \cap L_H^j(E')$ if $(i, E) < (j, E')$, or $\mathbb{F}$ if $(i, E) = (j, E')$, or 0 otherwise.

(3) The operations are such that $1 \in \mathcal{O}(L_H^n(E), L_H^n(E))$ is a strict unit, and for every sequence $(i_0, E_0) < \cdots < (i_d, E_d)$, for every sequence of intersection points

$$(x_1, \dots, x_d) \in (L_H^{i_0}(E_0) \cap L_H^{i_1}(E_1)) \times \cdots \times (L_H^{i_{d-1}}(E_{d-1}) \cap L_H^{i_d}(E_d)),$$

we have

$$\mu_{\mathcal{O}}(x_1, \dots, x_d) = \sum_{x_0 \in L_H^{i_0}(E_0) \cap L_H^{i_d}(E_d)} \#\mathcal{M}_{x_d, \dots, x_1, x_0}(\boldsymbol{L}_H, j) x_0.$$

4.5.2 The quasi-autoequivalence γ Before defining the A_∞-functor $\gamma\colon \mathcal{O} \to \mathcal{O}$, we recall Legendrian contact homology as defined in [16]. To each generic Legendrian Λ in $\mathbb{R} \times P$, the authors associate a semifree DG-algebra $A = A(\Lambda, j)$ generated by the self-intersection points of $\Pi_P(\Lambda)$, with a differential $\partial\colon A \to A$ defined using j-holomorphic discs in P. In our case, the differential of $A\big(\bigsqcup_k \Lambda_H^k(E), j\big)$ on a generator $x_0 \in L_H^{i_0}(E_0) \cap L_H^{i_d}(E_d)$ is given by

$$\partial x_0 = \sum_{(x_1, \dots, x_d)} \#\mathcal{M}_{x_d, \dots, x_1, x_0}(\boldsymbol{L}_H, j) x_d \cdots x_1,$$

where the sum is over the sequences

$$(x_1, \dots, x_d) \in (L_H^{i_0}(E_0) \cap L_H^{i_1}(E_1)) \times \cdots \times (L_H^{i_{d-1}}(E_{d-1}) \cap L_H^{i_d}(E_d)).$$

According to [10, Theorem 2.1], Legendrian contact homology as defined in [16] coincides with the version exposed in Section 3:

$$A(\Lambda, j) = \mathrm{CE}_*(\Lambda, (D\Pi_P)|_\xi^* j, \alpha).$$

We introduced this version only because it makes clearer the fact that some operations are defined using pseudoholomorphic polygons in the base P.

Definition 4.23 We denote by $\gamma\colon \mathcal{O} \to \mathcal{O}$ the A_∞-functor defined as follows:

(1) On objects, $\gamma(L_H^n(E)) = L_H^{n+1}(E)$.

(2) On morphisms, the map

$$\gamma\colon \mathcal{O}(L_H^{i_0}(E_0), L_H^{i_1}(E_1)) \otimes \cdots \otimes \mathcal{O}(L_H^{i_{d-1}}(E_{d-1}), L_H^{i_d}(E_d)) \to \mathcal{O}(L_H^{i_0+1}(E_0), L_H^{i_d+1}(E_d))$$

is obtained by dualizing the components of the DG-isomorphism

$$A\Big(\bigsqcup_{k=i_0}^{i_d}\Lambda_H^{k+1}, j\Big)=\mathrm{CE}_{-*}\Big(\mathbb{R}\times\bigsqcup_{k=i_0}^{i_d}\Lambda_H^{k+1},(D\Pi_P)|_\xi^* j,\alpha\Big)$$
$$\to \mathrm{CE}_{-*}\Big(\mathbb{R}\times\bigsqcup_{k=i_0}^{i_d}\Lambda_H^{k},(D\Pi_P)|_\xi^* j,\alpha\Big)=A\Big(\bigsqcup_{k=i_0}^{i_d}\Lambda_H^{k}, j\Big)$$

induced by the Legendrian isotopy $\big(\bigsqcup_{k=i_0}^{i_d}\Lambda_H^{k+1-t}\big)_{0\le t\le 1}$ (see Theorem 3.15).

Remark (1) The A_∞-functor $\gamma\colon\mathcal{O}\to\mathcal{O}$ is a quasi-equivalence because it is defined by dualizing the components of a DG-isomorphism.

(2) The $\mathbb{Z}$-splitting

$$\mathbb{Z}\times\mathcal{E}\xrightarrow{\sim}\mathrm{ob}(\mathcal{O}),\quad (n,E)\mapsto L_H^n(E),$$

is compatible with the quasi-autoequivalence γ in the sense of Definition 2.2. As explained there, this turns $\mathcal{O}$ into an Adams-graded A_∞-category.

4.5.3 Relation with the previous invariants We now explain how the pairs $(\mathcal{A}_2,\tau_2)$ (Definition 4.17) and $(\mathcal{O},\gamma)$ are related. See Figure 4, where we illustrate the action of the projection Π_P in the case

$$(P,\lambda)=(T^*S^1,p\,dq),\quad L=0_{S^1},\quad\text{and}\quad H(q,p)=h(q),$$

where $h\colon S^1\to\mathbb{R}$ is a Morse function.

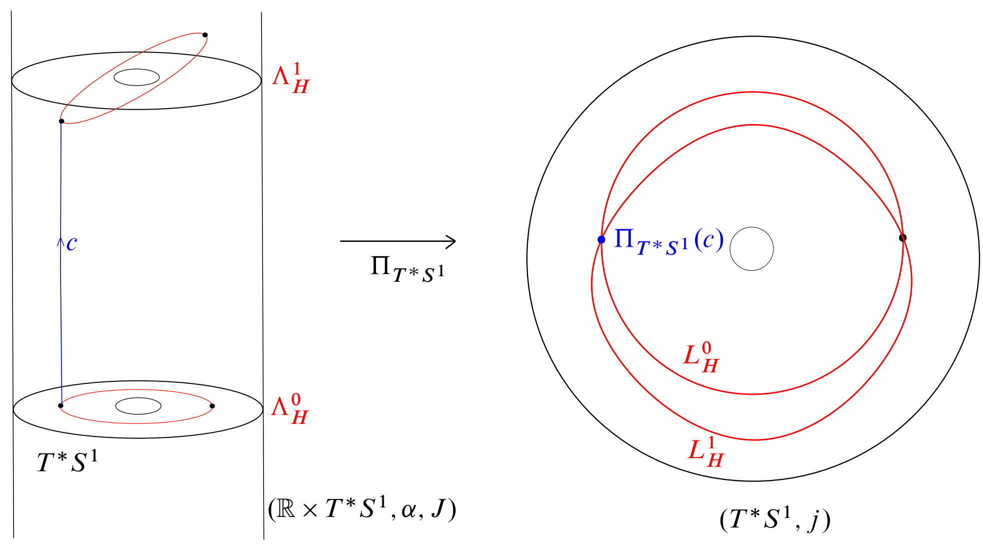


Figure 4: Action of the projection $\Pi_{T^*S^1}$.

Lemma 4.24 *There is a strict A_∞-isomorphism $\zeta_2\colon \mathcal{A}_2 \to \mathcal{O}$ defined as follows*:

(1) *On objects*, $\zeta_2(\Lambda^n_H(E)) = L^n_H(E)$.

(2) *On morphisms*, ζ_2 *sends a Reeb chord c in $\mathcal{A}_2(\Lambda^i_H(E), \Lambda^j_H(E'))$ to the intersection point*

$$\zeta_2(c) = \Pi_P(c)$$

in $\mathcal{O}(L^i_H(E), L^j_H(E'))$.

Proof We have to show that ζ_2 is an A_∞-map. Since $J = (D\Pi_P)|^*_\xi j$, it follows from [10, Theorem 2.1] that the map

$$u = (\sigma, v) \mapsto \Pi_P \circ v$$

induces a bijection

$$\mathcal{M}_{c_d,\dots,c_1,c_0}(\mathbb{R} \times \Lambda_H, J, \alpha) \xrightarrow{\sim} \mathcal{M}_{\Pi_P(c_d)\dots\Pi_P(c_1),\Pi_P(c_0)}(\boldsymbol{L}_H, j).$$

This implies the result. $\square$

Lemma 4.25
$$\gamma = \zeta_2 \circ \tau_2 \circ \zeta_2^{-1}.$$

Proof This follows from the definitions of τ_2, γ, ζ_2 and the fact that $J = (D\Pi_P)|^*_\xi j$. $\square$

Lemma 4.26 *The mapping torus of $\tau_2\colon \mathcal{A}_2 \to \mathcal{A}_2$ is quasi-equivalent to the mapping torus of $\gamma\colon \mathcal{O} \to \mathcal{O}$* (*see Definition 2.1*).

Proof According to Lemma 4.25 the following diagram of Adams-graded A_∞-categories is commutative:

$$\begin{tikzcd}
\mathcal{A}_2 \arrow[d, "\zeta_2"] & \mathcal{A}_2 \sqcup \mathcal{A}_2 \arrow[l, "\mathrm{id}\sqcup\mathrm{id}"'] \arrow[r, "\mathrm{id}\sqcup\tau_2"] \arrow[d, "\zeta_2\sqcup\zeta_2"] & \mathcal{A} \arrow[d, "\zeta_2"] \\
\mathcal{O} & \mathcal{O} \sqcup \mathcal{O} \arrow[l, "\mathrm{id}\sqcup\mathrm{id}"'] \arrow[r, "\mathrm{id}\sqcup\gamma"] & \mathcal{O}
\end{tikzcd}$$

Moreover, each vertical arrow is a quasi-equivalence according to Lemma 4.24. Thus, the result follows from Proposition 1.22. $\square$

4.6 Mapping torus of γ

In this section, we show that we can apply Theorem 2.5 (Theorem A in the introduction) in order to compute the mapping torus of $\gamma\colon \mathcal{O} \to \mathcal{O}$. This allows us to finish the proof of Theorem 4.3.

Recall that we have fixed a contactomorphism ϕ_H of V such that $\phi_H^*\alpha_H = \alpha$. Also recall that if θ is some real number, then

$$\Lambda^\theta(E) = \{(f_E(x) + \theta, x) \mid x \in L\}, \quad \Lambda^\theta_H(E) = \phi_H^{-1}(\Lambda^\theta(E)), \quad \text{and} \quad L^\theta_H(E) = \Pi_P(\Lambda^\theta_H(E)).$$

4.6.1 Continuation elements We denote by $\mathcal{O}_{2r}$ the A_∞-category obtained from $\mathcal{O}$ by applying the functor of Definition 1.27. We denote by

$$\Gamma = \left\{c_n(E) \in \mathcal{O}_{2r}(L^n_H(E), L^{n+1}_H(E)) \mid (n, E) \in \mathbb{Z} \times \mathcal{E}\right\}$$

the set of continuation elements in $\mathcal{O}_{2r}$ induced by the exact Lagrangian isotopies $(L^{n+t}_H)_{0\le t\le 1}$ (see for example [23, Section 3.3]).

Recall that if $\mathcal{C}$ is an A_∞-category equipped with a $\mathbb{Z}$-splitting of ob($\mathcal{C}$), we denote by $\mathcal{C}^0$ the full A_∞-subcategory of $\mathcal{C}$ whose set of objects corresponds to $\{0\} \times \mathcal{E}$.

Lemma 4.27 *There are quasi-equivalences of A_∞-categories*

$$\mathcal{O}^0_{2r} \simeq \overrightarrow{\mathcal{F}uk}(\boldsymbol{L}_H) \quad \textit{and} \quad \mathcal{O}_{2r}[\Gamma^{-1}]^0 \simeq \mathcal{F}uk(\boldsymbol{L}_H).$$

Proof First observe that we actually have $\mathcal{O}^0_{2r} = \overrightarrow{\mathcal{F}uk}(\boldsymbol{L}_H)$.

The second equivalence follows from the results of [37, Lecture 10; 23] about Fukaya categories and localization of A_∞-categories. More precisely, consider the subcategory $\mathcal{F}$ of $\mathcal{F}uk(P)$ with objects the Lagrangians $L^n(E)$. There is a trivial A_∞-functor $\mathcal{O}_{2r} \to \mathcal{F}$ (which is the identity on objects and on morphisms in $\mathcal{O}(L^i_H(E), L^j_H(E'))$ whenever $(i, E) < (j, E')$). Moreover, this functor sends continuation elements of $\mathcal{O}_{2r}$ to quasi-invertible morphisms in $\mathcal{F}$, and therefore induces an A_∞-functor $\mathcal{O}_{2r}[\Gamma^{-1}] \to \mathcal{F}$. Since the map

$$\mathcal{O}_{2r}(L^i_H(E), L^j_H(E')) \to \mathcal{O}_{2r}[\Gamma^{-1}](L^i_H(E), L^j_H(E'))$$

is a quasi-isomorphism whenever $(i, E) < (j, E')$, it follows that the functor $\mathcal{O}_{2r}[\Gamma^{-1}] \to \mathcal{F}$ is a quasi-equivalence. Thus we get

$$\mathcal{O}_{2r}[\Gamma^{-1}]^0 \simeq \mathcal{F}^0 = \mathcal{F}uk(\boldsymbol{L}_H). \qquad \square$$

4.6.2 The $\mathcal{O}_{2r}$-bimodule map In order to apply Theorem 2.5, we need a degree 0 closed $\mathcal{O}_{2r}$-module map $f : \mathcal{O}_{2r}(-,-) \to \mathcal{O}_{2r}(-, \gamma(-))$ such that the elements in f(units) satisfy certain hypotheses. As usual, we would like to find such an f geometrically, ie using some Lagrangian (or Legendrian) isotopy. However, here the unit $1 = e_{L^k_H(E)} \in \mathcal{O}(L^k_H(E), L^k_H(E))$, which is not an intersection point between Lagrangians, is supposed to be sent by f to something in $\mathcal{O}(L^k_H(E), L^{k+1}_H(E))$, which is generated by the intersection points between $L^k_H(E)$ and $L^{k+1}_H(E)$. Therefore, we need to somehow replace this unit by some intersection point between Lagrangians. The idea is that we will slightly perturb $L^k_H(E)$ to $L^{k+\delta}_H(E)$, and then replace $e_{L^k_H(E)}$ by the continuation element in the vector space generated by $L^k_H(E) \cap L^{k+\delta}_H(E)$.

Observe that if δ is small enough, $L^{k+\delta}_H(E)$ is a small perturbation of $L^k_H(E)$. Therefore, in a Weinstein neighborhood of $L^k_H(E)$, the Lagrangian $L^{k+\delta}_H(E)$ is the graph of $dh_{\delta,k,E}$, where $h_{\delta,k,E}$ is some Morse function on $L(E)$. In particular, the intersection points between $L^k_H(E)$ and $L^{k+\delta}_H(E)$ correspond to the critical points of $h_{\delta,k,E}$. Moreover, the continuation element in the vector space generated by $L^k_H(E) \cap L^{k+\delta}_H(E)$ corresponds to the sum of the minima of $h_{\delta,k,E}$.

Example 4.28 Assume that we are in the case

$$(P,\lambda)=(T^*M,p\,dq),\quad L=0_M,\quad \text{and}\quad H(q,p)=h(q),$$

where $h\colon M\to\mathbb{R}$ is a Morse function. As explained in Example 4.12, in this case we have

$$L_H^\theta=\Pi_{T^*M}(j^1(\theta e^h))=\operatorname{graph}(d(\theta e^h)).$$

Thus, $L_H^{k+\delta}$ is the graph of $d(\delta e^h)$ over L_H^k.

We will need the following result about moduli spaces of discs with boundary on small perturbations of the Lagrangians.

Lemma 4.29 *Let $g:=-d\lambda(-,j-)$ be the metric on P induced by j and $(-d\lambda)$. For every positive integer n, there exists $\delta_n>0$ such that the following holds for every $\delta\in\,]0,\delta_n]$. For every sequence*

$$(-n,E_0)\le(j_0,E_0)<\cdots<(j_p,E_p)\le(\ell_0,E_0')<\cdots<(\ell_q,E_q')\le(n,E_q'),\quad p,q\ge 0,$$

the rigid j-holomorphic discs in P with boundary on

$$L_H^{j_0}(E_0)\cup\cdots\cup L_H^{j_p}(E_p)\cup L_H^{\ell_0+\delta}(E_0')\cup\cdots\cup L_H^{\ell_q+\delta}(E_q')$$

are

(1) *in bijection with the rigid j-holomorphic discs in P with boundary on*

$$L_H^{j_0}(E_0)\cup\cdots\cup L_H^{j_p}(E_p)\cup L_H^{\ell_0}(E_0')\cup\cdots\cup L_H^{\ell_q}(E_q')$$

if $(j_p,E_p)<(\ell_0,E_0')$, or

(2) *in bijection with the rigid j-holomorphic discs in P with boundary on*

$$L_H^{j_0}(E_0)\cup\cdots\cup L_H^{j_{p-1}}(E_{p-1})\cup L_H^{\ell_0}(E_0')\cup L_H^{\ell_1}(E_1')\cup\cdots\cup L_H^{\ell_q}(E_q')$$

with a flow line of $(-\nabla_g h_{\delta,k,E_0'})$ attached on the component in $L_H^{\ell_0}(E_0')$ if $(j_p,E_p)=(\ell_0,E_0')$.

Proof The case $j_p<\ell_0$ follows from transversality of the moduli spaces in consideration. The case $j_p=\ell_0$ follows from the main analytic theorem of [13] (Theorem 3.6). □

In order to define the $\mathcal{O}_{2r}$-bimodule map f properly, we will use Lemma 4.29 to modify the A_∞-category $\mathcal{O}_{2r}$. In the following, we fix a *decreasing* sequence of positive real numbers $(\delta_n)_{n\ge1}$ such that, for every n,

(1) Lemma 4.29 holds with δ_n, and
(2) δ_n is small enough so that there is no handle slide instant in the Legendrian isotopy

$$\bigcup_{\ell=-n}^{n}\Lambda_H^{\ell+\delta_n t}=\bigcup_{\ell=-n}^{n}\bigcup_{E\in\mathcal{E}}\Lambda_H^{\ell+\delta_n t}(E),\quad t\in[0,1].$$

We define two families of A_∞-categories $(\mathcal{O}_{n,k})_{n,k}$ and $(\widetilde{\mathcal{O}}_{n,k})_{n,k}$ indexed by the couples (n,k), where $n\ge1$ and $-n\le k\le n$. The A_∞-category $\mathcal{O}_{n,k}$ is basically obtained from $\mathcal{O}_{2r}$ by restricting to objects $L_H^i(E)$, $-n\le i\le n$, and adding a copy of the object $L_H^k(E)$.

Definition 4.30 For every $(j, E) \in \mathbb{Z} \times \mathcal{E}$, let $\bar{L}_H^j(E)$ be a copy of $L_H^j(E)$. We denote by $\mathcal{O}_{n,k}$ the A_∞-category defined as follows:

(1) The set of objects of $\mathcal{O}_{n,k}$ is

$$\mathrm{ob}(\mathcal{O}_{n,k}) = \{L_H^j(E) \mid -n \le j \le k,\ E \in \mathcal{E}\} \cup \{\bar{L}_H^\ell(E) \mid k \le \ell \le n,\ E \in \mathcal{E}\}.$$

(2) The spaces of morphisms in $\mathcal{O}_{n,k}$ are the corresponding spaces of morphisms in $\mathcal{O}_{2r}$ when we replace $\bar{L}_H^\ell(E),\ k \le \ell \le n$, by $L_H^\ell(E)$, except that

$$\mathcal{O}_{n,k}(\bar{L}_H^k(E), L_H^k(E)) = \{0\}.$$

(3) The operations are the same as in $\mathcal{O}_{2r}$.

The A_∞-category $\widetilde{\mathcal{O}}_{n,k}$ is obtained from $\mathcal{O}_{n,k}$ by perturbing the objects $\bar{L}_H^\ell(E),\ k \le \ell \le n$, to $L_H^{\ell+\delta_n}(E)$.

Definition 4.31 Let

$$\Theta_{n,k} := \{-n, \dots, k\} \cup \{\ell + \delta_n \mid k \le \ell \le n\} \subset \mathbb{R}, \quad \text{and} \quad \widetilde{\boldsymbol{L}}_H := (L_H^\theta(E))_{(\theta,E)\in\Theta_{n,k}\times\mathcal{E}}.$$

We denote by $\widetilde{\mathcal{O}}_{n,k}$ the A_∞-category defined as follows:

(1) The objects of $\widetilde{\mathcal{O}}_{n,k}$ are the Lagrangians $L_H^\theta(E)$, $(\theta, E) \in \Theta_{n,k} \times \mathcal{E}$.

(2) The space of morphisms from $L_H^\theta(E)$ to $L_H^{\theta'}(E')$ is either generated by $L_H^\theta(E) \cap L_H^{\theta'}(E')$ if $(\theta, E) < (\theta', E')$, or $\mathbb{F}$ if $(\theta, E) = (\theta', E')$, or 0 otherwise.

(3) The operations are such that $e_{L_H^\theta(E)} = 1 \in \widetilde{\mathcal{O}}_{n,k}(L_H^\theta(E), L_H^\theta(E))$ is a strict unit, and for every sequence $(\theta_0, E_0) < \cdots < (\theta_d, E_d)$, for every sequence of intersection points

$$(x_1, \dots, x_d) \in (L_H^{\theta_0}(E_0) \cap L_H^{\theta_1}(E_1)) \times \cdots \times (L_H^{\theta_{d-1}}(E_{d-1}) \cap L_H^{\theta_d}(E_d)),$$

we have

$$\mu_{\widetilde{\mathcal{O}}_{n,k}}(x_1, \dots, x_d) = \sum_{x_0 \in L_H^{\theta_0}(E_0)\cap L_H^{\theta_d}(E_d)} \#\mathcal{M}_{x_d,\dots,x_1,x_0}(\widetilde{\boldsymbol{L}}_H, j)x_0.$$

These A_∞-categories being defined, Lemma 4.29 implies the following result.

Lemma 4.32 *There is a strict A_∞-functor $\rho_{n,k} : \mathcal{O}_{n,k} \to \widetilde{\mathcal{O}}_{n,k}$ defined as follows:*

(1) *On objects, $\rho_{n,k}(L_H^j(E)) = L_H^j(E)$ if $-n \le j \le k$ and $\rho_{n,k}(\bar{L}_H^\ell(E)) = L_H^{\ell+\delta_n}(E)$ if $k \le \ell \le n$.*

(2) *On morphisms, $\rho_{n,k}$ sends the unit of $\mathcal{O}_{n,k}(L_H^k(E), \bar{L}_H^k(E)) = \mathbb{F}$ to the continuation element in $\widetilde{\mathcal{O}}_{n,k}(L_H^k(E), L_H^{k+\delta_n}(E))$, and it sends any other morphism of $\mathcal{O}_{n,k}$ to the corresponding one in $\widetilde{\mathcal{O}}_{n,k}$.*

Proof Consider a sequence $(x_0, \dots, x_{d-1})$ of morphisms in $\mathcal{O}_{n,k}$. If in this sequence there is no morphism from $L_H^k(E)$ to $\bar{L}_H^k(E)$, then the relation

$$\mu_{\widetilde{\mathcal{O}}_{n,k}}(\rho_{n,k}x_0, \dots, \rho_{n,k}x_d) = \rho_{n,k}(\mu_{\mathcal{O}_{n,k}}(x_0, \dots, x_d))$$

follows directly from the first item of Lemma 4.29. Now assume that there is $p \in \{0, \dots, d-1\}$ such that $x_p = e_{L^k_H(E)} \in \mathcal{O}_{n,k}(L^k_H(E), \bar{L}^k_H(E))$. Recall that the continuation element in $\widetilde{\mathcal{O}}_{n,k}(L^k_H(E), L^{k+\delta_n}_H(E))$ corresponds to the sum of the minima of $h_{\delta_n,k,E}$. Then the second item of Lemma 4.29 implies that

$$\mu_{\widetilde{\mathcal{O}}_{n,k}}(\rho_{n,k}x_0, \dots, \rho_{n,k}x_d) = \begin{cases} \rho_{n,k}x_1 & \text{if } d=1 \text{ and } p=0, \\ \rho_{n,k}x_0 & \text{if } d=1 \text{ and } p=1, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the A_∞-relation for $\rho_{n,k}$ is still satisfied according to the behavior of the operations $\mu_{\mathcal{O}_{n,k}}$ with respect to the unit $e_{L^k_H(E)}$. □

We can now define geometrically an A_∞-functor that will finally allow us to define the $\mathcal{O}_{2r}$-bimodule map f.

Definition 4.33 We denote by $\nu_{n,k} : \widetilde{\mathcal{O}}_{n,k} \to \mathcal{O}_{2r}$ the A_∞-functor defined as follows:

(1) On objects, $\nu_{n,k}(L^j_H(E)) = L^j_H(E)$ if $-n \le j \le k$, and $\nu_{n,k}(L^{\ell+\delta_n}_H(E)) = L^{\ell+1}_H(E)$ if $k \le \ell \le n$.

(2) On morphisms, $\nu_{n,k}$ is obtained by dualizing the components of the DG-isomorphism

$$A\Big(\bigsqcup_{i=-n}^{n+1} \Lambda^i_H\Big) \xrightarrow{\sim} A\Big(\bigsqcup_{j=-n}^{k} \Lambda^j_H \sqcup \bigsqcup_{\ell=k}^{n} \Lambda^{\ell+\delta_n}_H\Big).$$

induced by the Legendrian isotopy

$$\Big(\bigsqcup_{j=-n}^{k} \Lambda^j_H\Big) \sqcup \Big(\bigsqcup_{\ell=k}^{n} \Lambda^{\ell+1-t(1-\delta_n)}_H\Big), \quad t \in [0,1]$$

(see Theorem 3.15 or [16, Proposition 2.6]).

Remark 4.34 We point out some properties of the A_∞-functor

$$\sigma_{n,k} := \nu_{n,k} \circ \rho_{n,k} : \mathcal{O}_{n,k} \to \mathcal{O}_{2r}.$$

(1) Let $n \le p$ be two positive integers, and let $k \in \{-n, \dots, n\}$. Recall that we have chosen δ_n small enough so that there is no handle slide instant in the Legendrian isotopy

$$\bigsqcup_{\ell=-n}^{n} \Lambda^{\ell+\delta_n t}_H, \quad 0 \le t \le 1.$$

Since $\delta_p \le \delta_n$, neither is there any handle slide instant in the Legendrian isotopy

$$\bigsqcup_{\ell=-n}^{n} \Lambda^{\ell+\delta_p t}_H, \quad 0 \le t \le 1.$$

Therefore, $\sigma_{p,k}$ agrees with $\sigma_{n,k}$ on $\mathcal{O}_{n,k} \subset \mathcal{O}_{p,k}$.

(2) Consider a sequence of integers

$$-n \le j_0 < \cdots < j_p \le k_1 < k_2 \le \ell_0 < \cdots < \ell_q \le n,$$

and a sequence of morphisms

$$\begin{aligned}(x_0, \dots, x_{p-1}, u, y_0, \dots, y_{q-1}) \\ \in \mathcal{O}_{n,k_i}(L_H^{j_0}(E_0), L_H^{j_1}(E_1)) \times \cdots \times \mathcal{O}_{n,k_i}(L_H^{j_{p-1}}(E_{p-1}), L_H^{j_p}(E_p)) \times \mathcal{O}_{n,k_i}(L_H^{j_p}(E_p), \bar{L}_H^{\ell_0}(E_0')) \\ \times \mathcal{O}_{n,k_i}(\bar{L}_H^{\ell_0}(E_0'), \bar{L}_H^{\ell_1}(E_1')) \times \cdots \times \mathcal{O}_{n,k_i}(\bar{L}_H^{\ell_{q-1}}(E_{q-1}'), \bar{L}_H^{\ell_q}(E_q')).\end{aligned}$$

Since the Legendrian isotopy defining ν_{n,k_i} is

$$\Bigl(\bigsqcup_{j=-n}^{k_i} \Lambda_H^j\Bigr) \sqcup \Bigl(\bigsqcup_{\ell=k_i}^{n} \Lambda_H^{\ell+1-t(1-\delta_n)}\Bigr), \quad t \in [0,1],$$

we have

$$\begin{aligned}&\sigma_{n,k_1}(x_0, \dots, x_{p-1}) = \delta_{1p} x_0,\\ &\sigma_{n,k_2}(y_0, \dots, y_{q-1}) = \gamma(y_0, \dots, y_{q-1}),\\ &\sigma_{n,k_2}(x_0, \dots, x_{p-1}, u, y_0, \dots, y_{q-1}) = \sigma_{n,k_1}(x_0, \dots, x_{p-1}, u, y_0, \dots, y_{q-1}).\end{aligned}$$

(3) By construction, the A_∞-functor $\nu_{n,k}$ sends the continuation element in $\widetilde{\mathcal{O}}_{n,k}(L_H^k(E), L_H^{k+\delta_n}(E))$ (corresponding to the sum of the minima of $h_{\delta_n,k,E}$) to the continuation element

$$c_k(E) \in \mathcal{O}_{2r}(L_H^k(E), L_H^{k+1}(E)).$$

In other words, $\sigma_{n,k}$ sends the unit $e_{L_H^k(E)} \in \mathcal{O}_{n,k}(L_H^k(E), \bar{L}_H^k(E))$ to $c_k(E)$.

(4) The map $\sigma_{n,k}\colon \mathcal{O}_{n,k}(L_H^j(E), \bar{L}_H^k(E')) \to \mathcal{O}_{2r}(L_H^j(E), L_H^{k+1}(E'))$ is a quasi-isomorphism for every $j < k$ and $E, E' \in \mathcal{E}$.

We can now state and prove the desired result.

Lemma 4.35 *There exists a degree* 0 *closed $\mathcal{O}_{2r}$-bimodule map $f\colon \mathcal{O}_{2r}(-,-) \to \mathcal{O}_{2r}(-,\gamma(-))$ which sends the unit $e_{L_H^k(E)} \in \mathcal{O}_{2r}(L_H^k(E), L_H^k(E))$ to the continuation element*

$$c_k(E) \in \mathcal{O}_{2r}(L_H^k(E), L_H^{k+1}(E)) \cap \Gamma,$$

and such that $f\colon \mathcal{O}_{2r}(L_H^j(E), L_H^k(E')) \to \mathcal{O}_{2r}(L_H^j(E), L_H^{k+1}(E'))$ is a quasi-isomorphism for every $j < k$ and $E, E' \in \mathcal{E}$.

Proof Consider a sequence

$$(j_0, E_0) < \cdots < (j_p, E_p) \le (k, E) = (\ell_0, E_0') < \cdots < (\ell_q, E_q'),$$

and a sequence of morphisms

$$\begin{aligned}(x_0,\dots,x_{p-1},u,y_0,\dots,y_{q-1})&\\ \in \mathcal{O}_{2r}(L_H^{j_0}(E_0),L_H^{j_1}(E_1))\times\cdots\times\mathcal{O}_{2r}(L_H^{j_{p-1}}(E_{p-1}),L_H^{j_p}(E_p))\times\mathcal{O}_{2r}(L_H^{j_p}(E_p),L_H^{k}(E_0'))&\\ \times\mathcal{O}_{2r}(L_H^{k}(E_0'),L_H^{\ell_1}(E_1'))\times\cdots\times\mathcal{O}_{2r}(L_H^{\ell_{q-1}}(E_{q-1}'),L_H^{\ell_q}(E_q')).&\end{aligned}$$

We choose $n\geq 1$ such that $-n\leq j_0\leq \ell_q\leq n$, and we set

$$f(x_0,\dots,x_{p-1},u,y_0,\dots,y_{q-1}):=\sigma_{n,k}(x_0,\dots,x_{p-1},u,y_0,\dots,y_{q-1})\in\mathcal{O}_{2r}(L_H^{j_0}(E_0),\gamma L_H^{\ell_q}(E_q')),$$

where on the right-hand side we consider that

$$\begin{aligned}(x_0,\dots,x_{p-1},u,y_0,\dots,y_{q-1})&\\ \in \mathcal{O}_{n,k}(L_H^{j_0}(E_0),L_H^{j_1}(E_1))\times\cdots\times\mathcal{O}_{n,k}(L_H^{j_{p-1}}(E_{p-1}),L_H^{j_p}(E_p))\times\mathcal{O}_{n,k}(L_H^{j_p}(E_p),\bar{L}_H^{k}(E_0'))&\\ \times\mathcal{O}_{n,k}(\bar{L}_H^{k}(E_0'),\bar{L}_H^{\ell_1}(E_1'))\times\cdots\times\mathcal{O}_{n,k}(\bar{L}_H^{\ell_{q-1}}(E_{q-1}'),\bar{L}_H^{\ell_q}(E_q')).&\end{aligned}$$

Observe that f is well defined (it does not depend on the choice of n) according to the first item of Remark 4.34.

We now verify that f is closed. According to Definition 1.4, we have

$$\begin{aligned}&\mu^1_{\mathrm{Mod}_{\mathcal{C},\mathcal{C}}}(f)(x_0,\dots,x_{p-1},u,y_0,\dots,y_{q-1})\\ &=\sum\sigma_{n,k}(\dots,\mu_{\mathcal{O}_{2r}}(\dots),\dots,u,\dots)+\sum\sigma_{n,\ell_s}(\dots,\mu_{\mathcal{O}_{2r}}(x_r,\dots,x_{p-1},u,y_0,\dots,y_{s-1}),\dots)\\ &\qquad+\sum\sigma_{n,k}(\dots,u,\dots,\mu_{\mathcal{O}_{2r}}(\dots),\dots)+\sum\mu_{\mathcal{O}_{2r}}(\dots,\sigma_{n,k}(\dots,u,\dots),\gamma(\dots),\dots,\gamma(\dots)).\end{aligned}$$

Now according to the second item of Remark 4.34, we have

$$\begin{aligned}&\sum\sigma_{n,\ell_s}(\dots,\mu_{\mathcal{O}_{2r}}(x_r,\dots,x_{p-1},u,y_0,\dots,y_{s-1}),\dots)\\ &\qquad=\sum\sigma_{n,k}(\dots,\mu_{\mathcal{O}_{2r}}(x_r,\dots,x_{p-1},u,y_0,\dots,y_{s-1}),\dots)\end{aligned}$$

and

$$\begin{aligned}&\sum\mu_{\mathcal{O}_{2r}}(\dots,\sigma_{n,k}(\dots,u,\dots),\gamma(\dots),\dots,\gamma(\dots))\\ &\qquad=\sum\mu_{\mathcal{O}_{2r}}\big(\sigma_{n,k}(\dots),\dots,\sigma_{n,k}(\dots),\sigma_{n,k}(\dots,u,\dots),\sigma_{n,k}(\dots),\dots,\sigma_{n,k}(\dots)\big).\end{aligned}$$

Therefore, we get

$$\mu^1_{\mathrm{Mod}_{\mathcal{C},\mathcal{C}}}(f)(x_0,\dots,x_{p-1},u,y_0,\dots,y_{q-1})=0$$

from the fact that $\sigma_{n,k}$ is an A_∞-functor.

Now f sends the unit $e_{L_H^k(E)}\in\mathcal{O}_{2r}(L_H^k(E),L_H^k(E))$ to the continuation element

$$c_k(E)\in\mathcal{O}_{2r}(L_H^k(E),L_H^{k+1}(E))\cap\Gamma$$

according to the third item of Remark 4.34. Finally, the map

$$f\colon \mathcal{O}_{2r}(L_H^j(E), L_H^k(E')) \to \mathcal{O}_{2r}(L_H^j(E), L_H^{k+1}(E'))$$

is a quasi-isomorphism for every $j < k$ and $E, E' \in \mathcal{E}$ according to the last item of Remark 4.34. □

4.6.3 Proof of the main result We end the section with the proof of Theorem 4.3 (Theorem B in the introduction).

Recall that we denote by $\mathbb{F}[t_m]$ the augmented Adams-graded associative algebra generated by a variable t_m of bidegree $(m,1)$, and by $t_m\mathbb{F}[t_m]$ its augmentation ideal (or equivalently, the ideal generated by t_m). The key result is the following.

Lemma 4.36 *The mapping torus of γ is quasi-equivalent to the Adams-graded A_∞-category*

$$\overrightarrow{\mathcal{F}uk}(\boldsymbol{L}) \oplus (t_{2r}\mathbb{F}[t_{2r}] \otimes \mathcal{F}uk(\boldsymbol{L})).$$

Proof Let $f\colon \mathcal{O}_{2r}(-,-) \to \mathcal{O}_{2r}(-,\gamma(-))$ be the degree 0 closed bimodule map of Lemma 4.35. According to the latter, the hypotheses of Theorem 2.5 are satisfied, and $f(\text{units}) = \Gamma$. Thus the mapping torus of γ is quasi-equivalent to the Adams-graded A_∞-algebra $\mathcal{O}_{2r}^0 \oplus (t_{2r}\mathbb{F}[t_{2r}] \otimes \mathcal{O}_{2r}[\Gamma^{-1}]^0)$ (recall that if $\mathcal{C}$ is an A_∞-category equipped with a $\mathbb{Z}$-splitting $\mathbb{Z}\times\mathcal{E} \simeq \text{ob}(\mathcal{C})$, we denote by $\mathcal{C}^0$ the full A_∞-subcategory of $\mathcal{C}$ whose set of objects corresponds to $\{0\}\times\mathcal{E}$). According to Lemma 4.27 we have

$$\mathcal{O}_{2r}^0 \simeq \overrightarrow{\mathcal{F}uk}(\boldsymbol{L}_H) \quad \text{and} \quad \mathcal{O}_{2r}[\Gamma^{-1}]^0 \simeq \mathcal{F}uk(\boldsymbol{L}_H).$$

The result follows from invariance of the Fukaya category (see [36, Section (10a)])

$$\overrightarrow{\mathcal{F}uk}(\boldsymbol{L}_H) \simeq \overrightarrow{\mathcal{F}uk}(\boldsymbol{L}) \quad \text{and} \quad \mathcal{F}uk(\boldsymbol{L}_H) \simeq \mathcal{F}uk(\boldsymbol{L}).$$ □

We now give the proof of Theorem 4.3 (Theorem B in the introduction). According to [29, Theorem 2.4], Koszul duality holds for the augmented Adams-graded DG-algebra $\mathrm{CE}^r_{-*}(\boldsymbol{\Lambda}^\circ)$ because it is *Adams connected* (see [29, Definition 2.1]). Indeed, recall from Section 4.1.2 that the Adams degree in $\mathrm{CE}^r_{-*}(\boldsymbol{\Lambda}^\circ)$ of a Reeb chord c is the number of times c winds around the fiber. Besides, recall from Section 4.1.2 that there is a coaugmented Adams-graded A_∞-cocategory $\mathrm{LC}_*(\boldsymbol{\Lambda}^\circ)$ such that

$$\mathrm{CE}^r_{-*}(\boldsymbol{\Lambda}^\circ) = \Omega(\mathrm{LC}_*(\boldsymbol{\Lambda}^\circ)) \quad \text{and} \quad \mathrm{LA}^*(\boldsymbol{\Lambda}^\circ) = \mathrm{LC}_*(\boldsymbol{\Lambda}^\circ)^\#.$$

Since there is a quasi-isomorphism $B(\Omega C) \simeq C$ for every A_∞-cocategory C (see [17, Section 2.2.2]), it follows that

$$E(\mathrm{CE}^r_{-*}(\boldsymbol{\Lambda}^\circ)) = B(\mathrm{CE}^r_{-*}(\boldsymbol{\Lambda}^\circ))^\# \simeq \mathrm{LC}_*(\boldsymbol{\Lambda}^\circ)^\# = \mathrm{LA}^*(\boldsymbol{\Lambda}^\circ)$$

(graded dual preserves quasi-isomorphisms). Now the quasi-equivalence

$$\mathrm{LA}^*(\boldsymbol{\Lambda}^\circ) \simeq \overrightarrow{\mathcal{F}uk}(\boldsymbol{L}) \oplus (t_{2r}\mathbb{F}[t_{2r}] \otimes \mathcal{F}uk(\boldsymbol{L}))$$

follows from Lemmas 4.10, 4.16, 4.20, 4.26 and 4.36. This concludes the proof.

References

[1] **C Abbas**, *An introduction to compactness results in symplectic field theory*, Springer (2014) MR Zbl

[2] **J F Adams**, *On the cobar construction*, Proc. Nat. Acad. Sci. USA 42 (1956) 409–412 MR Zbl

[3] **J F Adams**, **P J Hilton**, *On the chain algebra of a loop space*, Comment. Math. Helv. 30 (1956) 305–330 MR Zbl

[4] **W M Boothby**, **H C Wang**, *On contact manifolds*, Ann. of Math. 68 (1958) 721–734 MR Zbl

[5] **F Bourgeois**, **T Ekholm**, **Y Eliashberg**, *Effect of Legendrian surgery*, Geom. Topol. 16 (2012) 301–389 MR Zbl With an appendix by S Ganatra and M Maydanskiy

[6] **F Bourgeois**, **Y Eliashberg**, **H Hofer**, **K Wysocki**, **E Zehnder**, *Compactness results in symplectic field theory*, Geom. Topol. 7 (2003) 799–888 MR Zbl

[7] **B Chantraine**, **G Dimitroglou Rizell**, **P Ghiggini**, **R Golovko**, *Geometric generation of the wrapped Fukaya category of Weinstein manifolds and sectors*, Ann. Sci. École Norm. Sup. 57 (2024) 1–85 MR Zbl

[8] **Y Chekanov**, *Differential algebra of Legendrian links*, Invent. Math. 150 (2002) 441–483 MR Zbl

[9] **G Dimitroglou Rizell**, *Legendrian ambient surgery and Legendrian contact homology*, J. Symplectic Geom. 14 (2016) 811–901 MR Zbl

[10] **G Dimitroglou Rizell**, *Lifting pseudo-holomorphic polygons to the symplectisation of $P \times \mathbb{R}$ and applications*, Quantum Topol. 7 (2016) 29–105 MR Zbl

[11] **S K Donaldson**, *Symplectic submanifolds and almost-complex geometry*, J. Differential Geom. 44 (1996) 666–705 MR Zbl

[12] **T Ekholm**, *Rational symplectic field theory over $\mathbb{Z}_2$ for exact Lagrangian cobordisms*, J. Eur. Math. Soc. 10 (2008) 641–704 MR Zbl

[13] **T Ekholm**, **J B Etnyre**, **J M Sabloff**, *A duality exact sequence for Legendrian contact homology*, Duke Math. J. 150 (2009) 1–75 MR Zbl

[14] **T Ekholm**, **J Etnyre**, **M Sullivan**, *The contact homology of Legendrian submanifolds in $\mathbb{R}^{2n+1}$*, J. Differential Geom. 71 (2005) 177–305 MR Zbl

[15] **T Ekholm**, **J Etnyre**, **M Sullivan**, *Non-isotopic Legendrian submanifolds in $\mathbb{R}^{2n+1}$*, J. Differential Geom. 71 (2005) 85–128 MR Zbl

[16] **T Ekholm**, **J Etnyre**, **M Sullivan**, *Legendrian contact homology in $P \times \mathbb{R}$*, Trans. Amer. Math. Soc. 359 (2007) 3301–3335 MR Zbl

[17] **T Ekholm**, **Y Lekili**, *Duality between Lagrangian and Legendrian invariants*, Geom. Topol. 27 (2023) 2049–2179 MR Zbl

[18] **T Ekholm**, **L Ng**, *Legendrian contact homology in the boundary of a subcritical Weinstein 4-manifold*, J. Differential Geom. 101 (2015) 67–157 MR Zbl

[19] **T Ekholm**, **A Oancea**, *Symplectic and contact differential graded algebras*, Geom. Topol. 21 (2017) 2161–2230 MR

[20] **Y Eliashberg**, *Invariants in contact topology*, from "Proceedings of the International Congress of Mathematicians, II" (G Fischer, U Rehmann, editors), Deutsche Mathematiker Vereinigung, Berlin (1998) 327–338 MR Zbl

[21] **Y Eliashberg**, **A Givental**, **H Hofer**, *Introduction to symplectic field theory*, Geom. Funct. Anal. (2000) 560–673 MR Zbl

[22] **S Ganatra**, *Symplectic cohomology and duality for the wrapped Fukaya category*, preprint (2013) arXiv 1304.7312

[23] **S Ganatra**, **J Pardon**, **V Shende**, *Covariantly functorial wrapped Floer theory on Liouville sectors*, Publ. Math. Inst. Hautes Études Sci. 131 (2020) 73–200 MR Zbl

[24] **S Ganatra**, **J Pardon**, **V Shende**, *Sectorial descent for wrapped Fukaya categories*, J. Amer. Math. Soc. 37 (2024) 499–635 MR Zbl

[25] **Y B Kartal**, *Distinguishing open symplectic mapping tori via their wrapped Fukaya categories*, Geom. Topol. 25 (2021) 1551–1630 MR Zbl

[26] **Y B Kartal**, *Dynamical invariants of mapping torus categories*, Adv. Math. 389 (2021) art. id. 107882 MR Zbl

[27] **B Keller**, *On triangulated orbit categories*, Doc. Math. 10 (2005) 551–581 MR Zbl

[28] **B Keller**, *On differential graded categories*, from "Proceedings of the International Congress of Mathematicians, II" (M Sanz-Solé, J Soria, J L Varona, J Verdera, editors), Eur. Math. Soc., Zürich (2006) 151–190 MR Zbl

[29] **D M Lu**, **J H Palmieri**, **Q S Wu**, **J J Zhang**, *Koszul equivalences in A_∞-algebras*, New York J. Math. 14 (2008) 325–378 MR Zbl

[30] **V Lyubashenko**, **S Ovsienko**, *A construction of quotient A_∞-categories*, Homology Homotopy Appl. 8 (2006) 157–203 MR Zbl

[31] **Y Pan**, **D Rutherford**, *Functorial LCH for immersed Lagrangian cobordisms*, J. Symplectic Geom. 19 (2021) 635–722 MR Zbl

[32] **A Petr**, *Invariants of the Legendrian lift of an exact Lagrangian submanifold in the circular contactization of a Liouville manifold*, PhD thesis, Nantes Université (2022) Available at `https://hal.science/tel-03793906`

[33] **J Robbin**, **D Salamon**, *The Maslov index for paths*, Topology 32 (1993) 827–844 MR Zbl

[34] **J M Sabloff**, *Invariants of Legendrian knots in circle bundles*, Commun. Contemp. Math. 5 (2003) 569–627 MR Zbl

[35] **P Seidel**, *A_∞-subalgebras and natural transformations*, Homology Homotopy Appl. 10 (2008) 83–114 MR Zbl

[36] **P Seidel**, *Fukaya categories and Picard–Lefschetz theory*, Eur. Math. Soc., Zürich (2008) MR Zbl

[37] **P Seidel**, *Lectures on categorical dynamics and symplectic topology*, lecture notes (2013) Available at `https://math.mit.edu/~seidel/texts/937-lecture-notes.pdf`

Center for Quantum Mathematics, University of Southern Denmark (SDU)
Odense, Denmark

adrianpetrpetr@gmail.com

Received: 15 December 2022 Revised: 24 August 2023

Algebraic & Geometric Topology **25**:1 (2025) 563–644
DOI: 10.2140/agt.2025.25.563
Published: 24 March 2025

Infinite-type loxodromic isometries of the relative arc graph

Carolyn Abbott
Nicholas Miller
Priyam Patel

An infinite-type surface Σ is admissible if it has an isolated puncture p and admits shift maps. This includes all infinite-type surfaces with an isolated puncture outside of two sporadic classes. Given such a surface, we construct an infinite family of intrinsically infinite-type mapping classes that act loxodromically on the relative arc graph $\mathcal{A}(\Sigma, p)$. J Bavard produced such an element for the plane minus a Cantor set, and our result gives the first examples of such mapping classes for all other admissible surfaces. The elements we construct are the composition of three shift maps on Σ, and we give an alternative characterization of these elements as a composition of a pseudo-Anosov on a finite-type subsurface of Σ and a standard shift map. We then explicitly find their limit points on the boundary of $\mathcal{A}(\Sigma, p)$ and their limiting geodesic laminations. Finally, we show that these infinite-type elements can be used to prove that $\mathrm{Map}(\Sigma, p)$ has an infinite-dimensional space of quasimorphisms.

20F65, 57K20

1 Introduction

A surface Σ is of finite-type if $\pi_1(\Sigma)$ is finitely generated, and otherwise Σ is of infinite-type. Recently, there has been a surge of interest in infinite-type surfaces and their mapping class groups $\mathrm{Map}(\Sigma)$, which arise naturally in a variety of contexts in low-dimensional topology, dynamics, and even descriptive set theory. See the overview by Aramayona and Vlamis [4] for a survey of recent results on infinite-type mapping class groups.

For finite-type surfaces Σ, Nielsen [21] and Thurston [27] give a powerful classification of the elements of $\mathrm{Map}(\Sigma)$: every element is periodic, reducible, or pseudo-Anosov. The action of $\mathrm{Map}(\Sigma)$ by isometries on the (infinite-diameter and hyperbolic) curve graph $\mathcal{C}(\Sigma)$ captures a coarser classification of the elements of $\mathrm{Map}(\Sigma)$ since elements are either *elliptic* or *loxodromic*. These two classifications, which are both interesting in their own right, have a strong relationship; the loxodromic elements are exactly the pseudo-Anosovs. In this way, the most interesting and complex mapping classes correspond to the dynamically richest actions.

The situation for infinite-type surfaces is more complicated for a few reasons. First, the exact analog of the Nielsen–Thurston classification is no longer valid in this setting since some elements are neither periodic, reducible, nor pseudo-Anosov in the traditional sense. Second, the curve graph of an infinite-type surface has finite diameter unlike for finite-type surfaces. This paper is motivated by one of the biggest open problems for infinite-type surfaces, which is to give an analog of the Nielsen–Thurston classification for infinite-type mapping classes. We work towards this goal by studying the action of $\mathrm{Map}(\Sigma)$ on a different hyperbolic graph.

When Σ is an infinite-type surface with at least one isolated puncture p, the *relative arc graph*, $\mathcal{A}(\Sigma, p)$, plays the role of $\mathcal{C}(\Sigma)$ and is defined as follows: the vertices correspond to isotopy classes of simple arcs that begin and end at p and edges connect vertices for arcs admitting disjoint representatives. The subgroup $\mathrm{Map}(\Sigma, p)$ of $\mathrm{Map}(\Sigma)$ that fixes the isolated puncture p acts on $\mathcal{A}(\Sigma, p)$ by isometries. This graph was first defined by D Calegari [12], who initiated its study by asking whether, for the plane minus a Cantor set, this graph was infinite diameter and whether any element of $\mathrm{Map}(\Sigma, p)$ acted loxodromically. In [5], J Bavard carried out Caelgari's program for the plane minus a Cantor set and, for that surface, showed that $\mathcal{A}(\Sigma, p)$ is both infinite-diameter and hyperbolic. Aramayona, Fossas, and Parlier [2] then showed that these properties for $\mathcal{A}(\Sigma, p)$ hold more generally for any infinite-type surface with at least one isolated puncture.

Given that the trichotomy of the Nielsen–Thurston classification does not exactly hold for infinite-type surfaces, it is necessary to redefine reducible, and therefore irreducible, mapping classes in this setting. One of the most promising ways to motivate a new definition is to classify the elements of infinite-type mapping class groups that are loxodromic with respect to the action of $\mathrm{Map}(\Sigma)$ on a hyperbolic graph since these elements correspond to infinite-order irreducibles in the finite-type setting. In order to classify these elements, we must first construct them. We restrict to surfaces with an isolated puncture and their associated graphs $\mathcal{A}(\Sigma, p)$ in this paper because the relative arc graph is one of the few known graphs associated to an infinite-type surface that is both infinite-diameter and hyperbolic. In general, proving results for infinite-type surfaces typically involves a piecemeal approach. Constructing loxodromic isometries for infinite-type surfaces without a puncture will require a different graph, and thus, a genuinely different approach.

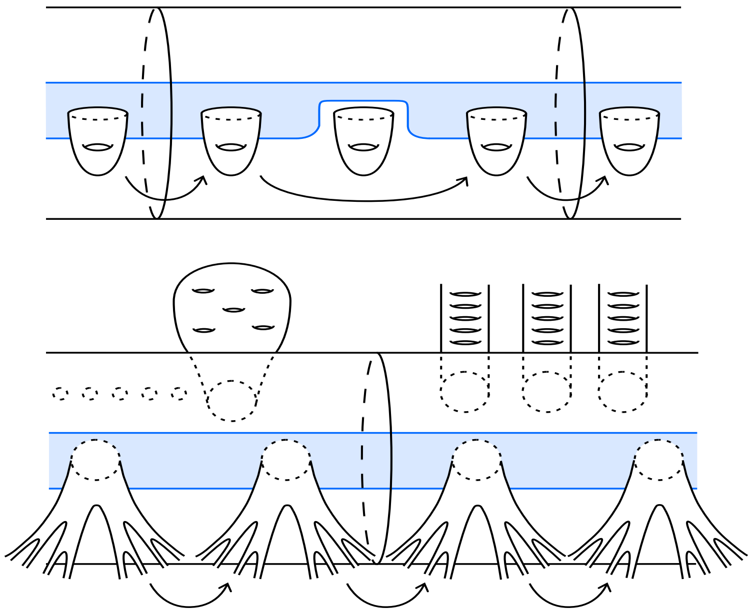


Figure 1: Above: a handleshift on an infinite-type surface. Below: a shift map on an infinite-type surface.

When Σ is the sphere minus a Cantor set with an isolated puncture p (ie Σ is the plane minus a Cantor set), Bavard [5] constructed an *intrinsically infinite-type* mapping class that is loxodromic with respect to the action of $\mathrm{Map}(\Sigma)$ on $\mathcal{A}(\Sigma, p)$, and for several years, this was the only known such example. In this paper, we give a new construction of mapping classes that are loxodromic with respect to the action of $\mathrm{Map}(\Sigma, p)$ on the relative arc graph $\mathcal{A}(\Sigma, p)$ for all infinite-type surfaces with an isolated puncture (outside of two sporadic classes). This class of surfaces is uncountable.

Theorem 1.1 *For any admissible surface Σ, there is an infinite family of intrinsically infinite-type homeomorphisms $\{g_n\}_{n\in\mathbb{N}}$ in $\mathrm{Map}(\Sigma, p)$ such that each g_n is loxodromic with respect to the action of $\mathrm{Map}(\Sigma, p)$ on $\mathcal{A}(\Sigma, p)$.*

In addition, we explore other dynamical and geometric properties of these mapping classes by demonstrating the convergence of a simple closed curve to a geodesic lamination on Σ under iterates of the maps and constructing an infinite-dimensional space of quasimorphisms of $\mathrm{Map}(\Sigma, p)$ using these elements (see below).

Each mapping class in our construction is the composition of three homeomorphisms called *shift maps*. Shift maps are generalizations of the *handleshift* homeomorphisms constructed by the third author and N Vlamis in [22] (see Figure 1 for examples of both). Roughly, an infinite-type surface Σ with an isolated puncture p is admissible if there is a proper embedding of the biinfinite flute surface containing p into Σ such that certain shift maps on the flute surface induce shift maps on Σ. Such an embedding allows us to reduce the proof of Theorem 1.1 to the case of the biinfinite flute surface. See Section 2.4 for more

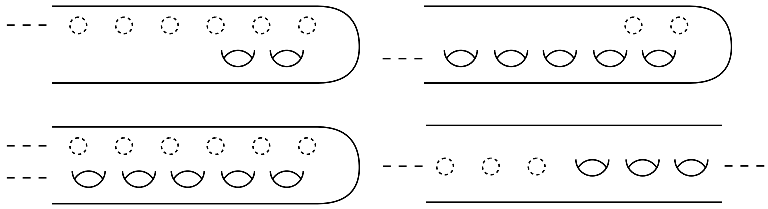


Figure 2: Examples of sporadic infinite-type surfaces that are not admissible. The first is a flute with finite genus, the other three are fluted Loch Ness monster surfaces.

details and Figure 3 for some examples of admissible surfaces. In Lemma 2.7, we show that a surface with an isolated puncture is admissible if and only if it admits shift maps. This set of surfaces consists of all infinite-type surfaces with an isolated puncture except a flute surface with finite (possibly zero) genus and a *fluted Loch Ness monster*. We call these two classes *sporadic surfaces* in this context. See Figure 2 for examples of sporadic surfaces and Lemma 2.9 for a proof of this fact. Since sporadic surfaces are exactly the small class of surfaces with an isolated puncture that do not admit shift maps, different methods will need to be developed in order to prove an analog of Theorem 1.1 for these surfaces. It would be interesting to understand how elements of the mapping class groups of sporadic surfaces act on the relative arc graph.

The handleshift homeomorphisms mentioned above have proven to be crucial in understanding various aspects of infinite-type mapping class groups. For example, it is shown in [22] that they are needed to topologically generate the *pure mapping class group* whenever Σ has at least two nonplanar ends, and in [3] Aramayona, Patel, and Vlamis showed that they are used to show that the pure mapping class groups of such surfaces surject onto $\mathbb{Z}$. With this paper, we emphasize the importance of more general shift maps to the theory of infinite-type mapping class groups. Inspired by Bavard's work in [5], we choose the shift maps in our construction carefully so that their composition mimics some of the behavior of pseudo-Anosov maps in the finite-type setting. In fact, we show that there is an alternative description of our homeomorphisms as the composition of a pseudo-Anosov homeomorphism on a finite-type subsurface and a *standard shift map* on Σ in Theorem 8.5. Additionally, in Section 10, we use the work of D Šarić [26] to prove the following theorem regarding geodesic laminations for the mapping classes constructed in Theorem 1.1.

Theorem 1.2 *If Σ is an admissible surface equipped with its conformal hyperbolic metric that is equal to its convex core, then there exists a simple closed curve c_0 on Σ such that the sequence $(g_n^i(c_0))_{i\in\mathbb{N}}$ converges to a geodesic lamination on Σ.*

In particular, we produce a train track on Σ and show the geodesic lamination from this theorem is weakly carried by this train track.

We emphasize that the elements arising from our construction are of *intrinsically infinite-type*, that is, they do not lie in the closure of the compactly supported mapping class group $\overline{\mathrm{Map}_c(\Sigma)}$, where the closure is taken with respect to the compact-open topology on $\mathrm{Map}(\Sigma)$. These are the first examples of intrinsically

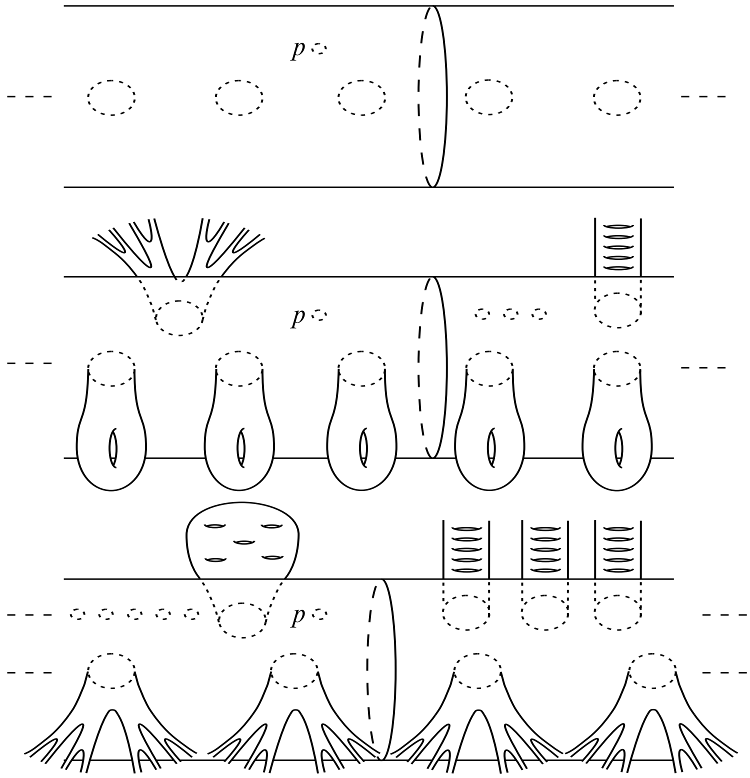


Figure 3: Examples of admissible surfaces, the first of which is the biinfinite flute surface S itself.

infinite-type loxodromic isometries for all admissible surfaces outside of the plane minus a Cantor set. Additionally, we emphasize that our construction does not rely on, and is not a generalization of, one of the few known methods for constructing pseudo-Anosov mapping classes for finite-type surfaces.

The most obvious candidates for mapping classes that are loxodromic with respect to the action on $\mathcal{A}(\Sigma, p)$ are those that are pseudo-Anosov on a finite-type subsurface $\Sigma' \subset \Sigma$ containing the special puncture p, that extend via the identity map to the rest of Σ (these are compactly supported mapping classes). In [7], Bavard and Walker prove that these types of mapping classes do indeed act loxodromically on a graph that is quasi-isometric to $\mathcal{A}(\Sigma, p)$. In that paper they point out that, though their class of examples is interesting, it will be even more important to construct mapping classes of intrinsically infinite-type that act loxodromically on $\mathcal{A}(\Sigma, p)$; this remark was one of the main points of inspiration for writing this paper. The intrinsically infinite-type elements of $\mathrm{Map}(\Sigma)$ are more mysterious since tools from finite-type surface theory do not directly generalize when studying these elements.

Remark 1.3 Morales and Valdez [20], building off of the work of Hooper [16], have also produced noncompactly supported elements that are loxodromic, but their elements are in the closure of the compactly supported mapping class group. Their method is a generalization of the Thurston–Veech construction of pseudo-Anosovs in the finite-type setting.

Aside from the motivation provided by a Nielsen–Thurston classification for infinite-type mapping classes, Bestvina and Fujiwara [10] show that constructing elements of $\mathrm{Map}(\Sigma)$ that act loxodromically on hyperbolic graphs can be used to understand the second bounded cohomology $H_b^2(\mathrm{Map}(\Sigma), \mathbb{R})$ of $\mathrm{Map}(\Sigma)$. In particular, they show that, for a compact surface Σ, there exist elements acting loxodromically on $\mathcal{C}(\Sigma)$ that are *weakly properly discontinuous* (WPD). These elements are used to prove that the space of quasimorphisms of $\mathrm{Map}(\Sigma)$ is infinite-dimensional, which is sufficient to conclude that $H_b^2(\mathrm{Map}(\Sigma), \mathbb{R})$ is, as well.

Along these lines, M Bestvina asked the following question at the AIM workshop on infinite-type surfaces [1, Problem 4.7]: "For $\Sigma = \mathbb{R}^2 - C$ (where C is a Cantor set) is it true that every subgroup of $\mathrm{Map}(\Sigma)$ has either infinite-dimensional space of quasimorphisms or is amenable?"[1] More generally, we would like to characterize the infinite-type mapping classes that can be used to produce quasimorphisms of $\mathrm{Map}(\Sigma)$. In Section 9, we show that the elements constructed in Theorem 1.1 can be used to give a new proof of the following theorem, originally due to Bavard [5] in the case of a plane minus a Cantor set and Bavard and Walker [7] in the general case.

Theorem 1.4 *Let Σ be an admissible surface. The space of nontrivial quasimorphisms on* $\mathrm{Map}(\Sigma, p)$ *is infinite-dimensional.*

When Σ is infinite type, $\mathrm{Map}(\Sigma)$ does not contain WPD elements; see the demonstration by Bavard and Genevois [6]. Despite this, we are still able to build nontrivial quasimorphisms using a weaker condition on loxodromic elements introduced by Bestvina and Fujiwara [10], called being antialigned, and an approach similar to that of Bavard in [5] which involves defining an intersection pairings on a specific class of arcs on Σ. In [8], Bavard and Walker use the same condition of being antialigned to show that homeomorphisms that are pseudo-Anosov on finite-type subsurfaces Σ' and extend via the identity to the rest of Σ can be used to produce quasimorphisms of $\mathrm{Map}(\Sigma)$. There is also a weaker version of the WPD condition, called WWPD, which was introduced by Bestvina, Bromberg, and Fujiwara in [9]; WWPD elements are always antialigned. A Rasmussen shows in [24] that for a surface Σ with an isolated puncture p, an element of $\mathrm{Map}(\Sigma, p)$ is WWPD with respect to the action on $\mathcal{A}(\Sigma, p)$ if and only if it stabilizes a finite-type subsurface Σ' containing the puncture p and restricts to a pseudo-Anosov on Σ'. In particular, the elements used by Bavard and Walker to construct nontrivial quasimorphisms are WWPD elements. The elements we construct in Theorem 1.1 do not fix any finite-type subsurface and thus are *not* WWPD. Our construction gives subgroups of $\mathrm{Map}(\Sigma)$ that do not contain WWPD elements but do have an infinite-dimensional space of quasimorphisms.

Plan of the paper In order to prove Theorem 1.1, we explicitly compute the images of a particular arc on Σ under iterates of each homeomorphism g_n and prove that these images form a quasigeodesic axis for the action of $\langle g_n \rangle$ on $\mathcal{A}(\Sigma, p)$. Though some of the methods in our paper are inspired by Bavard's work in [5], we note that there are a variety of additional challenges in proving Theorem 1.1 for such a

[1] This question has since been answered by Fournier-Facio, Lodha, and Zaremsky [14].

wide class of surfaces. In fact, we first prove the theorem for the biinfinite flute surface S and then use the fact that the inclusion of $\mathcal{A}(S, p)$ into $\mathcal{A}(\Sigma, p)$ is a $(2,0)$-quasi-isometric embedding (see Lemma 2.10) to extend the theorem to all admissible surfaces. One of the first challenges in proving Theorem 1.1 is rigorously coding arcs on S, which we do in Section 3.1, in order to quantify how long two arcs on S fellow travel. We note that the coding of our arcs is done in way that is "geometrically intuitive". What we mean is that, given the code for an arc, one can easily draw the corresponding arc on the surface. For other ways to code arcs see [5; 7]. We then introduce *standard position* for an arc on S in Section 3.2 so that we can use the code for an arc to find its image under our shift maps in a well-defined way. Most importantly, we must understand when segments of arcs become trivial under our shift maps, and in Section 4.1 we introduce a kind of cancellation in the image of the code for a segment which we call *cascading cancellation*. This kind of cancellation will cause technical problems throughout the paper and much of Section 6 is devoted to understanding how to control it.

The rest of Section 4 is devoted to proving Theorem 4.8, which answers the question of when a segment in an arc becomes trivial under our shift maps. We define the homeomorphisms g_n of Theorem 1.1 in Section 5, show that we have *"starts like" functions* in Section 6, and show that we have *highways* in Section 7. Finally, we prove Theorem 1.1 in Section 8, introduce an intersection pairing for arcs and prove Theorem 1.4 in Section 9, and prove the convergence to a geodesic lamination from Theorem 1.2 in Section 10.

Acknowledgements The authors are grateful to Chris Leininger for pointing out the work of Verberne and to Ekta Patel for helping us with computational work related to this project. The authors would also like to thank Alden Walker and Nick Vlamis for helpful discussions as well as Elizabeth Field for helpful comments on an earlier draft. Abbott was partially supported by NSF DMS–1803368 and DMS–2106906, Miller was partially supported by NSF DMS–2005438 and DMS–2300370, and Patel was partially supported by NSF DMS–1840190 and DMS–2046889. The authors would also like to thank the American Institute of Mathematics and the organizers of the *Infinite-type surfaces* workshop that was held there, during which some of this work was completed.

2 Background

2.1 Space of ends and classification of infinite-type surfaces

Central to the classification of infinite-type surfaces is the definition of the space of ends $E(\Sigma)$ of an infinite-type surface Σ. Informally, an end of Σ is a way to escape or go off to infinity in Σ. More formally we have:

Definition 2.1 An *exiting sequence* in Σ is a sequence $\{U_n\}_{n\in\mathbb{N}}$ of connected open subsets of Σ satisfying

(1) $U_n \subset U_m$ whenever $m < n$;

(2) U_n is not relatively compact for any $n \in \mathbb{N}$, that is, the closure of U_n in Σ is not compact;

(3) the boundary of U_n is compact for each $n \in \mathbb{N}$; and

(4) any relatively compact subset of Σ is disjoint from all but finitely many U_n.

Two exiting sequences $\{U_n\}_{n\in\mathbb{N}}$ and $\{V_n\}_{n\in\mathbb{N}}$ are equivalent if for every $n \in \mathbb{N}$ there exists $m \in \mathbb{N}$ such that $U_m \subset V_n$ and $V_m \subset U_n$. An *end* of Σ is an equivalence class of exiting sequences.

The *space of ends* $E(\Sigma)$, or simply E, of Σ is the set of ends of Σ equipped with a natural topology for which it is totally disconnected, Hausdorff, second countable, and compact. In particular, $E(\Sigma)$ is homeomorphic to a closed subset of the Cantor set. To describe the topology, let V be an open subset of Σ with compact boundary, define $\widehat{V} = \{[\{U_n\}_{n\in\mathbb{N}}] \in E : U_n \subset V \text{ for some } n \in \mathbb{N}\}$ and let $\mathcal{V} = \{\widehat{V} : V \subset \Sigma \text{ is open with compact boundary}\}$. The set E becomes a topological space by declaring $\mathcal{V}$ a basis for the topology.

We note that ends can be isolated or not and can be *planar* (if there exists an i such that U_i is homeomorphic to an open subset of the plane $\mathbb{R}^2$) or *nonplanar* (if every U_i has infinite genus). The set of nonplanar ends of Σ is a closed subspace of $E(\Sigma)$ and will be denoted by $E_g(\Sigma)$.

Kerékjártó [17] and Richards [25] showed that the homeomorphism type of an orientable infinite-type surface is determined by the quadruple

$$(g, b, E_g(\Sigma), E(\Sigma)),$$

where $g \in \mathbb{Z}_{\geq 0} \cup \{\infty\}$ is the genus of Σ and $b \in \mathbb{Z}_{\geq 0}$ is the number of (compact) boundary components of Σ.

Of particular interest to us is the infinite-type surface called the *biinfinite flute* obtained from an infinite cylinder by deleting a countable discrete sequence of points exiting both ends of the cylinder (see Figure 3). By the classification theorem of Kerékjártó and Richards, this surface can also be obtained from $\mathbb{S}^2$ by deleting $\{x_i\}$, $\{y_i\}$, x, and y, where $\{x_i\}$ and $\{y_i\}$ are countable discrete sequences of points converging to distinct points x and y, respectively. Note that S has two special nonisolated ends.

2.2 Mapping class groups and arc graphs

The *mapping class group*, $\mathrm{Map}(\Sigma)$, of a surface Σ is the group of orientation-preserving homeomorphisms of Σ up to isotopy. The natural topology on any group of homeomorphisms is the compact-open topology and $\mathrm{Map}(\Sigma)$ is endowed with the quotient topology with respect to the compact-open topology on the space of homeomorphisms of Σ. When Σ is a finite-type surface, this topology agrees with the discrete topology on $\mathrm{Map}(\Sigma)$, but when Σ is of infinite type it does not. There are several important subgroups of $\mathrm{Map}(\Sigma)$: $\mathrm{Map}_c(\Sigma)$ is the subgroup consisting of mapping classes with compact support, $\mathrm{PMap}(\Sigma)$ is the pure mapping class group consisting of mapping classes which fix the set of ends pointwise,

$\overline{\mathrm{Map}_c(\Sigma)} < \mathrm{PMap}(\Sigma)$ is the closure of the compactly supported mapping class group with respect to the topology described above, and when Σ has an isolated puncture p, $\mathrm{Map}(\Sigma, p)$ is the subgroup of mapping classes that fix p.

When Σ is finite-type, $\mathrm{Map}(\Sigma)$ is algebraically generated by finitely many Dehn twists [19]. Infinite-type mapping class groups, sometimes called big mapping class groups, are uncountable groups, so there is no countable algebraic generating set. However, one can consider topological generating sets (countable dense subsets of $\mathrm{Map}(\Sigma)$) and in [22], Vlamis and the third author prove that for many infinite-type surfaces, Dehn twists are not sufficient in topologically generating even $\mathrm{PMap}(\Sigma)$. They show that in addition to Dehn twists, a new class of homeomorphisms called *handleshifts* (defined in Section 2.4) are often needed to topologically generate $\mathrm{PMap}(\Sigma)$. In a subsequent paper with Aramayona [3], Vlamis and the third author give an algebraic description of $\mathrm{PMap}(\Sigma)$ that will be relevant in Section 8. When Σ is an infinite-type surface with $n > 1$ nonplanar ends, they prove that $\mathrm{PMap}(\Sigma) = \overline{\mathrm{Map}_c(\Sigma)} \rtimes \mathbb{Z}^{n-1}$, where $\mathbb{Z}^{n-1}$ is generated by $n-1$ handleshifts with disjoint support. In particular, when Σ has exactly 2 nonplanar ends (for example when Σ is the ladder surface), $\mathrm{PMap}(\Sigma) = \overline{\mathrm{Map}_c(\Sigma)} \rtimes \mathbb{Z}$, where $\mathbb{Z} = \langle H \rangle$ and H is the standard handleshift, shifting each genus of Σ over to the right by one.

In this paper we are primarily concerned with mapping classes of intrinsically infinite-type.

Definition 2.2 An element $f \in \mathrm{Map}(\Sigma)$ is of *intrinsically infinite-type* if $f \notin \overline{\mathrm{Map}_c(\Sigma)}$.

More specifically, we are interested in how such elements act on a particular graph of arcs called the *relative arc graph*.

Let Σ be a connected, orientable surface with empty boundary, and let $\Pi \subset \Sigma$ be the set of punctures of Σ, which we assume to be nonempty. In this subsection, it is convenient to regard Π as a set of marked points on Σ. By a *proper arc* on Σ we mean a map $\alpha\colon [0,1] \to \Sigma$ such that $\alpha^{-1}(\Pi) = \{0,1\}$. We often conflate an arc with its image in Σ. An arc is *simple* if it is an embedding when restricted to the open interval $(0,1)$.

The *arc graph* $\mathcal{A}(\Sigma)$ is the simplicial graph whose vertices are isotopy classes of simple arcs on Σ, where we only consider isotopies rel endpoints, and two (isotopy classes of) arcs are connected by an edge if they can be realized disjointly away from Π. The mapping class group $\mathrm{Map}(\Sigma)$ acts on $\mathcal{A}(\Sigma)$ by isometries. Hensel, Przytycki, and Webb [15] show that when Σ has finite-type, the graph $\mathcal{A}(\Sigma)$ is infinite diameter and 7-hyperbolic. On the other hand, when Σ is infinite-type with infinitely many punctures, it is straightforward to see that $\mathcal{A}(\Sigma)$ has diameter 2, and so this graph is not particularly useful for studying $\mathrm{Map}(\Sigma)$.

Assuming that Π contains a nonempty set of isolated punctures, Aramayona, Fossas, and Parlier [2] construct a particular subgraph of the arc graph which has interesting geometry, even when Π is infinite. We are interested in a special case of this construction, involving a single isolated puncture p.

Definition 2.3 The *relative arc graph* $\mathcal{A}(\Sigma, p)$ is the subgraph of $\mathcal{A}(\Sigma)$ spanned by arcs which start and end at p. More precisely, the vertices of $\mathcal{A}(\Sigma, p)$ are isotopy classes of arcs on Σ with endpoints on p, where we allow only isotopy rel endpoints. There is an edge between two (isotopy classes of) arcs if they can be realized disjointly away from p.

Aramayona, Fossas, and Parlier show that $\mathcal{A}(\Sigma, p)$ is connected, has infinite diameter, and is 7-hyperbolic (see [2, Theorem 1.1]). While $\mathrm{Map}(\Sigma)$ does not necessarily act on $\mathcal{A}(\Sigma, p)$, the subgroup $\mathrm{Map}(\Sigma, p)$ that fixes the puncture p does act by isometries on this graph. When Σ has only one isolated puncture p, $\mathrm{Map}(\Sigma) = \mathrm{Map}(\Sigma, p)$.

2.3 Metric spaces and loxodromic isometries

We now introduce some basics of metric spaces and isometries of a hyperbolic metric space. Given a metric space X, we denote by d_X the distance function on X. A map $f\colon X \to Y$ between metric spaces X and Y is a (K, C)*-quasi-isometric embedding* if there is are constants $K \geq 1$, $C \geq 0$ such that for all $x, y \in X$,

$$\frac{1}{K} d_X(x, y) - C \leq d_Y(f(x), f(y)) \leq K d_X(x, y) + C.$$

A *geodesic* in X is an isometric embedding of an interval into X and a (K, C)*-quasigeodesic* in X is a (K, C)-quasi-isometric embedding of an interval into X. We call the constants K, C the *quality* of the quasigeodesic. By an abuse of notation, we often conflate a (quasi)geodesic and its image in X.

Definition 2.4 Given an action by isometries of a group G on a hyperbolic space X, an element $g \in G$ is *elliptic* if it has bounded orbits; *loxodromic* if the map $\mathbb{Z} \to X$ given by $n \mapsto g^n x_0$ for some (equivalently, any) $x_0 \in X$ is a quasi-isometric embedding; and *parabolic* otherwise.

Any biinfinite quasigeodesic in X which is preserved by a loxodromic isometry $g \in G$ is called an *axis* of g. An axis always exists; for any $x_0 \in X$, the set $\{g^n x_0 \mid n \in \mathbb{Z}\}$ is a (discrete) quasigeodesic preserved by g. If X is a geodesic metric space, in the sense that there exists a geodesic connecting any two points of X, then we may construct a continuous quasigeodesic axis as follows. Fix a geodesic $[x_0, gx_0]$ from x_0 to gx_0. Then g stabilizes the path formed by concatenating the geodesics $g^n[x_0, gx_0]$; this path is a quasigeodesic axis of g in X. Varying the point x_0 will change the quality of the quasigeodesic. Let $g^+ = \lim_{n\to\infty} g^n x_0$ and $g^- = \lim_{n\to-\infty} g^n x_0$ be points in the Gromov boundary ∂X of X. The *limit set* of $\langle g\rangle$ is the subset $\{g^+, g^-\} \subseteq \partial X$; this set is fixed pointwise by g. It is straightforward to show that the limit set $\{g^+, g^-\}$ does not depend on the choice of $x_0 \in X$.

2.4 Shift maps and the biinfinite flute surface

A handleshift was first defined in [22] as follows. Consider the surface S' defined by taking the strip $\mathbb{R} \times [-1, 1]$, removing a disk of radius $\frac{1}{2}$ with center $(n, 0)$ for each $n \in \mathbb{Z}$, and attaching a torus with one boundary component to the boundary of each such disk. A handleshift on S' is the homeomorphism that acts like a translation, sending (x, y) in S to $(x + 1, y)$ and which tapers to the identity on $\partial S'$. Given a surface of infinite-genus Σ with at least two nonplanar ends and a proper embedding of S' into

Σ so that the two ends of the strip correspond to two distinct ends of Σ, the handleshift on S' induces a handleshift on Σ, where the homeomorphism acts as the identity on the complement of S'. In this paper, more flexibility is allowed, and we define the following generalization.

Definition 2.5 Let S' be the surface defined by taking the strip $\mathbb{R} \times [-1, 1]$, removing a closed disk of radius $\frac{1}{4}$ with center $(n, 0)$ for $n \in \mathbb{Z}$, and attaching any fixed topologically nontrivial surface with exactly one boundary component to the boundary of each such disk. A *shift* on S' is the homeomorphism that acts like a translation, sending (x, y) in S' to $(x + 1, y)$ and which tapers to the identity on $\partial S'$.

Lanier and Loving use two particular cases of this generalization in [18]. Naming the full generalization a "shift" is in line with their paper. Note that it is essential for the same surface to be glued to the boundary component of each disk in order for the shift to be a homeomorphism of the surface.

As above, given a surface Σ with a proper embedding of S' into Σ so that the two ends of the strip correspond to two different ends of Σ, the shift on S' induces a shift on Σ, where the homeomorphism acts as the identity on the complement of S'. Given a shift h on Σ, the embedded copy of S' in Σ is called the *domain* of h. In this paper, we produce special homeomorphisms that can be obtained as a composition of three shift maps on such a surface Σ with an isolated puncture p and that are loxodromic with respect to the action of $\mathrm{Map}(\Sigma, p)$ on $\mathcal{A}(\Sigma, p)$. Instead of working generally with surfaces that admit shift maps, we begin by letting S be the biinfinite flute surface. Then, S admits shift maps which shift a countable collection of punctures on S. To prove Theorem 1.1 we first construct mapping classes that are loxodromic with respect to the action of $\mathrm{Map}(S, p)$ on $\mathcal{A}(S, p)$. We then use this surface as a template for constructing the desired mapping classes for more general surfaces Σ by extending the shift maps on S to shift maps on Σ as follows.

Definition 2.6 Let S be the biinfinite flute surface. A surface Σ with an isolated puncture p is *admissible* if there exists a proper embedding $S \hookrightarrow \Sigma$ where S contains p, the two nonisolated ends of S correspond to distinct ends of Σ, and such that a countably infinite collection of connected components of $\Sigma \setminus S$ are of the same (nontrivial) topological type. Note that when the components are once-punctured disks, there are countably many isolated punctures of S that remain isolated punctures when embedded in Σ. Denote this special class of connected components of $\Sigma \setminus S$ by $\mathcal{U}$, so that the elements of $\mathcal{U}$ are all homeomorphic to a fixed surface Σ_0 with one boundary component. See Figure 1 for some examples of admissible surfaces.

Given a shift map h on S, the support of h is a strip $\mathbb{R} \times [-1, 1]$ with countably many punctures. When the set of punctures in the support of h only consists of those corresponding to elements of $\mathcal{U}$, we can glue copies of Σ_0 onto the punctures of this strip to produce a shift map on a surface S' as in Definition 2.5. The embedding of S in Σ therefore gives an embedding of S' in Σ and the shift on S' in Σ is extended via the identity on $\Sigma \setminus S'$ as usual. From this construction, we immediately have one direction of the following lemma.

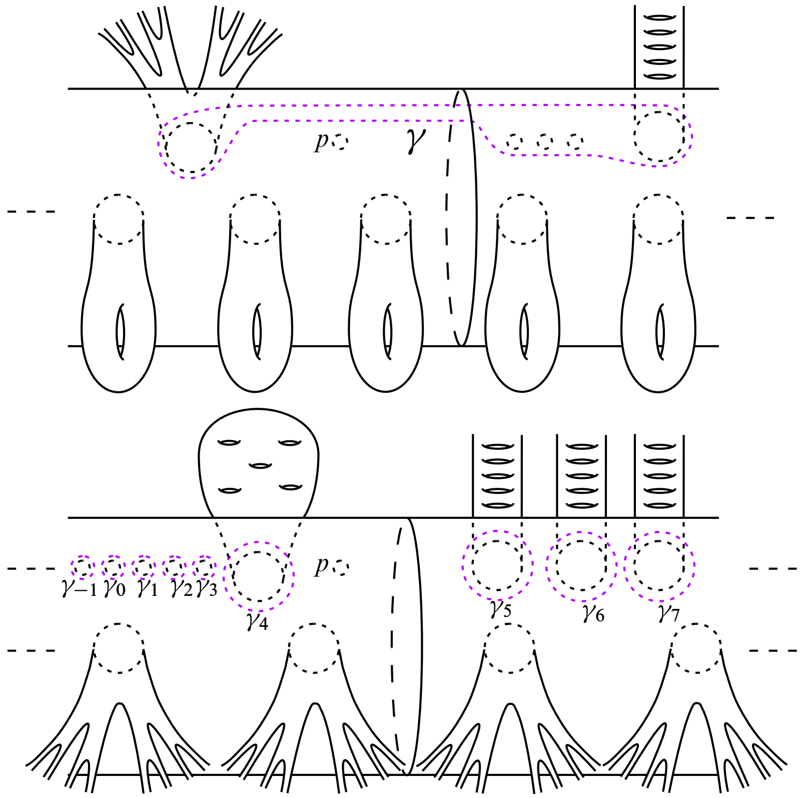


Figure 4: Examples showing that surfaces with shift maps are always admissible. For the first surface, only one curve γ is needed to cut away the extra topology of Σ. In the second case, a countable collection of curves $\{\gamma_i\}$ is needed.

Lemma 2.7 *Given a surface Σ with an isolated puncture, Σ is admissible if and only if Σ admits shift maps.*

Proof It is left to show that if Σ has an isolated puncture p and admits a shift map, then Σ is admissible. To see this, we consider the proper embedding of S' into Σ. Recall that S' is obtained from a punctured strip by gluing on countably many copies of any surface Σ_0 with exactly one boundary component. Let $\mathcal{T} = \{T_i\}$ denote the corresponding countable collection of subsurfaces homeomorphic to Σ_0 in Σ, indexed by $\mathbb{Z}$.

Note that $E(S')$ is a closed subset of $E(\Sigma)$, as is $X = E(S') \cup \{p\}$, and thus $X^c = E(\Sigma) \setminus X$ is open in $E(\Sigma)$. The second countability of the topology on $E(\Sigma)$ implies that X^c is the union of countably many basis elements. If X^c is in fact clopen in $E(\Sigma)$, then X^c is compact and is therefore a finite union of basis elements. In this case, there exists a simple closed curve γ in Σ with the following property: there exists a connected component K of $\Sigma \setminus \gamma$ such that the end space of $\overline{K} = K \cup \gamma$ is exactly X^c. In this way, γ cuts away the ends of Σ that are in X^c (see Figure 4). We then have that

$$\Sigma \setminus \big(\overline{K} \cup \big(\textstyle\bigcup_i T_i\big)\big)$$

is homeomorphic to the biinfinite flute surface and Σ is admissible with $\mathcal{T}$ playing the role of $\mathcal{U}$ in the definition of an admissible surface.

In general, we can only assume that X^c is open, not clopen, so that it can be expressed as the union of countably many basis elements for the topology. Then, there exists a countable collection of simple closed curves $\{\gamma_i\}$ and a countable collection of connected components K_i of $\Sigma \setminus \gamma_i$ such that the end space of $\bigcup_i \overline{K}_i = \bigcup_i K_i \cup \gamma_i$ is exactly X^c (see Figure 4). In this case,

$$\Sigma \setminus \big(\big(\bigcup_i \overline{K}_i\big) \cup \big(\bigcup_i T_i\big)\big)$$

is homeomorphic to the biinfinite flute surface and Σ is admissible. □

Given this equivalent definition for an admissible surface, we can show that this class includes all infinite-type surfaces with an isolated puncture outside of two sporadic classes that do not admit shift maps. We will need the following definition.

Definition 2.8 The *Loch Ness monster* is the infinite-type surface with no planar ends and exactly one nonplanar end. An infinite-type surface is a *fluted Loch Ness monster* if it is obtained from the Loch Ness monster in one of the two following ways:

(1) by deleting a finite, nonzero collection of isolated points, or

(2) deleting a countably infinite collection of isolated points accumulating to exactly one point, which we also delete from the surface, or accumulating onto the end of the Loch Ness monster.

See Figure 2 for examples of fluted Loch Ness monsters.

Lemma 2.9 *Let Σ be an infinite-type surface with an isolated puncture. Then Σ is admissible unless Σ is a flute surface with finite (possibly zero) genus or is a fluted Loch Ness monster surface.*

Proof Let Σ be an infinite-type surface with an isolated puncture p. If Σ has at least two nonplanar ends, then Σ admits a shift map (in fact a handleshift). Similarly, if Σ has at least two nonisolated planar ends, then Σ admits a shift map with these two ends corresponding to the two ends of the strip S' in Definition 2.5. Thus, if Σ does not admit a shift map, Σ has exactly one nonisolated planar end and finite genus, ie a flute surface with finite genus, or has exactly one nonplanar end and up to one nonisolated planar end, ie a fluted Loch Ness monster. □

Going back to the original definition of an admissible surface, there are a few more notable remarks regarding the relationship between S and Σ. First, there is not necessarily an embedding of $\mathrm{Map}(S, p)$ into $\mathrm{Map}(\Sigma, p)$ since if the support of a shift h on S contains punctures that do not correspond to elements of $\mathcal{U}$, then there may not be a way to extend that shift to Σ. In particular, if h shifts one puncture x to another puncture x' but the topology of the surfaces glued to x and x' are different, there is no extension of h to a shift of Σ. This will not affect our arguments since there are countably many punctures of S corresponding to the elements of $\mathcal{U}$ which we move to the *front* of the cylinder for S along with p, and we move all other punctures to the *back* of the cylinder. Here we are choosing one nonisolated end of S to correspond to the left direction on the surface, and the other nonisolated end to correspond to moving right on the surface so that there is a well-defined notion of the front and back of S. In our constructions, we use shift maps on S whose support only contains the punctures on the front of S so that all of these shifts extend to Σ.

Second, and most importantly, we now show that proving Theorem 1.1 for admissible surfaces Σ can be reduced to the case of the biinfinite flute S. In fact, this is the motivation for the original definition of an admissible surface. For simplicity, given a surface M with an isolated puncture p and any points $a, b \in \mathcal{A}(M, p)$, we write $d_M(a, b)$ for the distance between a and b in $\mathcal{A}(M, p)$

Lemma 2.10 *Let Σ be an admissible infinite-type surface with an isolated puncture p. Then the inclusion of $\mathcal{A}(S, p)$ into $\mathcal{A}(\Sigma, p)$ is a $(2,0)$-quasi-isometric embedding.*

Proof As $S \subset \Sigma$, it is clear that $d_\Sigma(a, b) \le d_S(a, b)$ for any $a, b \in \mathcal{A}(S, p)$.

To obtain the other inequality, let $S_{a,b} \subset S$ be a finite-type subsurface of S which contains a, b, the puncture p, and has complexity at least 2. Note that $S_{a,b}$ is then a finite-type subsurface of Σ as well. Thus by [2, Corollary 4.3] applied to $S_{a,b} \subset \Sigma$ and to $S_{a,b} \subset S$, we have

$$d_S(a, b) \le d_{S_{a,b}}(a, b) \le 2 d_\Sigma(a, b).$$

Together, these imply that

$$d_\Sigma(a, b) \le d_S(a, b) \le 2 d_\Sigma(a, b),$$

completing the proof. □

In particular, let $g \in \mathrm{Map}(S, p)$ be loxodromic with respect to the action of $\mathrm{Map}(S, p)$ on $\mathcal{A}(S, p)$ with a (K, C)-quasigeodesic axis. If g can be extended to an element of $\mathrm{Map}(\Sigma, p)$, then this extension is loxodromic with respect to the action of $\mathrm{Map}(\Sigma, p)$ on $\mathcal{A}(\Sigma, p)$, and the extension will have a $(2K, C)$-quasigeodesic axis.

3 Coding arcs and standard position

Let S be the biinfinite flute surface with a distinguished isolated puncture p, and let $\{p_i\}_{i \in \mathbb{Z}}$ be any countably infinite discrete collection of punctures on S which exits both ends of the cylinder and does not contain p. As described in Section 2.4, we choose one nonisolated end of S to correspond to the left direction and one to correspond to the right direction, which gives a well-defined notion of a front and back of the cylinder for S. We move all of the punctures in $\{p_i \mid i \in \mathbb{Z}\} \cup \{p\}$ to the front of the cylinder for S and all other punctures to the back. We also move the distinguished puncture p so that it lies to the right of p_{-1} and to the left of p_0. We will consider the collection $\{p_i \mid i \in \mathbb{Z}\} \cup \{p\}$ of punctures. We index this set with $\mathbb{Z} \cup \{P\}$, which we give the ordering consisting of the usual ordering on $\mathbb{Z}$ with the additional requirement that $-1 < P < 0$. The index P corresponds to the distinguished puncture p.

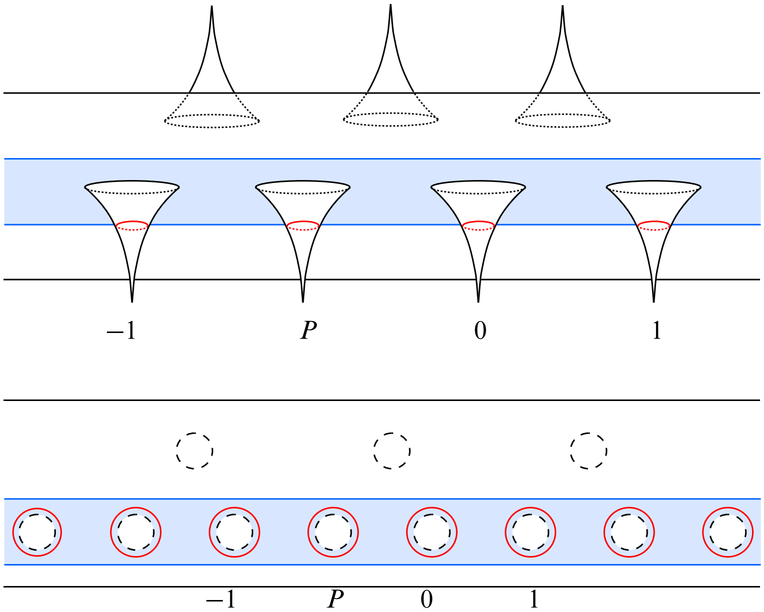


Figure 5: The curves B_i are in red. The blue region is the domain of the shift map H.

Fix the simple closed curve B_0 bounding the puncture p_0 on S shown in Figure 5. More formally, to define B_0 we fix a complete hyperbolic metric on S and let B_0 be a horocycle at a height sufficiently far out the cusp. Fix a shift map H on S whose domain contains exactly the collection $\{p_i\}$ for $i \in \mathbb{Z} \cup P$ and which shifts p_i to p_{i+1} for all $i \in \mathbb{Z} \cup P$.

Definition 3.1 Define the simple closed curves $B_i := H^i B_0$ for $i \in \mathbb{Z} \cup P$. Then B_i is a simple closed curve bounding the puncture p_i, where $p_P = p$.

Our choice of left/right also gives a well-defined notion of an arc passing *over* or *under* a puncture (or equivalently some B_i). In all pictures of S throughout the paper, we denote the special puncture p by an "X", and rather than drawing the punctures p_i, we draw the simple closed curves B_i in S. We will use these simple closed curves to put arcs into *standard position* as described later in this section.

3.1 Coding arcs

We use the simple closed curves B_i to describe a way to code simple arcs on S starting and ending at p. We will use this code to quantify how long two arcs fellow travel, which will be essential for proving the results of this paper.

Suppose that γ is an oriented arc on S starting and ending at p such that γ can be homotoped to be completely contained on the front of S. We code γ as follows. First homotope γ so that it is disjoint from all B_i with $i \in \mathbb{Z} \cup P$, with the exception that γ starts and ends at the puncture p and therefore intersects B_P exactly twice. The code for γ always starts and ends with the character P_s (which stands for "puncture start") and contains either the character k_o or the character k_u, where $k \in \mathbb{Z} \cup \{P\}$, whenever γ

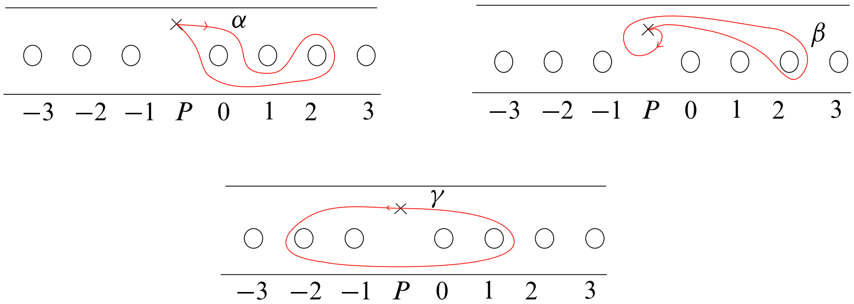


Figure 6: Pictured are arcs on the front of the surface S. The X denotes the puncture p and the elements $k \in \mathbb{Z} \cup \{P\}$ shown under S denote the subscript on the simple closed curves B_k. The codes for arcs α, β, γ are given in Example 3.2.

passes over or under the simple closed curve B_k for $k \in \mathbb{Z} \cup P$. These characters appear in the code for γ in the same order in which γ passes over/under the curves B_k. For example, since $-1 < P < 0$, the second character of the code for γ must be either 0_o, 0_u, $(-1)_o$, $(-1)_u$, P_o, or P_u, because if γ doesn't immediately wrap around p (which would lead to the second character being P_o or P_u), it must pass over or under either B_0 or B_{-1} before it can pass over or under B_k for any $k \neq 0, -1$. Similarly, if the character 2_o or 2_u appears in the code, each adjacent character must be one of $1_o, 1_u, 2_o, 2_u, 3_o$, or 3_u. To simplify notation, we write $k_{o/u}$ to mean that the character could be k_o or k_u. We will write $k_{o/u}k_{u/o}$ to mean that the two adjacent characters are either $k_o k_u$ or $k_u k_o$; the $k_{u/o}$ is used to emphasize that the second character has the opposite subscript as the first one.

Example 3.2 Consider the arcs shown in Figure 6. The elements $k \in \mathbb{Z} \cup \{P\}$ shown under S denote the subscript on the simple closed curves B_k. The code for α is $P_s 0_o 1_u 2_o 2_u 1_u 0_u P_s$, the code for β is $P_s P_u P_o 0_o 1_o 2_o 2_u 1_o 0_o P_s$, and the code for γ is $P_s(-1)_o(-2)_o(-2)_u(-1)_u P_u 0_u 1_u 1_o 0_o P_s$.

Now suppose γ is an oriented arc on S starting and ending at p such that no arc in its homotopy class is contained on the front of S. Since γ starts and ends at p, which is on the front of the surface, every time γ leaves the front of S it must eventually reenter the front. We give the code C to any subpath of γ which is on the back of S. Up to homotopy, we may assume that each time γ exits then enters the front of S, it does so "between" two simple closed curves B_k and B_{k+1}. In other words, there is an arc γ' in the homotopy class of γ whose code contains either $k_{o/u}C(k+1)_{o/u}$ or $k_{o/u}Ck_{o/u}$ each time γ' leaves the front of S. We give γ the same code as γ'. We emphasize that this implies that the code of an arc does not distinguish the behavior of arcs γ on the back of S.

By an abuse of notation, we typically blur the distinction between an arc and its code, writing, for example, $\alpha = P_s 0_o 1_u 2_o 2_u 1_u 0_u P_s$.

Definition 3.3 Let γ be an oriented arc on S starting and ending at p. A code for γ is *reduced* if no two adjacent characters in the code are the same and if the character immediately following the initial P_s or preceding the terminal P_s is not $P_{o/u}$.

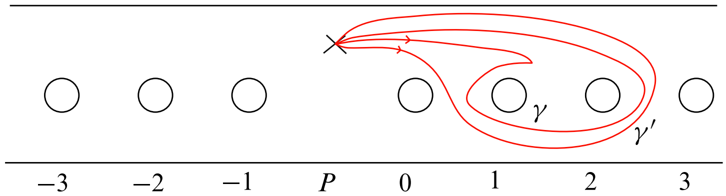


Figure 7: The arcs γ and γ' are homotopic and have the same reduced code: γ' is formed from γ by the removal of the pair $1_o 1_o$.

The appearance of repeated characters in the code of an arc indicates backtracking in the arc. The following lemma is immediate.

Lemma 3.4 *If there are two arc γ and γ', starting and ending at p, whose codes differ only by the removal of two adjacent characters which are equal, ie $k_o k_o$ or $k_u k_u$, then γ and γ' are homotopic.*

Example 3.5 The arcs γ and γ' in Figure 7, with codes $P_s 0_o 1_o 1_o 1_u 2_u 2_o 1_o 0_o P_s$ and $P_s 0_o 1_u 2_u 2_o 1_o 0_o P_s$, respectively, are homotopic.

Note that if a triple appears in the code for an arc, it is reduced to a single character according to our convention, as only *pairs* of repeated characters are removed. For example, $P_s 0_o 1_o 1_o 1_o 1_u 0_o P_s$ is reduced to $P_s 0_o 1_o 1_u 0_o P_s$.

Each homotopy class of curves on S determines a reduced code, in the sense that any two homotopic curves have the same reduced code. We write that two codes are equal if they determine homotopic arcs. For example, we write

$$P_s 0_o 1_o 1_o 1_u 2_u 1_u 0_u P_s = P_s 0_o 1_u 2_u 1_u 0_u P_s.$$

The converse of this fact is not true, however, because the code does not encode the behavior of arcs on the back of S; hence there can be nonhomotopic arcs with the same reduced code. This will not cause any problems in this paper.

Definition 3.6 The *code length* of an arc γ, denoted $\ell_c(\gamma)$, is the number of characters in a reduced code for γ.

Convention 3.7 When giving the code of an arc for which the numerical values of the characters are unimportant (or unknown), we will use variables in the code. Our convention is to use Roman letters to represent single characters and Greek letters to represent strings of characters whose length is (possibly) greater than one. For example, $\ell_c(a_1 a_2 a_3) = 3$ while $\ell_c(a\gamma b) = \ell_c(\gamma) + 2$.

Given a string of characters $\alpha = a_1 a_2 \dots a_n$, we denote by $\bar{\alpha}$ the reverse of α, so that $\bar{\alpha} = a_n a_{n-1} \dots a_2 a_1$. If α is an arc, then $\bar{\alpha}$ is the same arc with the opposite orientation.

3.2 Standard position

In this section, we describe how to use the code for an arc to find its image under a general class of shifts which we call "permissible".

Definition 3.8 We say a shift shifting to the right is a *right shift*, while a shift shifting to the left is a *left shift*. A right shift is *permissible* if its domain D stays on the front of our subsurface and contains a *turbulent region* (n_1, n_2), that is, there exist $n_1, n_2 \in \mathbb{Z} \cup \{P\}$ with $n_1 < n_2$ such that D contains B_k for all $k \in (-\infty, n_1] \cup [n_2, \infty)$ but does not contain B_k for any $k \in (n_1, n_2)$. We call $(-\infty, n_1) \cup [n_2, \infty)$ the *shift region* of h. See Figure 11. Analogously, a left shift is *permissible* if its domain D stays on the front of our subsurface and contains a *turbulent region* (n_2, n_1), that is, there exist $n_1, n_2 \in \mathbb{Z} \cup \{P\}$ with $n_2 < n_1$ such that D contains B_k for all $k \in (-\infty, n_2] \cup [n_1, \infty)$ but does not contain B_k for any $k \in (n_2, n_1)$. The shift region for a left shift is $(-\infty, n_2] \cup (n_1, \infty)$.

Convention 3.9 Throughout the paper, we will use both left and right shifts. For notational simplicity, all general results about shifts will be stated for right shifts. All statements of results, proofs, and figures will make this assumption as well. However, all of our definitions and results (and their proofs) also hold for left shifts, by modifying any proof for a right shift so that we essentially replace all instances of n_1 with n_2 and vice versa and replace all instances of the word "increasing" by the word "decreasing" and vice versa. The only subtleties are that:

- We retain the convention that $h(B_{n_1}) = B_{n_2}$.
- For the shift region intervals $(-\infty, n_1) \cup [n_2, \infty)$ that appear for a right shift, we use $(-\infty, n_2] \cup (n_1, \infty)$ for the left shift. In particular, the n_2 is always contained in the shift region.

Remark 3.10 It is worthwhile to mention that Convention 3.9 is equivalent to simply redefining the order, given by the symbol $<_{\mathrm{rev}}$, on $\mathbb{Z} \cup \{P\}$ to be the opposite of the standard meaning of the inequality sign $<$. For example, in this "reversed order" we would have $5 <_{\mathrm{rev}} 3$ and so on. Given this and using the standard meanings for "increasing" and "decreasing" with respect to $<_{\mathrm{rev}}$, all of the proofs for shifts that shift to the left would go through identically as shifts that shift to the right when one replaces each instance of $<$ with an $<_{\mathrm{rev}}$. Despite the simplicity of this reversed order, we found writing proofs with it to be more confusing to the reader than applying the above convention.

In order to find the image of an arc using only its code, we will need to consider paths whose endpoints are not on p.

Definition 3.11 A *segment* is a simple path with at least one endpoint which is not a puncture, and no endpoints on a puncture other than p. We code a segment in an analogous way as we did arcs in Section 3.1. If a segment begins or ends on p, then the initial or terminal character of the code is P_s, respectively.

Note that a segment can have at most one instance of P_s in its code. Given a segment γ, we denote the initial and terminal character of its code by γ^i and γ^t, respectively. A segment is *supported on* an interval $(a,b) \subset \mathbb{Z} \cup \{P\}$ if the numerical value of every character in its reduced code is contained in (a,b). A subsegment of γ which is supported on an interval (a,b) is denoted $\gamma|_{(a,b)}$, with similar notation for half-open and closed intervals. A segment is (*strictly*) *monotone* if the numerical value of the characters in its reduced code are (strictly) monotone as a subset of $\mathbb{Z} \cup \{P\}$. A segment with code C is called a *back loop*.

The notion of left/right on the front of S induces an orientation on strictly monotone segments contained on the front of S in the following way. If the terminal endpoint of such a segment γ is to the right of the initial endpoint, then γ is *oriented to the right*. Similarly, if the terminal endpoint is to the left of the initial endpoint, then γ is *oriented to the left*. Since γ is strictly monotone, one of the above two possibilities must occur. We note that single characters of a code represent strictly monotone segments and so can be oriented in this way.

We will use the code for an arc or segment to find the image of the arc or segment under certain homeomorphisms of S. The process can be complicated. We now introduce a new way of concatenating strings of characters which will be more suited to finding the image of an arc or segment in certain situations.

Definition 3.12 Given two segments α and β such that the terminal character of α agrees with the initial character of β and such that these two characters have the same orientation, the *efficient concatenation* of α and β, denoted $\alpha + \beta$, is formed by removing the terminal character of α to form a new string α' and concatenating this new string with β, resulting in $\alpha'\beta$.

For example,

$$P_s 0_o + 0_o 1_o = P_s 0_o 1_o,$$

and

$$P_s 0_o 1_o 2_o 2_u + 2_u 2_u 2_o 1_o 1_u = P_s 0_o 1_o 2_o 2_u 2_u 2_o 1_o 1_u = P_s 0_o 1_u,$$

where the middle term is an unreduced code and the final term is a reduced code. See Figure 8. We note that if α and β can be efficiently concatenated, then they cannot be concatenated, because α^t and β^i have the same orientation. By a similar reasoning, if two segments can be concatenated, then they cannot be efficiently concatenated. Throughout the paper, we only (efficiently) concatenate two segments when it is possible.

As written, the code of a segment is not well behaved under homotopy because every segment is homotopically trivial or homotopic into a puncture. We will introduce a standard position for segments on S with the property that any two segments that are homotopic rel endpoints will, in standard position, have the same reduced code. Standard position will also allow us to find the image of a segment under a permissible shift using only its code.

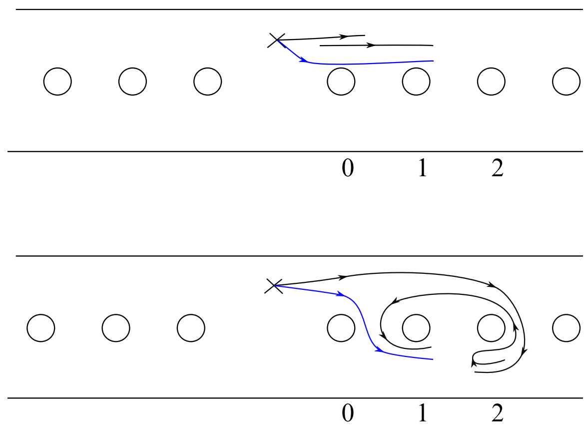


Figure 8: The two examples of efficient concatenation following Definition 3.12. In each, the (reduced) code for the blue segment is the efficient concatenation of the codes of the two black segments.

Definition 3.13 Fix a simple closed curve S_j as in Figure 9 for each $j \in \mathbb{Z} \cup \{P\}$. Let

$$\mathcal{C} = \{S_j \mid j \in \mathbb{Z} \cup \{P\}\}.$$

For each $j \in \mathbb{Z} \cup \{P\}$, we orient the simple closed curve B_j clockwise and identify B_j with the subset $\mathbb{S}^1$ of $\mathbb{C}$ by a homeomorphism which preserves this orientation. Fix points $b_L^j, b_R^j \in B_j$ corresponding to $-1, 1 \in \mathbb{S}^1$, respectively. Here L and R stand for *left* and *right*.

To describe standard position, we will sometimes move endpoints of segments γ to lie on various boundary components. When we do this, we will use the following convention. Suppose $\gamma^i = k_{o/u}$ and we want to move the initial endpoint of γ onto the boundary component B_k. If γ is oriented to the right, then we move the initial endpoint of γ to b_L^k, and if it is oriented to the left, we move the initial endpoint to b_R^k. On the other hand, suppose $\gamma^t = (k')_{o/u}$ and we want to move the terminal endpoint of γ onto the boundary component $B_{k'}$. If γ is oriented to the right, we move the terminal endpoint of γ to $b_R^{k'}$, and if it is oriented to the left, we move the terminal endpoint to $b_L^{k'}$. See Figure 10. Moving endpoints in this way does not change the code for γ.

It will be useful to understand how a given segment interacts with the domain of a permissible shift.

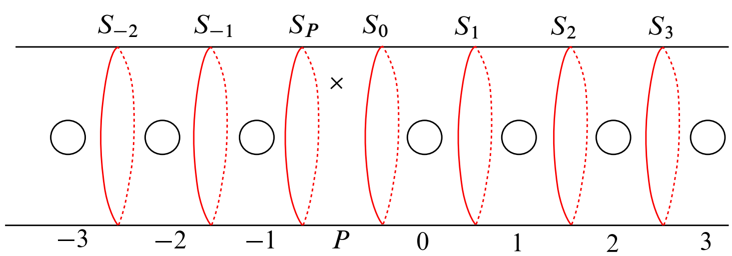


Figure 9: Some of the simple separating curves in $\mathcal{C}$ defined in Definition 3.13.

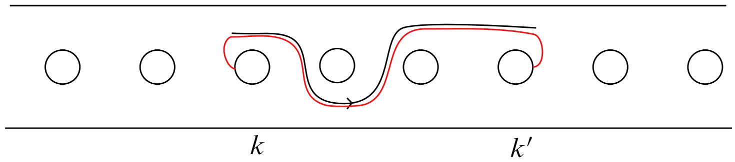


Figure 10: Adjusting the endpoints of γ (black) results in the red segment.

Definition 3.14 Let h be a permissible right shift with domain D and turbulent region (n_1, n_2). Then both (the upper and lower) boundary components of D have the same reduced code on (n_1, n_2). For any $k, k' \in (n_1, n_2)$, we let $\partial D|_{[k,k']}$ be the reduced code of the boundary components of D on the interval $[k, k']$ with similar notation for open and half-open intervals. If a segment γ has the same code as ∂D on an interval $(k_1, k_2) \subset (n_1, n_2)$, we say that γ *follows* ∂D or *agrees with* ∂D on that interval. As noted in the convention above, there is an analogous definition when h is a left shift. When a segment γ supported on $[n_1, n_2]$ intersects one component of ∂D, we call this a *half crossing*. When such a γ intersects both components of ∂D so that the code for the subsegment of γ between these two half crossings is empty, in the sense that this subsegment does not pass over or under B_k for any k, we call this a *full crossing*. See Figure 11.

Even though S is a straightforward surface, standard position for segments on S is necessarily complicated. Before introducing it formally in the next two subsections, we briefly give an intuitive idea of standard position in the following remark. On a first reading of this paper, we strongly suggest the reader read this remark and study Figures 12–15 instead of reading the formal definition of standard position given in Sections 3.2.1 and 3.2.2. The reader may then safely skip to Section 3.3. If, later in the paper, the image of a particular segment seems counterintuitive, likely this is because we put the segment in standard position before taking its image. This would be a good time to look back at Sections 3.2.1 and 3.2.2 with that example of a segment in mind.

Remark 3.15 Let h be a permissible right shift with domain D and turbulent region (n_1, n_2), and let γ be a segment. To put γ in standard position with respect to h, we first homotope subsegments of γ that are contained in the region $(-\infty, n_1) \cup (n_2, \infty)$ to be completely contained in D. For the subsegments of γ contained in the region $[n_1, n_2]$, we homotope γ so that crossings are full crossings whenever possible. We will always be able to make crossings full except near n_1 or n_2, because n_1 and n_2 are where γ leaves

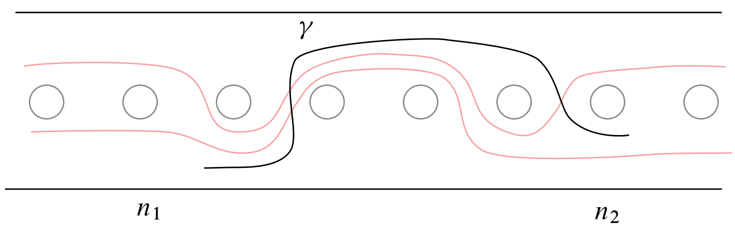


Figure 11: The domain D for a permissible shift is shown in red. The segment γ follows ∂D on $[n_1 + 1, n_1 + 3]$, has a half crossing between B_{n_2-1} and B_{n_2}, and has a full crossing between B_{n_1+1} and B_{n_1+2}.

the turbulent region and enters the shift region, where we have already ensured that it is contained in D. We also homotope γ to minimize the number of full crossings. If γ contains a loop $(n_1)_{o/u}(n_1)_{u/o}$ or $(n_2)_{o/u}(n_2)_{u/o}$, we homotope γ so that it crosses one connected component of ∂D immediately before the loop and the other connected component immediately after. In other words, γ has to enter one side of D before the loop and exit the other side of D after the loop. If γ contains the character C, we homotope it so that the subsegment of γ represented by C exits and reenters the front of S at the same place, in the sense that the endpoints of this subsegment lie on the same simple closed curve S_i. Finally, we also homotope each endpoint of γ to lie either on the nearest B_k (if it is in the shift region) or the nearest S_k (if it is in the turbulent region).

Standard position will be slightly different for those segments which contain back loops. We first discuss standard position for segments that do not contain back loops.

3.2.1 Segments without back loops Given a segment γ whose reduced code does not contain C and a permissible right shift h with domain D and turbulent region (n_1, n_2), we put γ into *standard position with respect to h* as follows.

If the endpoints of γ are not contained in (n_1, n_2) and it is possible to homotope γ completely inside of the domain of h, we do so, and we move the endpoints of γ to lie on the B_k curve numbered by the first and last characters of the code as described above.

Otherwise, γ can be written as the concatenation $\gamma_1 \dots \gamma_k$ of disjoint connected maximal subsegments such that each γ_i is either

(a) supported on either $(-\infty, n_1]$ or $[n_2, \infty)$; or

(b) supported on (n_1, n_2).

We now homotope each γ_i individually. If γ_i satisfies (a), then we move the endpoints of γ_i onto the B_k curve numbered by the first and last character of the code as above and homotope the interior of γ_i to lie completely inside the domain of h. We homotope segments γ_i satisfying (b) using the following procedure:

Step (i) If the initial character of the segment is $k_{o/u}$, move the initial endpoint of the segment onto the separating curve S_k if $k_{o/u}$ is oriented to the right and onto S_{k+1} if it is oriented to the left. If the terminal character of the segment is $k'_{o/u}$, move the terminal endpoint of the segment onto $S_{k'}$ if $k'_{o/u}$ is oriented to the left and onto $S_{k'+1}$ if it is oriented to the right. Now move the endpoints along the S_j containing them to reduce the number of full and half crossings, if possible, without creating any self-intersections. In particular, the endpoints should not lie in the domain of h. See Figure 12.

There is one caveat to the rule above. Note that when γ_i is type (b), then γ_{i+1} and γ_{i-1} are type (a), otherwise γ_i would not be maximal. In the case that either of these neighboring segments is exactly a

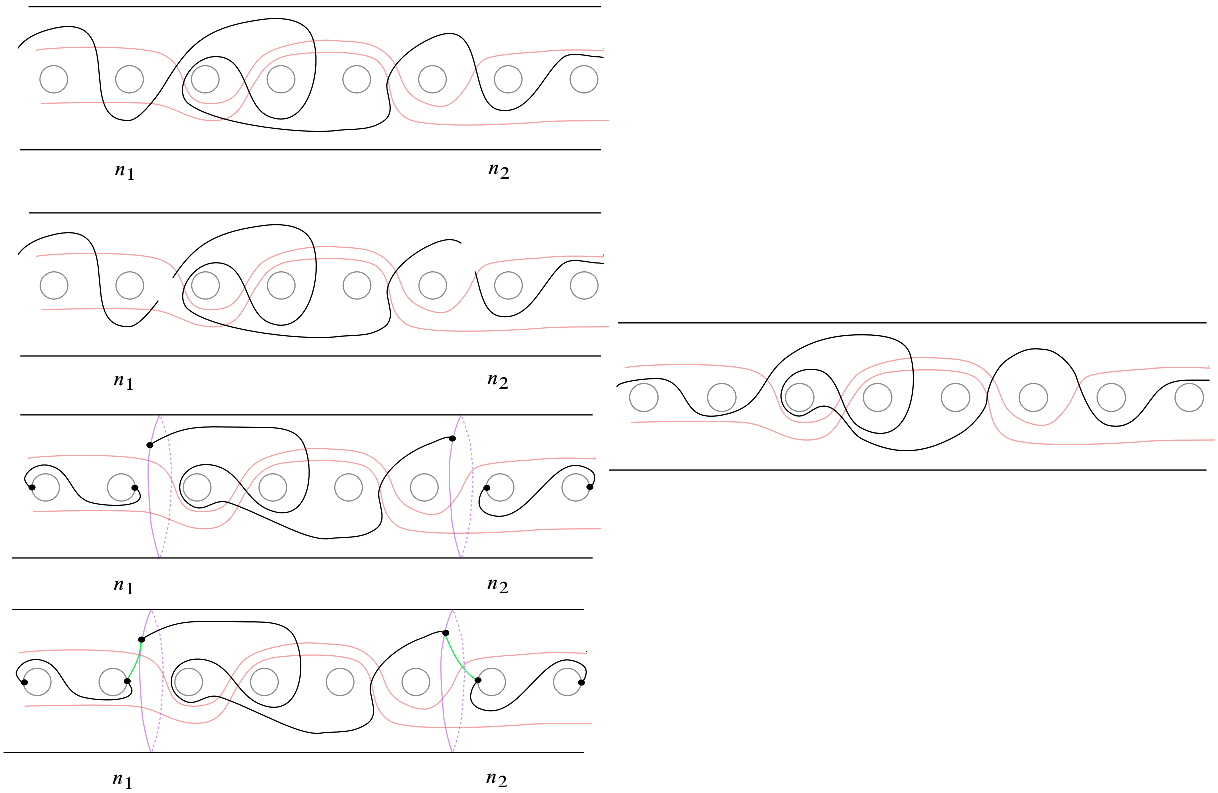


Figure 12: The domain D for a permissible shift is shown in red. In black is a segment γ that has no $(n_i)_{o/u}(n_i)_{u/o}$ for $i = 1, 2$. Left: putting γ in standard position. Right: a simple segment with the same code as γ which we say is in standard position.

loop around n_1 or n_2 we must adjust the position of the endpoints of γ_i along S_j. If $\gamma_{i+1} = (n_k)_o(n_k)_u$, for $k = 1$ or 2, we require that the terminal endpoint of γ_i is above D and require that the terminal endpoint is below D when $\gamma_{i+1} = (n_i)_u(n_i)_o$. Similarly, if $\gamma_{i-1} = (n_i)_o(n_i)_u$, for $i = 1$ or 2, we require that the initial endpoint of γ_i is below D and require that the initial endpoint of γ_i is above D when $\gamma_{i-1} = (n_i)_u(n_i)_o$. See Figure 13. Note that this repositioning of endpoints can cause additional crossings of D as in Figure 14, but this is the appropriate configuration for our calculations.

Step (ii) Homotope the segment rel endpoints to make all crossings full. Since Step (i) ensures that the endpoints of γ_i are always outside the domain D, this is always possible.

Step (iii) Homotope the segment rel endpoints to reduce the number of crossing by removing all bigons that bound disks and have one side on the segment γ_i and the other side on ∂D.

Step (iv) Finally, there may be a choice of where a full crossing occurs. If there is such a choice, then it will always be possible to homotope rel endpoints so that the crossing occurs between two adjacent characters $k_{o/u}$, $(k')_{o/u}$ with $k, k' \in (n_1, n_2)$ such that the o/u pattern of k and/or k' does not match that of ∂D, and our convention is to make this choice for the largest possible k, k'. For example, in Figure 12,

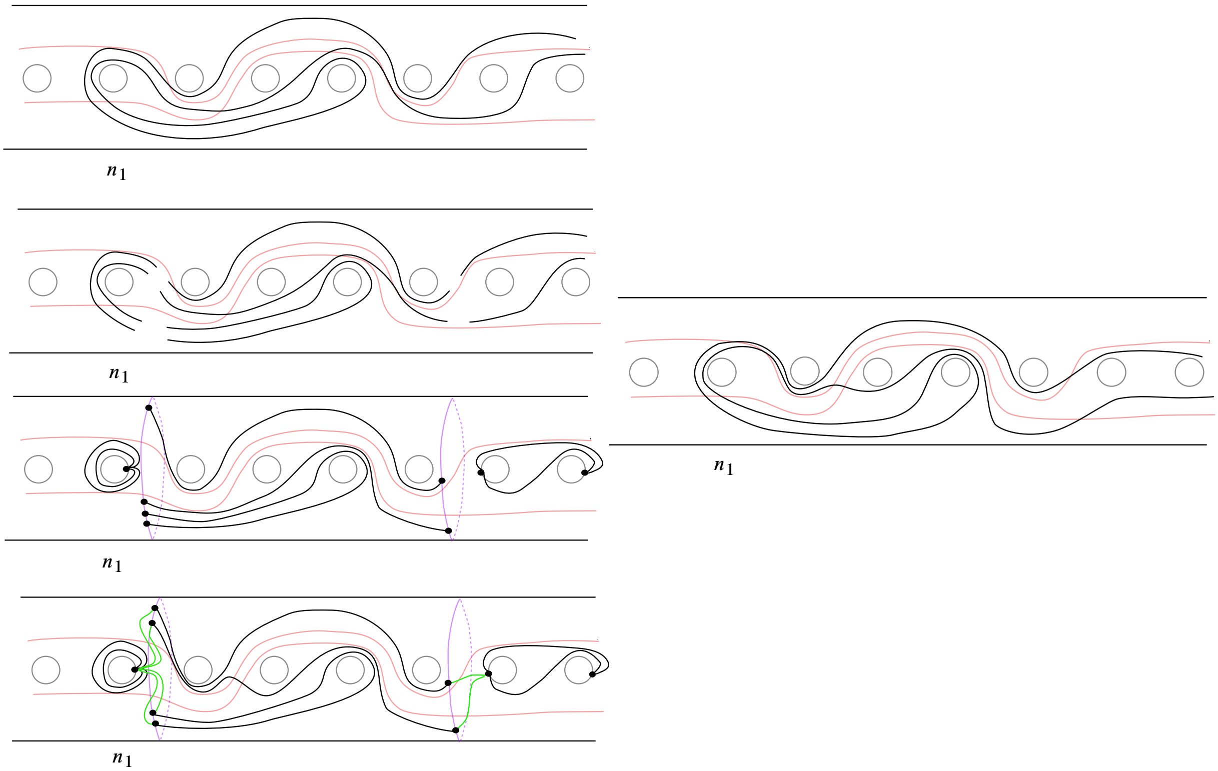


Figure 13: A segment γ that contains multiple copies occurrences of $(n_1)_o(n_1)_u$. Left: putting γ in standard position. Right: a simple segment with the same code as γ which we say is in standard position.

the bottom strand could fully cross the domain between $(n_1+1)_o$ and $(n_1+1)_u$ or between $(n_1+1)_u$ and $(n_1+2)_u$; we make the latter choice.

At this point, we have a collection of disjoint subsegments, each in standard position. The final step is to connect the endpoints of these segments, in the order they were originally connected, by segments called *connectors*. Connectors always occur between characters with numerical value n_1 and n_1+1 or n_2 and n_2-1. For the purposes of the code, we picture these segments extended slightly in either direction to overlap with the segments on either side so that the code of a connector will always be a pair $j_{o/u}(j+1)_{o/u}$ or $(j+1)_{o/u}j_{o/u}$. Thus, if a connector α connects disjoint subsegments δ_1 and δ_2 (in that order), then $\alpha = (\delta_1)^t(\delta_2)^i$ and we can write $\delta_1 + \alpha + \delta_2$. Note that this code is equivalent to the concatenation $\delta_1\delta_2$. By construction, connectors always have one endpoint inside and one endpoint outside the domain D of the shift. After applying this procedure, the resulting segment in standard position may no longer be simple. However, it will have the same code as our original γ.

The following lemma summarizes the above procedure.

Lemma 3.16 *A segment γ in standard position with respect to a permissible right shift h which is not completely contained in the domain of h can be written as the efficient concatenation of* (*possibly empty*)

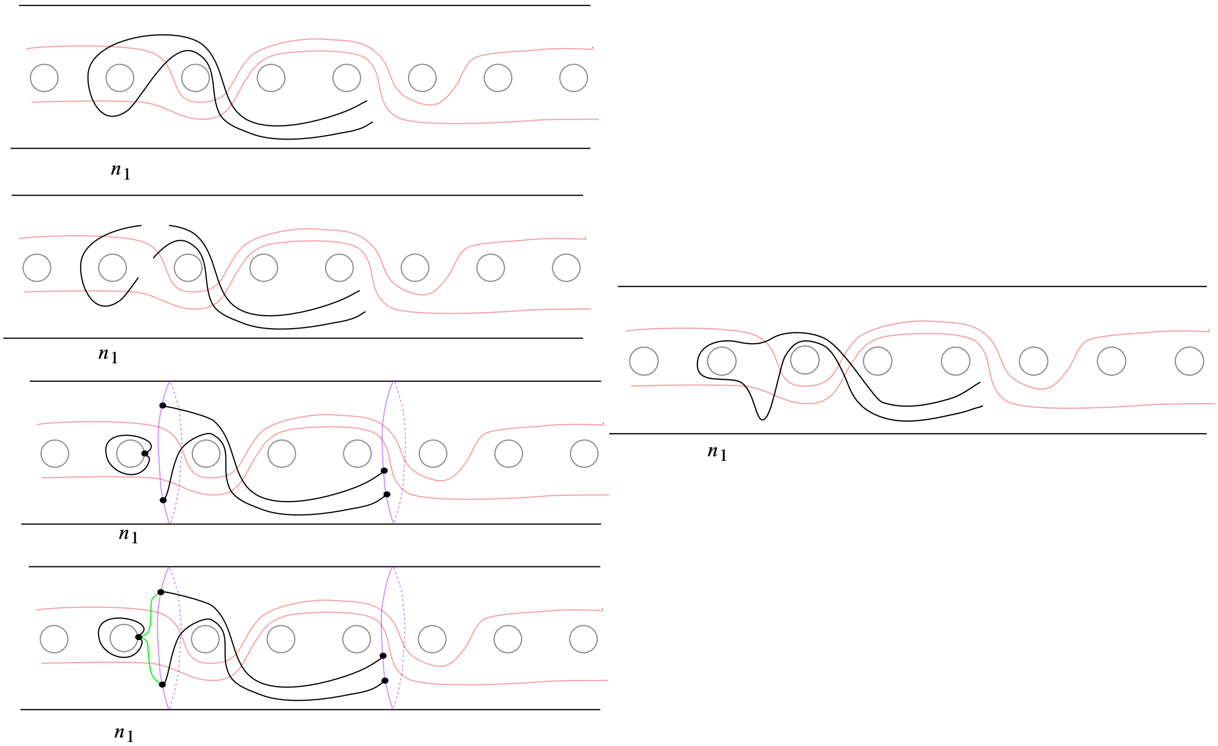


Figure 14: A segment γ that contains $(n_1)_o(n_1)_u$. Note that the requirements on the endpoints of γ_i in step (i) of the procedure for standard position require that we have one connector cross the top of ∂D and one connector cross the bottom of ∂D. Left: putting γ in standard position. Right: a simple segment with the same code as γ which we say is in standard position.

segments in the following way,

$$\gamma = \gamma_1^{\mathrm{tu}} + \gamma_1^{c1} + \gamma_1^{\mathrm{sh}} + \gamma_1^{c2} + \gamma_2^{\mathrm{tu}} + \gamma_2^{c1} + \cdots + \gamma_n^{\mathrm{sh}} + \gamma_n^{c2}, \tag{1}$$

where for each i,

- γ_i^{tu} *is supported on the turbulent region (n_1, n_2), has been put into standard position following steps* (i)–(iv) *above, and has both endpoints outside of the domain D;*
- γ_i^{sh} *is supported on the shift region $(-\infty, n_1] \cup [n_2, \infty)$, is completely contained in the domain D, and has both endpoints on the B_k curves; and*
- $\gamma_i^{c1}, \gamma_i^{c2}$ *are connectors, each of which has code length* 2*, is supported on either $[n_1, n_1+1]$ or $[n_2-1, n_2]$, and has one endpoint on B_{n_1} or B_{n_2} and one endpoint outside the domain D on S_{n_1+1} or S_{n_2}.*

We will sometimes use the notation γ_i^{conn} for a connector if it is not important if this subsegment is the first connector γ_i^{c1} or the second connector γ_i^{c2}.

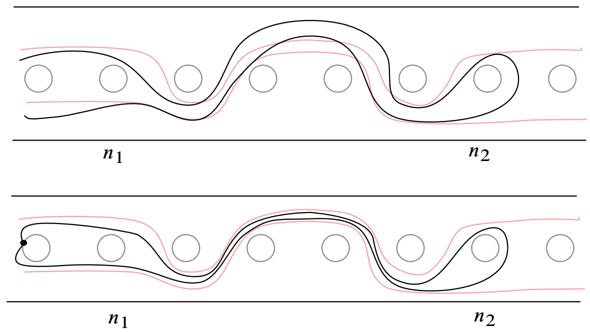


Figure 15: A segment γ that has endpoints in $(-\infty, n_1] \cup [n_2, \infty)$ and can be homotoped to be completely inside D.

It is often convenient to abuse notation and say that a simple segment is in "standard position" even if it is not the result of the above procedure because these segments are easier to draw and think about. What we mean by this is that the segment intersects the boundary of D in the same place as it would in standard position. In Figures 12–15, we give a segment, the steps to put it in standard position, and an example of a simple segment with the same code which we also say is in standard position.

Because each segment in (1) has fixed endpoints, its image under h is well-defined up to homotopy rel endpoints. Thus we may use the decomposition of γ to find its image under h as

$$h(\gamma) = h(\gamma_1^{\mathrm{tu}}) + h(\gamma_1^{c1}) + h(\gamma_1^{\mathrm{sh}}) + h(\gamma_1^{c2}) + h(\gamma_2^{\mathrm{tu}}) + h(\gamma_2^{c1}) + \cdots + h(\gamma_n^{\mathrm{sh}}) + h(\gamma_n^{c2}).$$

Since in standard position γ may not be simple, its image under h may not be simple. However, this code corresponds to a unique (homotopy class of) simple segment with the same endpoints. It is important that we use *efficient* concatenation when calculating $h(\gamma)$. Using regular concatenation, the code for γ is simply $\gamma_1^{\mathrm{tu}}\gamma_1^{\mathrm{sh}} \dots \gamma_n^{\mathrm{tu}}\gamma_n^{\mathrm{sh}}$. However, it is not true that $h(\gamma_1^{\mathrm{tu}})h(\gamma_1^{\mathrm{sh}}) \dots h(\gamma_n^{\mathrm{tu}})h(\gamma_n^{\mathrm{sh}})$ is a code for γ; in fact, much of the time this code does not define a segment. See Example 3.17. Most of the interesting behavior in the image of an arc or segment under a permissible shift actually comes from the full and half crossings. Since each connector contributes a half crossing, they are essential for determining the image of γ.

Example 3.17 Consider the permissible shift h shown in Figure 11 along with the segment γ. In Figure 16, we put γ in standard position and find its image under the shift. Using code, we have $\gamma = (n_1+1)_u(n_1+2)_o(n_1+3)_o(n_2-1)_o(n_2)_u$. Note that every character in this code except the terminal $(n_2)_u$ is fixed by h, and $h((n_2)_u) = (n_2+1)_u$. If we simply compute $h(\gamma)$ character by character, we

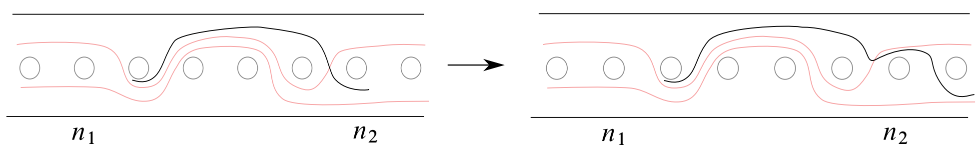


Figure 16: The segment γ from Example 3.17 in standard position (left) and its image under the shift whose domain is shown (right).

get $h(\gamma) = (n_1+1)_u(n_1+2)_o(n_1+3)_o(n_2-1)_o(n_2+1)_u$. However, this is not a well-defined code since n_2-1 and n_2+1 are not adjacent. On the other hand, using efficient concatenation, we see that

$$\begin{aligned} h(\gamma) &= h(\gamma^{\mathrm{tu}}) + h(\gamma^{\mathrm{conn}}) + h(\gamma^{\mathrm{sh}}) \\ &= h((n_1+1)_u(n_1+2)_o(n_1+3)_o(n_2-1)_o) + h((n_2-1)_o(n_2)_u) + h((n_2)_u) \\ &= (n_1+1)_u(n_1+2)_o(n_1+3)_o(n_2-1)_o + (n_2-1)_o(n_2)_o(n_2+1)_u + (n_2+1)_u \\ &= (n_1+1)_u(n_1+2)_o(n_1+3)_o(n_2-1)_o(n_2)_o(n_2+1)_u. \end{aligned}$$

Note that all but the final pair $(n_2-1)_o(n_2)_u$ are fixed by h. This final pair is a half crossing and in standard position it is a connector.

In fact, given a segment $\gamma = \gamma^{\mathrm{tu}}$ supported on (n_1, n_2), we can further decompose γ into subsegments which are disjoint from D and pairs which fully cross D in such a way that makes it straightforward to find its image under h. Write

$$\gamma = \gamma_1^d + \gamma_1^e + \cdots + \gamma_s^d + \gamma_s^e, \tag{2}$$

where each γ^d is a maximal subsegment disjoint from D, each γ^e fully crosses D, and $\ell_c(\gamma^d) \geq 2$, $\ell_c(\gamma^e) = 2$, when nonempty.

Using the above decomposition, in an unreduced code we have

$$h(\gamma) = \gamma_1^d + h(\gamma_1^e) + \cdots + \gamma_s^d + h(\gamma_s^e).$$

Every nonempty γ_j^e will have image which follows ∂D, loops around n_2, and follows ∂D back, so that

$$h(\gamma_j^e) = \partial D|_{[(\gamma_j^e)^i, n_2)}(n_2)_{o/u}(n_2)_{u/o}\overline{\partial D|_{[(\gamma_j^e)^t, n_2)}}.$$

As in Example 3.17, if we don't use efficient concatenation then we can write $\gamma = \gamma_1^d\gamma_2^d \dots \gamma_s^d$. Applying h to each of these subsegments individually would yield $h(\gamma) = \gamma$, since each of these subsegments is fixed by h, which is not the correct image.

3.2.2 Segments with back loops If γ is a segment with code equal to C, we require that γ has both endpoints on some separating curve S_i in our collection $\mathcal{C}$. We also assume that the endpoints of C lie outside the domain, and, moreover, that γ does not intersect D. There are two possibilities for γ, either γ both enters and exits the front of S at the top or bottom or (up to taking inverses) γ enters at the top and exits at the bottom of the front of the surface. Recall that we define the top/bottom of the front of S with respect to the notion of right/left on the front of S. In the first case, this implies that the endpoints of γ are both above or both below D, respectively, while in the second case one will be above and one will be below. In either case, this convention implies that $\gamma \cap D = \varnothing$ and $h(\gamma) = \gamma$.

Suppose next that γ is a segment whose code contains C but also contains other characters. Suppose for simplicity that the code for γ contains a single instance of C, so that $\gamma = \tau_1 C \tau_2$, where τ_i does not contain C for $i = 1, 2$. We put γ in standard position as follows. First note that by definition of the code, we must either have $(\tau_1)^t = (\tau_2)^i$ (if τ_1 and τ_2 have opposite orientations) or $(\tau_1)^t = (\tau_2)^i \pm 1$ (if $(\tau_1)^t$ and $(\tau_2)^i$

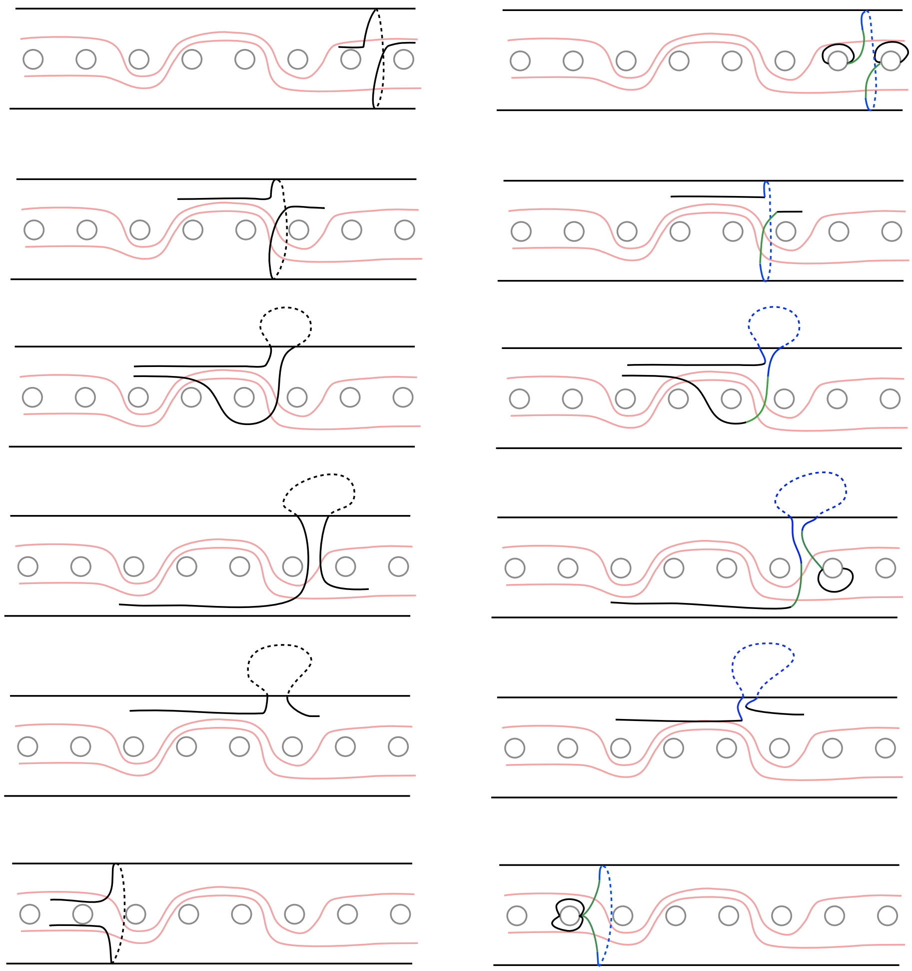


Figure 17: Various segments γ containing back loops. The segments on the right are in standard form. Back loops are blue and back loop connectors are green. Note that in the second, third, and fifth examples, at least one of $C^{\pm}$ is empty.

have the same orientation). See Figure 17. Without loss of generality, suppose τ_1^t is oriented to the right. Consider the (disjoint) segments τ_1 and τ_2, put them in standard position as in the previous section, and homotope the endpoints of C as in the previous paragraph. Note the endpoints of C will lie on the curve $S_{(\tau_1)^t+1}$. We now have three disjoint segments with codes τ_1, τ_2, and C. We will add (possibly empty) segments called *back loop connectors* from the terminal point of τ_1 to the initial point of C and from the terminal point of C to the initial point of τ_2, respectively, to form a connected segment. See Figure 17. We code these back loop connectors with the characters C^-, C^+, respectively, so that we can easily discuss their image. In particular, by a slight abuse of notation, we replace C in the code for γ with C^-CC^+.

Recall that without loss of generality, $(\tau_1)^t$ is oriented to the right. If the terminal endpoint of τ_1 lies outside D and on the same side of D as the initial point of the back loop C, then the back loop connector

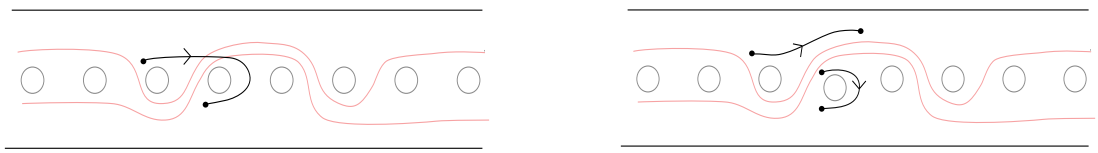


Figure 18: A segment γ (left) is split into subsegments $\gamma = \gamma_1 + \gamma_2$ and each γ_i is individually put into standard position with respect to the shift h whose domain is shown (right). This causes a loss of information when we take the images of γ_i individually under h: since $h(\gamma_i) = \gamma_i$ for $i = 1, 2$, we have $h(\gamma_1) + h(\gamma_2) = \gamma_1 + \gamma_2 = \gamma$, but γ is clearly not fixed by h.

C^- is empty. An analogous statement holds for C^+, where we use the initial endpoint of τ_2 in place of the terminal endpoint of τ_1. Each nonempty back loop connector $C^\pm$ either

(i) has one endpoint on $B_{(\tau_1)^t}$ or $B_{(\tau_2)^i}$ (depending on whether it is C^- or C^+), the other endpoint on $S_{(\tau_1)^t+1}$, and half-crosses D; or

(ii) has both endpoints on $S_{(\tau_1)^t+1}$ which are outside of D and fully crosses D.

3.3 Gaps in segments

When we use the code of a segment to find its image under a permissible shift, we first break it into smaller subsegments using standard position. When we do this, we always use efficient concatenation, so that the codes of the individual pieces overlap in a single character. The goal of efficient concatenation is to ensure that we do not lose any information about the segment by breaking it into pieces. However, we need to be careful when we do this. If we first break a segment into subsegments and then put each subsegment into standard position, it is possible that we will lose some information. In particular, we may cause there to be a "gap" in the segment. Based on standard position, these gaps can only occur when breaking a segment in the interior of the turbulent region (n_1, n_2).

To see this, suppose we break a segment $\gamma = \gamma_1 + \gamma_2$ into two subsegments such that the numerical value of $\gamma_1^t = \gamma_2^i$ is $j \in (n_1, n_2)$. If we put each γ_i into standard position individually, it is possible that γ_1^t and γ_2^i lie on opposite sides of D (see Figure 18). In this case, we have lost the full crossing between them. Recall that a shift fixes the surface outside of its domain. In the region (n_1, n_2), a segment in standard position is disjoint from the domain of the shift *except* where there is a full crossing (see the decomposition in equation (2)), so the full crossings are essential for determining the image of a segment. Thus we cannot use γ_1 and γ_2 to find the correct image of γ.

In order to ensure that we do not lose any information when working with a segment and subsegments in the turbulent region, we always first homotope the whole segment γ to be a simple segment in standard position. We then fix this particular representative of the homotopy class of γ for the remainder of the time we work with it. Thus, when we break γ into subsegments γ_1 and γ_2, we do *not* put these into standard position individually. To be precise, we choose the endpoints γ_1^t and γ_2^i to be the intersection of

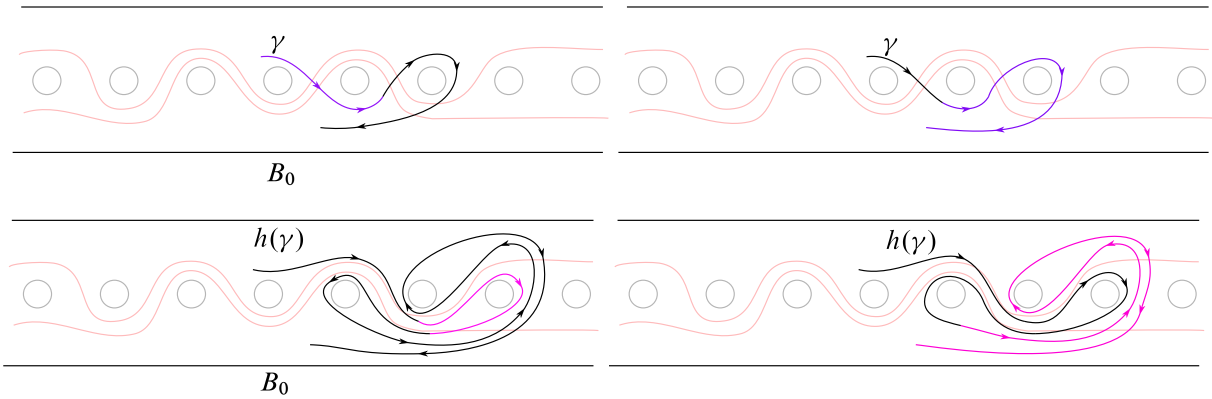


Figure 19: A segment γ and its image $h(\gamma)$ are shown. Left: a subsegment which does not have a preimage. The pink subsegment of $h(\gamma)$ does not have an inverse; it is a proper subpath of the image of the purple subsegment of γ. Right: a subsegment which has a preimage. The pink subsegment of $h(\gamma)$ has an inverse; it is the image of the purple subsegment of γ.

γ and the appropriate separating curves in our collection $\mathcal{C}$ and we do not allow any further homotopies of γ_1 or γ_2. This will always ensure that there are no gaps between γ_1 and γ_2.

In certain cases, it may be simpler to break γ into subsegments in a different way, and we do this whenever it will avoid technicalities. For example, in Figure 18, we may simply choose not to divide γ into subsegments at all. On the other hand, we could also choose to make γ_1 or γ_2 longer than is strictly necessary in a particular calculation in order to avoid a potential loss of information.

3.4 Taking inverses of segments

In general, if we have a segment γ and a subsegment ζ of $h(\gamma)$, there is not necessarily a subsegment of γ which we may call $h^{-1}(\zeta)$. In other words, not every subsegment of $h(\gamma)$ is the image of a subsegment of γ. It is important here that when we think of a subsegment of γ, we are fixing γ in standard position. That is, we are thinking of a *reduced* code for γ, rather than any (unreduced) code representing γ.

For example, consider the shift shown in Figure 19, left. Here, the pink subsegment ζ_1 of $h(\gamma)$ is not the image of any subsegment of γ. Rather, it is properly contained in the image of the purple subsegment of γ, specifically because it is in the image of the full crossing of the purple segment with D. Explicitly,

$$\gamma = 0_o 1_u 2_o 2_u 1_u, \quad h(\gamma) = 0_o 1_u 2_o 2_u 1_u 0_o 0_u 1_u 2_u 2_o 1_u 1_o 2_o 2_u 1_u 0_u,$$

and one can see that if the subsegment $\zeta_1 = 1_u 2_o 2_u 1_u$ of $h(\gamma)$ then there is no subsegment γ' of γ for which $h(\gamma') = \zeta_1$.

However, we may take an inverse image of a subsegment ζ of $h(\gamma)$ whenever we *know* that ζ is the image of a subsegment of γ. This is the case, for example, when $\zeta = h(\gamma)$; in other words, it is true that $\gamma = h^{-1}(h(\gamma))$. This can also happen when the initial and terminal characters of a subsegment of $h(\gamma)$

are the images of the initial and terminal characters of a subsegment of γ. For example, in Figure 19, right, a direct computation will show that the initial and terminal characters of ζ_2 (in pink) are the images of the initial and terminal characters of the purple subsegment of γ. Therefore ζ_2 is the image of the purple subsegment of γ, which is precisely the result of calculating $h^{-1}(\zeta_2)$.

4 Characterizing loops with trivial image

As in Section 3, we let S be the biinfinite flute surface with a distinguished puncture p and fix the collection of simple closed curves $\{B_i \mid i \in \mathbb{Z} \cup \{P\}\}$ as in Definition 3.1. Let h be a permissible shift (see Definition 3.8) and $k \in \mathbb{Z} \cup \{P\}$. By an abuse of notation, we may occasionally write $h(k)$, by which we mean that $h(k)$ is the label of $h(B_k)$. Thus, given any segment whose code is $k_{o/u}$, we have $h(k_{o/u}) = h(k)_{o/u}$.

Recall that a segment is a path which does not have both endpoints on p.

Definition 4.1 A segment in standard position is *trivial* if it can be homotoped rel endpoints to one of the following:

(a) a segment contained in one of the separating curves $S_i \in \mathcal{C}$; or

(b) a point.

We will use the notation $\varnothing$ to denote the reduced code for a trivial segment. For example, we write $k_o k_o = \varnothing$.

Definition 4.2 A *loop* is a segment that has one of the following forms:

(1) $\delta_1 a_{o/u} a_{u/o} \delta_2$ for some a where $\delta_1^i = \delta_2^t$,

(2) $a_{o/u} a_{u/o}$ for some a, or

(3) C.

See Figure 20 for examples of loops. A loop satisfying (1) is called a *regular loop*, a loop satisfying (2) is called an *over-under loop*, and, as in Definition 3.11, a loop satisfying (3) is called a *back loop*. Note

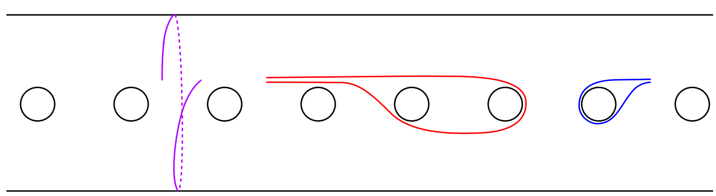


Figure 20: Some examples of loops from Definition 4.2. The loop in red fits case (1), the loop in blue fits case (2), and the loop in purple fits case (3).

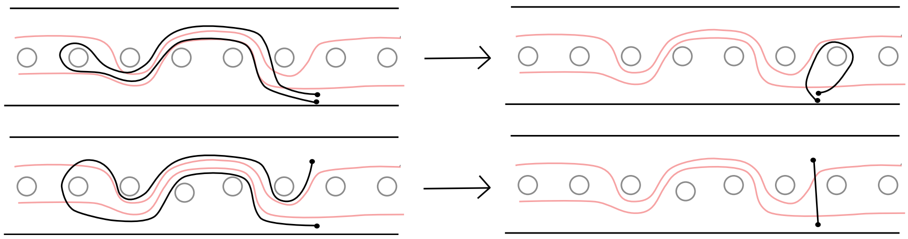


Figure 21: The permissible shift h translates to the right. The first two loops on the left are homotopic via homotopies which keep the endpoints on the same fixed separating curve. However, one image is trivial while the other is not.

that regular loops always contain an over-under loop but not all over-under loops can be extended to a regular loop. A *single loop* is a back loop, an over-under loop, or a regular loop such that δ_i does not contain a loop for $i = 1, 2$.

With the above definitions, the rest of this section is devoted to analyzing the following question.

Question 4.3 *Let h be a permissible shift. When does h send a loop to a trivial segment*?

The reason we introduced standard position is to ensure that this question is well defined. The issue is that homotopies of a loop can change whether or not its image is trivial, even if those homotopies keep the endpoints on a fixed separating closed curve (see Figures 21 and 22). Thus it is important that, given a loop, we first put it in standard position before applying the permissible shift h. This will remove any possible ambiguity in the image of the loop.

In this section, we first introduce a kind of cancellation in the image of a segment and its code which we call *cascading cancellation*. This kind of cancellation will cause technical problems throughout the paper, and much of Section 6 is devoted to understanding how to control it. We then prove Theorem 4.8, which answers Question 4.3. We end the section with a discussion of several technical consequences of the Theorem 4.8 which will be useful later.

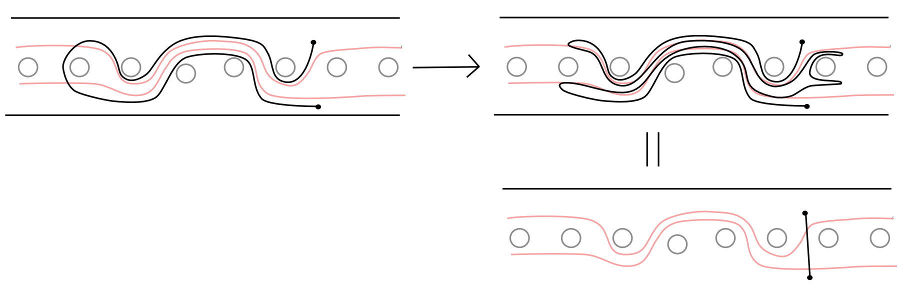


Figure 22: The second loop from Figure 21 is in standard position. We show why its image under h is trivial.

4.1 Cascading cancellation

Definition 4.4 We call an arc γ *symmetric* if $\gamma = \delta q_1 q_2 \bar{\delta}$ for any characters q_1, q_2. In other words, a reduced code for γ is palindromic with the exception of the middle two characters. Note in particular that this implies that q_1, q_2 have the same numerical value.

Recall that we find the image of a path with code $\alpha\beta$ under a permissible shift f as follows. Let q be the last character of α and q' be the first character of β. Then

$$f(\alpha\beta) = f(\alpha) + f(qq') + f(\beta).$$

While $f(\alpha)$, $f(qq')$, and $f(\beta)$ are all reduced codes, it is possible that the efficient concatenation will cause there to be cancellation. For example, if $f(\alpha) = 1_u 1_o 2_o 3_o$ and $f(qq') = 3_o 3_o 2_o 1_o 0_o P_o$, then $f(\alpha) + f(qq') = 1_u 0_o P_o$. When this type of cancellation occurs, that is, when a character of $f(qq')$ does not appear in a reduced code of the image, we say there is *cancellation involving* $f(qq')$. In our example, there is cancellation involving $f(qq')$ and $f(\alpha)$. When it is necessary to be more precise, we may also say there is (respectively, is not) cancellation involving a character s, if s appears in the unreduced code but not the reduced code (respectively, appears in both the unreduced code and the reduced code) of the image. Thus in our example, there is no cancellation involving 1_u but there is cancellation involving 2_o. Our goal is to understand, in general, when there is cancellation with a particular character in a path under a permissible shift.

Suppose $\alpha = \gamma q_1$ and h is a permissible shift. Let q be the terminal character of γ. It is tempting to believe that if we can show that there is no cancellation involving $h(q_1)$ within $h(qq_1)$, then there is no cancellation involving $h(q_1)$ at all. However, this is not sufficient, for it is possible that there is "cascading cancellation". Before giving a formal definition, we illustrate this phenomenon with an example.

Example 4.5 Consider the segment $\delta = 3_u 2_u 1_o 0_u$ and the permissible left shift h whose domain is shown in Figure 23. Then

$$h(\delta) = h(3_u 2_u) + h(2_u 1_o) + h(1_0 0_u).$$

We have

$$h(1_o 0_u) = 1_o 0_u, \quad h(2_u 1_o) = 2_u 1_o \quad \text{and} \quad h(3_u 2_u) = 3_u 2_u 1_o 0_u (-1)_u (-1)_o 0_u 1_o 2_u 2_u.$$

Putting this together, we obtain $h(\delta) = 3_u 2_u 1_o 0_u (-1)_u (-1)_o$.

There is no cancellation involving either of the terms when computing $h(2_u 1_0) + h(1_o 0_u)$. However, when computing $h(3_u 2_u) + h(2_u 1_0)$, we see that $h(2_u 1_o)$ completely cancels with an initial segment of $h(3_u 2_u 3)$, and $h(1_o 0_u)$ completely cancels with the next subsegment of $h(3_u 2_u)$. Therefore, there is in fact cancellation involving $h(0_u)$ in $h(\delta)$.

Definition 4.6 Formally, given an arc $\delta_1 \dots \delta_n$ and a permissible shift f, we say there is *cascading cancellation involving* $f(\delta_n)$ if there is cancellation involving $f(\delta_n)$ in $f(\delta_1 \delta_2 \dots \delta_n)$ but there is no cancellation involving $f(\delta_n)$ in $f(\delta_{n-1}\delta_n)$.

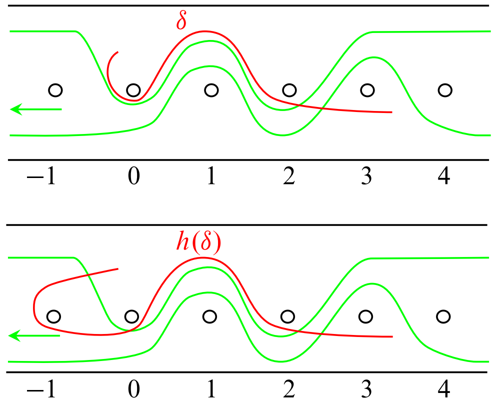


Figure 23: Above: the segment δ in standard position. The domain of the left shift h is shown in green. Below: the image $h(\delta)$.

Understanding and controlling cascading cancellation is the difficult part of many of the proofs in this paper. The remainder of this subsection is devoted to theorems that will allow us to control cascading cancellation for permissible shifts. We will often be in the following situation: there is a segment $\gamma = \gamma_1\gamma_2$ whose image under a shift h we would like to understand and we know that $h(\gamma_1)$ has some desired quality (such as containing a loop, for example). In order to show that the desirable behavior of $h(\gamma_1)$ persists in $h(\gamma)$, we need to ensure that there is no cancellation involving $h(\gamma_1)$ and $h(\gamma_2)$ by controlling cascading cancellation. In Section 6, we will revisit this topic and prove some additional results that allow us to control cascading cancellation for the particular homeomorphisms we construct, which are compositions of shifts.

Lemma 4.7 *Let h be a permissible right shift with domain D and turbulent region (n_1, n_2). Let α be a strictly monotone segment supported on (n_1, n_2). Then in a reduced code $h(\alpha)$ has n loops around n_2, where n is the number of times α fully crosses D.*

Proof Without loss of generality, we will assume that α is strictly monotone increasing, that is, the numerical value α^i is strictly less than that of α^t, as the conclusion is invariant under replacing α by $\bar{\alpha}$. As in Section 3.2.1, put α in standard position. Since no segment in standard position which is supported on (n_1, n_2) will be completely contained in the domain of the shift, we may write

$$\alpha = \alpha_1^d + \alpha_1^e + \cdots + \alpha_s^d + \alpha_s^e,$$

where each α_j^d is a maximal subsegment disjoint from D, each α_j^e fully crosses D, and $\ell_c(\alpha_j^d) \geq 2$, $\ell_c(\alpha_j^e) = 2$ when nonempty. Notice that if $\alpha_j^d \neq \varnothing$ and $\alpha_{j+1}^d \neq \varnothing$, then also $\alpha_j^e \neq \varnothing$ by maximality.

Under the above decomposition, in an unreduced code we have

$$h(\alpha) = \alpha_1^d + h(\alpha_1^e) + \cdots + \alpha_s^d + h(\alpha_s^e),$$

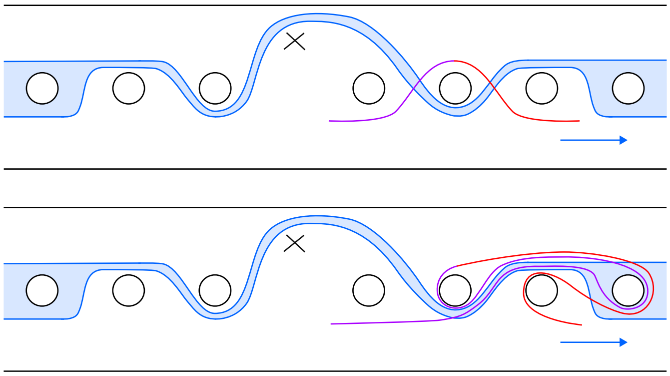


Figure 24: Above: The case where $\alpha_j^e + \alpha_j^d + \alpha_{j+1}^e = \alpha_j^e + \alpha_{j+1}^e$ in the proof of Lemma 4.7, where α_j^e is in purple and α_{j+1}^e is in red. Below: the nontrivial image of this segment under h.

where every nonempty α_j^e will have image

$$h(\alpha_j^e) = \partial D|_{[(\alpha_j^e)^i, n_2)}(n_2)_{o/u}(n_2)_{u/o}\overline{\partial D|_{[(\alpha_j^e)^t, n_2)}}.$$

Thus each full crossing α_j^e contributes a loop around n_2 in an *unreduced* code for $h(\alpha)$. We must show that such loops persist in a *reduced* code for $h(\alpha)$.

If $\alpha_j^e, \alpha_{j+1}^e \neq \varnothing$ and $\alpha_{j+1}^d = \varnothing$ then a calculation shows that in a reduced code

$$h(\alpha_j^e + \alpha_{j+1}^d + \alpha_{j+1}^e) = h(\alpha_j^e + \alpha_{j+1}^e) = h(\alpha_j^e) + h(\alpha_{j+1}^e).$$

In particular, there is no cancellation between $h(\alpha_j^e)$ and $h(\alpha_{j+1}^e)$ and the loops around n_2 persist. See Figure 24.

Now assume that α_{j+1}^d is nonempty. The maximality of α_{j+1}^d implies that α_j^e and α_{j+1}^e are nonempty as well. We must show that there is no cancellation between the loops around n_2 in $h(\alpha_j^e)$ and $h(\alpha_{j+1}^e)$ so that both loops persist in a reduced code for $h(\alpha_j^e) + \alpha_{j+1}^d + h(\alpha_{j+1}^e)$. Recall that in standard position, a full crossing occurs between two adjacent characters $k_{o/u}$, $(k')_{o/u}$ with $k, k' \in (n_1, n_2)$ such that the o/u pattern of k and/or k' does not match that of ∂D, and our convention is to make this choice for the largest possible k, k'.

The subtlety arises because the code for $h(\alpha_j^e) + \alpha_{j+1}^d$ may not be reduced if an initial subsegment of $\alpha_{j+1}^d = h(\alpha_{j+1}^d)$ agrees with ∂D, in which case $(\alpha_j^e)^t$ agrees with ∂D. If $(\alpha_j^e)^t$ is the character of the full crossing α_j^e that does *not* agree with ∂D, then this character will block cancellation between $h(\alpha_j^e)$ and α_{j+1}^d, so that the loops around n_2 in $h(\alpha_j^e)$ and $h(\alpha_{j+1}^e)$ cannot cancel. On the other hand, if $(\alpha_{j+1}^e)^i$ is the character of the full crossing α_{j+1}^e that does *not* agree with ∂D, then this character will block cancellation between the two n_2 loops, even if α_{j+1}^d fully cancels in a reduced code for $h(\alpha_j^e) + \alpha_{j+1}^d + h(\alpha_{j+1}^e)$.

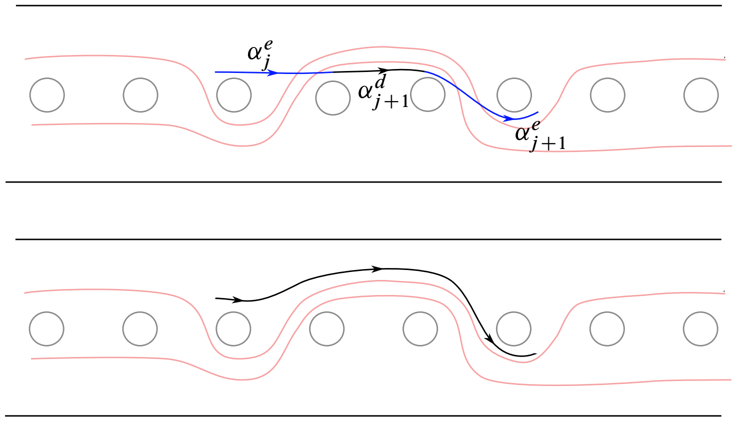


Figure 25: Above: A picture of $\alpha_j^e + \alpha_{j+1}^d + \alpha_{j+1}^e$, not in standard position, where $(\alpha_j^e)^t$ and $(\alpha_{j+1}^e)^i$ both agree with ∂D. Below: The same subsegment of α in standard position.

In the last case, where $(\alpha_j^e)^t$ and $(\alpha_{j+1}^e)^i$ both agree with ∂D, the largest possible choice of k, k' for the full crossing α_j^e would in fact result in a segment $\alpha_j^e + \alpha_{j+1}^d + \alpha_{j+1}^e$ that is fully disjoint from D, contradicting the maximality of α_{j+1}^d, so that this case cannot occur. See Figure 25.

Thus, there is no cancellation between the loops around n_2 in $h(\alpha_j^e)$ and $h(\alpha_{j+1}^e)$, which proves the lemma. $\square$

4.2 Loops with trivial image

In this subsection, we describe the form a loop must have if its image under a shift is trivial. In particular, the image of any loop which does not have the form as stated in the theorem is nontrivial. In addition, any segment that is not a loop cannot have trivial image under a shift since the numerical values the initial and terminal characters, and thus their images, differ.

Theorem 4.8 *Let h be a permissible right shift with turbulent region (n_1, n_2). Suppose β is a nontrivial loop such that $h(\beta) = \varnothing$. Then either $\beta = k_o \delta k_o$ or $\beta = k_u \delta k_u$, where*

(i) $k \in (n_1, n_2)$,

(ii) $\delta = \gamma (n_1)_{o/u} (n_1)_{u/o} \overline{\gamma}$, *and*

(iii) γ *follows ∂D between k and n_1.*

Proof Put β in standard position. We will consider the image $h(\beta)$. By the discussion in Section 3.4, we have $\beta = h^{-1}(h(\beta))$. By assumption, $h(\beta)$ is trivial, and therefore it is homotopic rel endpoints to either a segment contained in the separating curve S_k for some $k \in [n_1, n_2)$ or a point. If $h(\beta)$ is homotopic rel endpoints to a point, then $\beta = h^{-1}(h(\beta))$ is also homotopic rel endpoints to a point, in which case β is trivial, which is a contradiction. So suppose that $h(\beta)$ is homotopic rel endpoints to an embedded subsegment σ of the separating curve S_k for some $k \in (n_1, n_2]$. Then β is homotopic rel endpoints to $h^{-1}(\sigma)$.

Let D denote the domain of h. Since β is in standard position, the endpoints of σ (which are the endpoints of β) are not contained in D. There are two possibilities. If $\sigma \cap D = \varnothing$, then β is homotopic rel endpoints to $h^{-1}(\sigma) = \sigma$ and so β is trivial, which is a contradiction. On the other hand, if $\sigma \cap D \neq \varnothing$, then σ must fully cross D. A direct computation shows that, up to reversing orientation,

$$h^{-1}(\sigma) = \overline{\partial D}|_{[k,n_1)}(n_1)_{o/u}(n_1)_{u/o}\partial D|_{(n_1,k]},$$

for $k \in (n_1, n_2)$, as desired. □

We note one immediate consequence of Theorem 4.8.

Corollary 4.9 *Suppose β is any of the following*:

(1) *a loop containing more than one single loop*;

(2) *a loop containing a back loop*; *or*

(3) *a loop with endpoints outside of the turbulent region.*

Then $h(\beta)$ is nontrivial.

Some of these consequences may be surprising, so we now give a more intuitive (and informal) discussion of why the image of loops as in the corollary are nontrivial.

We first consider the case of segments which contain more than one single loop. At first glance, it may seem that if a segment β is composed of loops which each satisfy the conclusion of Theorem 4.8, then $h(\beta)$ will be trivial. The problem arises in how these loops fit together to form β. Suppose we have two loops, β_1 and β_2 which each satisfy the conclusion of Theorem 4.8. If the numerical value of β_1^t is not the same as that of β_2^i, then in order to "connect" β_1 to β_2, we must add a segment between the terminal point of β_1 and the initial point of β_2. This segment is either disjoint from D, in which case it is fixed by h, or it fully crosses D. In the former case, this segment will appear in the image of β, while in the latter case, the image of this full crossing will be nontrivial by Lemma 4.7 applied to the full crossing. If the numerical values are the same and $\beta = \beta_1\beta_2$, then β will be trivial. On the other hand, if there is some segment connecting the terminal point of β_1 to the initial point of β_2, then as before, the image of this segment will not be trivial and will force $h(\beta)$ to be nontrivial. See Figure 26 for an example of this.

Suppose next that β contains a back loop C. Intuitively, it seems reasonable that if $\beta = C\beta'\overline{C}$ and β' is a loop as in Theorem 4.8, then $h(\beta)$ is trivial. However, this is not the case. To see this, notice that if we put β' in standard position then it will have one endpoint above D and one below D. However, C and $\overline{C}$ are the same loop. If C is either above or below D, then this will cause β to have a full crossing between β' and one of C or $\overline{C}$. The image of this full crossing will be nontrivial by Lemma 4.7 applied to the full crossing, which will prevent any cancellation between $h(C) = C$ and $h(\overline{C}) = \overline{C}$. On the other hand, if C exits at the top and enters at the bottom, then $\overline{C}$ exits at the bottom and enters at the top. This will again force β to have a full crossing between β' and one of C or $\overline{C}$, preventing cancellation between the images of the back loops.

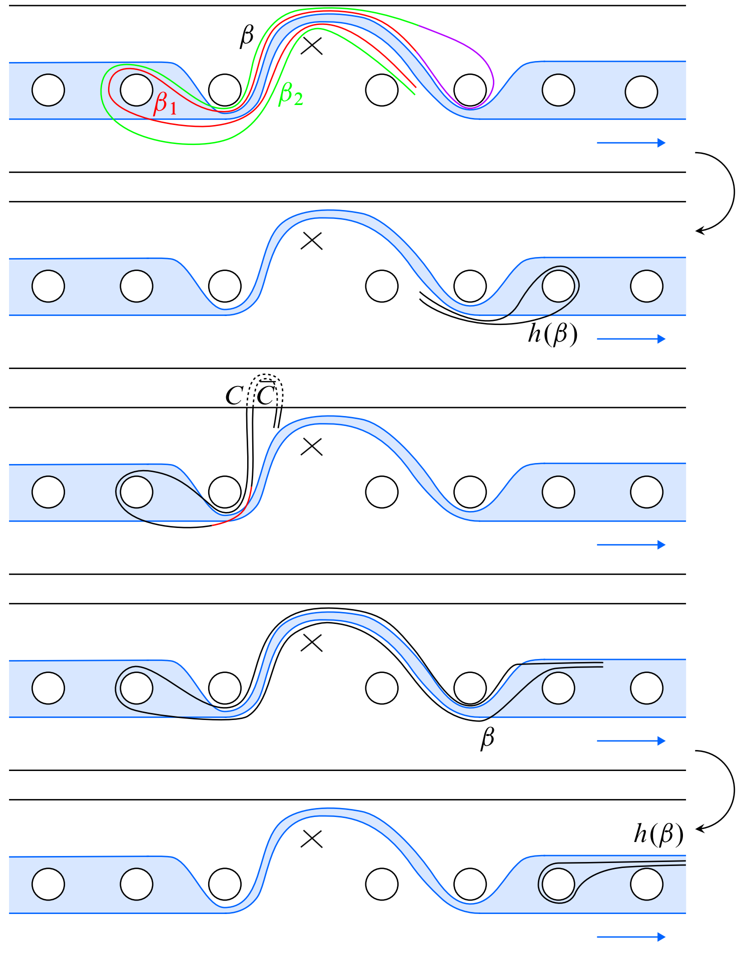


Figure 26: Examples of loops whose images are nontrivial. From top to bottom, these correspond to Corollary 4.9 (1), (2), and (3).

Finally, if β is a loop whose endpoints are outside the turbulent region, say both in $[n_2, \infty)$, and $\beta|_{(n_1,n_2)}$ has the form in the conclusion of Theorem 4.8, it seems possible that $h(\beta)$ is trivial. For example, suppose that $\beta = (n_2)_o\beta'(n_2)_o$, where β' is as in Theorem 4.8. Then since $h(\beta')$ is trivial, $h(\beta')$ will be a segment contained in the separating curve S_{n_2}. On the other hand, the images of the connectors $(n_2)_o(\beta')^i$ and $(\beta')^t(n_2)_o$ will each be of the form $(n_2+1)_o(n_2)_{o/u}(n_2-1)_{o/u}$ or $(n_2-1)_{o/u}(n_2)_{o/u}(n_2+1)_o$. In particular, we have $h(\beta) = (n_2+1)_o(n_2)_{o/u}(n_2)_{u/o}(n_2+1)_o$, which is nontrivial.

4.3 Consequences of Theorem 4.8

In this subsection, we record several (technical) consequences of Theorem 4.8 which will be useful in later sections. The first shows that loops in the shift region persist under the image of shifts.

Lemma 4.10 *Let h be a permissible right shift and γ any nontrivial, simple arc with no back loops. If $k \in (-\infty, n_1) \cup [n_2, \infty)$ and $k_{o/u}k_{u/o}$ appears in a reduced code for γ, then there is no cancellation involving $h(k_{o/u}k_{u/o})$ in an unreduced code for $h(\gamma)$. In particular, such a pair $k_{o/u}k_{u/o}$ will yield a pair $(k+1)_{o/u}(k+1)_{u/o}$ in a reduced code for $h(\gamma)$.*

Proof We argue for $k_o k_u$ without loss of generality. Fix k as above and put γ into standard position. If in standard position γ is completely contained in the domain of h, then the result is clear, so suppose this is not the case. We assume for contradiction that there is cancellation with $h(k_o k_u)$ and, without loss of generality, that it is with $h(k_u) = (k+1)_u$, as otherwise we replace γ with $\overline{\gamma}$ and argue identically for k_o instead. Let γ' be the minimal subsegment of γ beginning with $k_o k_u$ so that $h(\gamma')$ had cancellation with $h(k_u)$. The only way cancellation in an unreduced code can involve $h(k_u)$ is if a subsegment of the form $(k+1)_o(k+1)_u\delta(k+1)_u$ appears in an unreduced $h(\gamma')$ where a reduced code for δ is trivial. Since $k \in (-\infty, n_1) \cup [n_2, \infty)$, any $(k+1)_u$ in $h(\gamma')$ must be the image under h of k_u. Therefore, $\gamma' = k_o k_u \eta k_u$, where $h(k_u \eta k_u) = (k+1)_u \delta (k+1)_u$, which has trivial reduced code. This contradicts Theorem 4.8 as $k \notin (n_1, n_2)$. □

The second consequence of Theorem 4.8 shows that characters of a segment that lie in the turbulent region which disagree with the domain of a shift persist in a reduced code of the image of the segment under that shift.

Lemma 4.11 *Let h be a permissible right shift with domain D. Let δ be a simple segment whose support intersects (n_1, n_2) nontrivially, and suppose δ contains a character b with numerical value in (n_1, n_2) which disagrees with ∂D. Then b persists in a reduced code for $h(\delta)$.*

In other words, if $\delta = \delta_1 b \delta_2$, then $h(\delta) = \sigma_1 b \sigma_2$, where $\sigma_1 b = h(\delta_1 b)$ and $b\sigma_2 = h(b\delta_2)$.

Proof Write $b = k_{o/u}$, where by assumption $k \in (n_1, n_2)$.

Claim 4.12 *Since b disagrees with ∂D, any occurrence of b in $h(\delta)$ must also appear in δ.*

Proof Since b disagrees with ∂D, the segment δ cannot be homotoped rel endpoints to be completely contained in D. Thus in standard position, δ can be written as the efficient concatenation of subsegments which are disjoint from D, subsegments of length two which are either full or half crossings, and subsegments supported on $(-\infty, n_1] \cup [n_2, \infty)$ (see Lemma 3.16). We consider the images of each type of subsegment in turn. The subsegments which are disjoint from D are fixed by h. The image of subsegments supported on $(-\infty, n_1) \cup [n_2, \infty)$ have empty intersection with (n_1, n_2), and so cannot contain b. The remaining subsegments are half crossings, full crossings, or segments that include a character whose numerical value is n_1. Any subsegment of the image of any of these subsegments supported on (n_1, n_2) agrees with ∂D and so cannot contain b. This proves the claim. □

Now suppose towards a contradiction that b does not persist in $h(\delta)$. Then there must be another instance of b in an *unreduced code* for $h(\delta)$ which cancels with b. By the claim, δ must contain a subsegment of the form $b\sigma b$ whose image is trivial. Theorem 4.8 then implies that $b\sigma b = \overline{\partial D|_{(n_1,k]}}(n_1)_{o/u}(n_1)_{u/o}\partial D|_{(n_1,k]}$. However, $b = k_{o/u}$ disagrees with ∂D by assumption, and therefore $b\sigma b$ cannot have this form, which is a contradiction. $\square$

5 Constructing the homeomorphisms

As in Section 3, let S be the biinfinite flute with an isolated puncture p, and let $\{B_i \mid i \in \mathbb{Z} \cup P\}$ be the simple closed curves from Definition 3.1 bounding a collection of punctures $\{p_i \mid i \in \mathbb{Z} \cup P\}$, where $p_P = p$. As in that section, we move the punctures $\{p_i \mid i \in \mathbb{Z} \cup P\}$ to the front of S and all other punctures to the back of S.

In this section, we define a countable collection of elements $\{g_n\}_{n\in\mathbb{N}}$ in $\mathrm{Map}(S, p)$. In the following sections we will show that these elements are of intrinsically infinite type and are loxodromic with respect to the action of $\mathrm{Map}(S, p)$ on $\mathcal{A}(S, p)$.

Definition 5.1 For each $n \in \mathbb{N}$, define

(3) $$g_n := h_3^{(n)} \circ h_2^{(n)} \circ h_1,$$

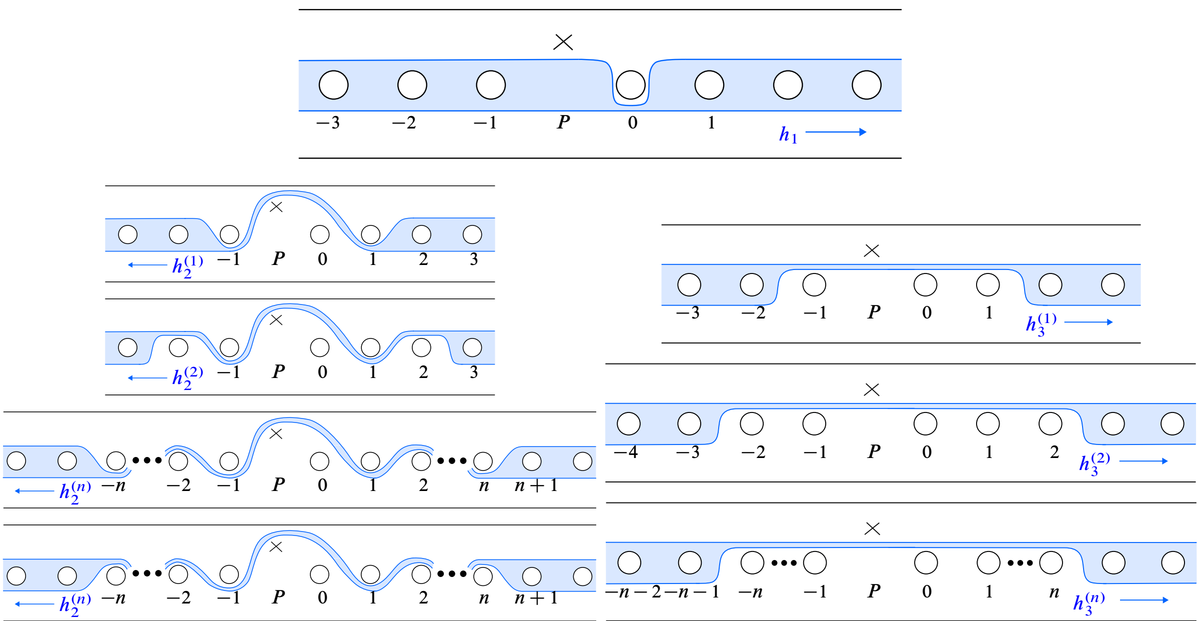


Figure 27: The various shift maps used in the construction of $g_n = h_3^{(n)} \circ h_2^{(n)} \circ h_1$. Top: the shift map h_1. Bottom left: the shift map $h_2^{(n)}$ for various n. In order from top to bottom, the first two pictures are the case of $n = 1, 2$ (respectively) and the bottom two are for general n odd and general n even (respectively). Bottom right: the shift map $h_3^{(n)}$ for various n. In order from top to bottom, the first two pictures are the case of $n = 1, 2$ and the bottom picture is the case of general n.

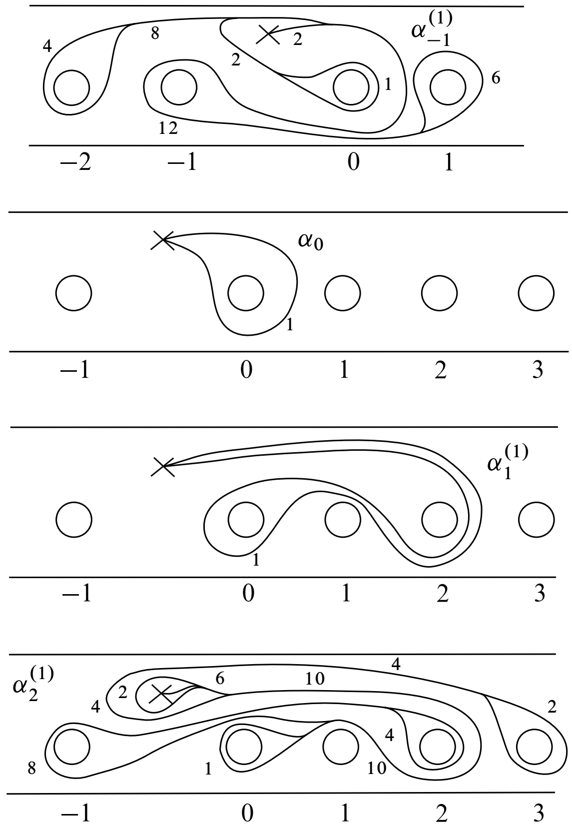


Figure 28: The arcs $\alpha_i^{(1)}$ for $i = -1, 0, 1, 2$ drawn as train tracks for simplicity.

where h_1 and $h_i^{(n)}$ for $i = 2, 3$ are permissible shifts defined as follows. Let h_1 be the right shift whose domain includes all simple closed curves B_j for $j \in \mathbb{Z} \setminus \{0\}$ and passes under p and B_0. Let $h_2^{(n)}$ be the left shift whose domain includes only $\{B_j \mid j \in (-\infty, -n-1] \cup [n+1, \infty)\}$, passes over p and B_0, and, when $j \in [-n, -1] \cup [1, n]$, passes under B_j for j odd and over B_j for j even. Finally, let $h_3^{(n)}$ be the right shift whose domain includes only $\{B_j \mid j \in (-\infty, -n-1] \cup [n+1, \infty)\}$ and passes under p and over B_j for all $j \in [-n, P) \cup (P, n]$. See Figure 27.

In the language of the previous sections, we have that $n_1 = -1$ and $n_2 = 1$ for h_1. For $h_2^{(n)}$, $n_1 = n + 1$ and $n_2 = -n - 1$. For $h_3^{(n)}$, $n_1 = -n - 1$ and $n_2 = n + 1$.

In order to prove that each g_n is loxodromic with respect to the action of $\mathrm{Map}(S, p)$ on $\mathcal{A}(S, p)$, we introduce a collection of arcs $\{\alpha_i^{(n)} \mid i \in \mathbb{Z}\}$ which are invariant under g_n. We will show in Proposition 8.2 that for $i \geq 0$, these arcs form a quasigeodesic half-axis for g_n in $\mathcal{A}(S, p)$.

Definition 5.2 For each fixed n, we define a sequence of arcs $\{\alpha_i^{(n)}\}_{i=-\infty}^{\infty}$ in $\mathcal{A}(S, p)$ as the images under successive applications of g_n of the fixed initial arc $\alpha_0 = P_s 0_o 0_u P_s$. That is to say that $\alpha_i^{(n)} = g_n^i(\alpha_0)$ for any $i \in \mathbb{Z}$. See Figure 28 for the case $n = 1$ and Figure 29 for general $n \geq 1$.

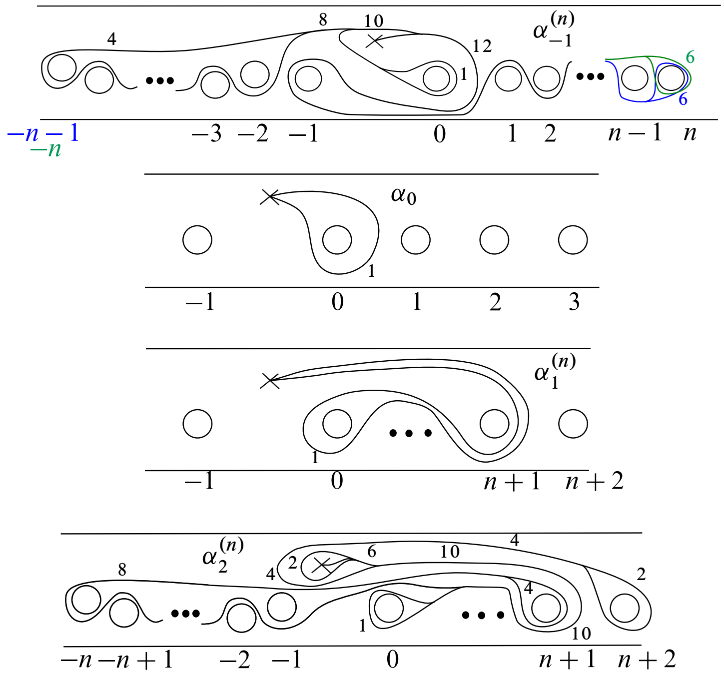


Figure 29: The arcs $\alpha_i^{(n)}$ for $i=-1,0,1,2$ and $n\geq 1$. In $\alpha_{-1}^{(n)}$, the blue appears when n is odd, while the green appears when n is even.

When $n=1$, the arcs $\alpha_i^{(1)}$ are the most straightforward and thus most useful for building intuition. We suggest that the reader keep these arcs in mind while reading the remainder of the paper. We now give the code for $\alpha_1^{(1)}$ and $\alpha_2^{(1)}$, which the reader can compare to Figure 28:

$$\begin{aligned}
\alpha_1^{(1)} &= P_s 0_o 1_o 2_o 2_u 1_o 0_o 0_u 1_o 2_u 2_o 1_o 0_o P_s,\\
\alpha_2^{(1)} &= P_s 0_o 1_o 2_o 2_u 1_o 0_o P_u (-1)_u (-1)_o P_u 0_o 1_o 2_o 2_u 1_o 0_o P_u (-1)_o (-1)_u P_u 0_o 1_o 2_u 2_o 1_o\\
&\qquad 0_o P_u P_o 0_o 1_o 2_o 3_o 3_u 2_o 1_o 0_o P_o P_u 0_o 1_o 2_o 2_u 1_o 0_o P_u (-1)_u (-1)_o P_u 0_o 1_o 2_u 2_o 1_o 0_o\\
&\qquad P_u (-1)_o (-1)_u P_u 0_o 1_o 2_u 2_o 1_o 0_o P_u P_o 0_o 1_o 2_o 2_u 1_o 0_o 0_u 1_o 2_u 2_o 1_o 0_o P_o P_u 0_o 1_o 2_o\\
&\qquad 2_u 1_o 0_o P_u (-1)_u (-1)_o P_u 0_o 1_o 2_o 2_u 1_o 0_o P_u (-1)_o (-1)_u P_u 0_o 1_o 2_u 2_o 1_o 0_o P_u P_o 0_o\\
&\qquad 1_o 2_o 3_u 3_o 2_o 1_o 0_o P_o P_u 0_o 1_o 2_o 2_u 1_o 0_o P_u (-1)_u (-1)_o P_u 0_o 1_o 2_u 2_o 1_o 0_o P_u (-1)_o (-1)_u\\
&\qquad P_u 0_o 1_o 2_u 2_o 1_o 0_o P_s.
\end{aligned}$$

While it is possible to compute the images of arcs under $g_n = h_3^{(n)} \circ h_2^{(n)} \circ h_1$ by hand, we have also written a computer program to implement this. The interested reader should contact the authors for more details.

Remark 5.3 In light of how complicated the arcs $\alpha_i^{(n)}$ are when $i \neq 0,1$, it may be surprising that only single loops can have trivial image under each of the shifts $h_1, h_2^{(n)}$, and $h_3^{(n)}$ (see Theorem 4.8 and Corollary 4.9). Figure 30 gives the intermediate steps so that one can see how the image under g_1 of a complicated arc such as $\alpha_{-1}^{(1)}$ becomes the straightforward arc α_0. The figure shows the collection of single loops with trivial image at each stage of computing $g_1(\alpha_{-1}^{(1)}) = \alpha_0$. A similar phenomenon happens for $n>1$.

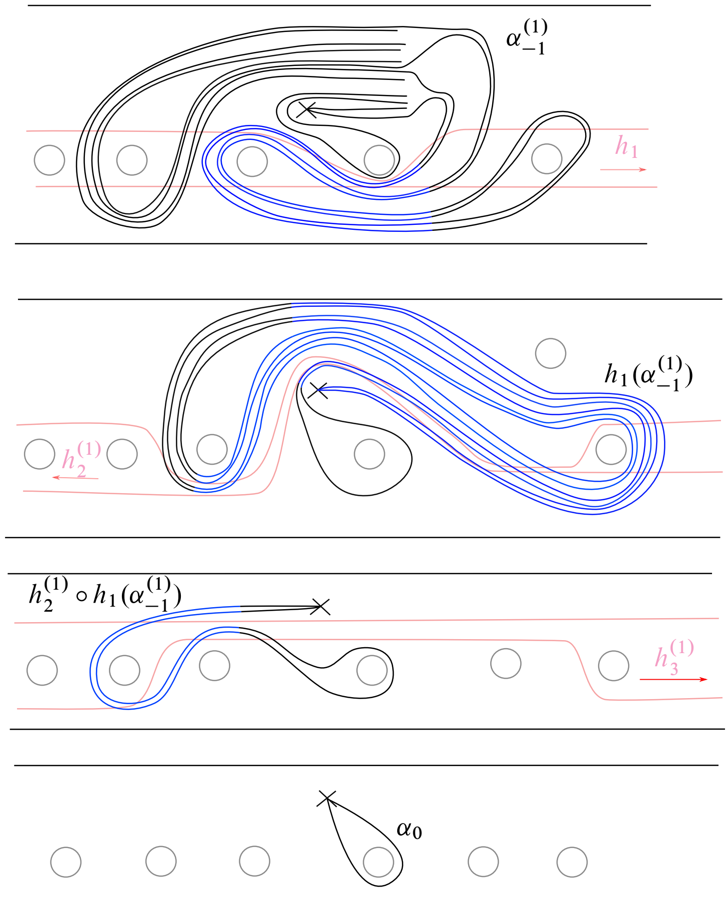


Figure 30: Each of the strands in blue are single loops whose image is trivial under the shift whose domain is shown in red. Each blue strand fits the form of Theorem 4.8.

Our method for using the arcs $\alpha_i^{(n)}$ to prove that g_n is a loxodromic isometry of $\mathcal{A}(S, p)$ is inspired by the foundational work of Bavard in [5]. Bavard constructs an element f of the mapping class group of the plane minus a Cantor set which is loxodromic with respect to the associated relative arc graph. We give a brief outline of Bavard's methods here.

Bavard constructs a collection of simple paths β_i which start at an isolated puncture and end at some point of the Cantor set. These paths are invariant under the action of a chosen homeomorphism f and have the property that $\beta_{i+1} = f(\beta_i)$. They are constructed so that for all i, the path β_{i+1} begins by following the same path as β_i. Roughly, if a path γ begins by following the same path as β_i then Bavard says the path γ "begins like" β_i (see [5, Section 2]). Bavard uses this definition to define a function from the vertex set of a certain graph of paths (defined similarly to the relative arc graph but with paths instead of arcs) to $\mathbb{Z}_{\geq 0}$ by sending a path γ to the maximal $i \in \mathbb{Z}_{\geq 0}$ so that γ begins like β_i. Aramayona, Fossas, and Parlier [2] show that the relative arc graph is quasi-isometric to the graph whose vertex set is isotopy class of paths with *at least one* endpoint on the distinguished isolated puncture. Therefore this function can then be used to estimate distances in the relative arc graph and ultimately to show that the collection of paths $\{\beta_i \mid i \in \mathbb{Z}_{\geq 0}\}$ forms a geodesic half-axis for the element f. The key fact that Bavard uses is that if δ is a path which begins like β_i and γ is any path disjoint from δ, then γ begins like β_{i-1} [5, Lemma 2.4].

Our arcs $\{\alpha_i^{(n)}\}_{i\in\mathbb{Z}}$ do not satisfy the same property as Bavard's paths. One notable difference is that our arcs start and end at the puncture p. Because of this, for a fixed n, $\alpha_{i+1}^{(n)}$ does not begin by following *the entirety* of $\alpha_i^{(n)}$. However, we will show that it does begin by following the *first half* of $\alpha_i^{(n)}$. In light of this, we modify Bavard's notion of "begins like" and make the following definition.

Definition 5.4 Fix an $n \in \mathbb{N}$. An arc δ *starts like* $\alpha_i^{(n)}$ if the maximal initial or the maximal terminal segment of δ which agrees with an initial or terminal segment of $\alpha_i^{(n)}$ (not necessarily respectively) has code length at least $\lfloor \frac{1}{2}\ell_c(\alpha_i^{(n)}) \rfloor$. Recall that code length was defined in Definition 3.6.

In Section 6, we investigate the properties of arcs that start like $\alpha_i^{(n)}$ for some i, and in Section 7 we use these properties to prove Theorem 7.10, which is an analog of [5, Lemma 2.4]. In our more general context, proving this result is quite a bit more involved than the proof of [5, Lemma 2.4]. One reason for this is that the behavior of our arcs $\alpha_i^{(n)}$ is much more complicated than the paths from [5].

6 Arcs that start like $\alpha_i^{(n)}$

The main goal of this section is to prove that the images of arcs which start like $\alpha_i^{(n)}$ under g_n must start like $\alpha_{i+1}^{(n)}$ (see Corollary 6.9). This will follow from Proposition 6.2 and Theorem 6.4. Proposition 6.2 states that the image of the first half of $\alpha_i^{(n)}$ is the first half of $\alpha_{i+1}^{(n)}$. We need to be careful about what we mean by this because of the possibility that we cause there to be a gap when we break $\alpha_i^{(n)}$ into its first and second half (see Section 3.3). Given a segment γ, let $\mathring{\gamma}$ be the initial subsegment of γ with code length $\lfloor \frac{1}{2}\ell_c(\gamma) \rfloor$. For a fixed i and n, we have $\alpha_i^{(n)} = \mathring{\alpha}_i^{(n)} + \alpha'$ for some α'. As in Section 3.3, it is possible that when we put $\mathring{\alpha}_i^{(n)}$ and α' (or their images) into standard position individually with respect to one of the shifts $h_1, h_2^{(n)}$, or $h_3^{(n)}$, we will lose information when we take its image. In fact, this does happen when we put $h_2^{(n)}(h_1(\mathring{\alpha}_i^{(n)}))$ and $h_2^{(n)}(h_1(\alpha'))$ into standard position with respect to $h_3^{(n)}$ (see Figure 31). We avoid this issue by extending $\mathring{\alpha}_i^{(n)}$ by one character and considering $\mathring{\alpha}_i^{(n)}0_u$. Note that 0_u is fixed by g_n. Thus when we say that the image of the first half of $\alpha_i^{(n)}$ is the first half of $\alpha_{i+1}^{(n)}$ we precisely mean that $g_n(\mathring{\alpha}_i^{(n)}0_u) = \mathring{\alpha}_{i+1}^{(n)}0_u$ as a subsegment of $g_n(\alpha_i^{(n)}) = \alpha_{i+1}^{(n)}$, in a reduced code. In other words, g_n sends everything before the central 0_u in $\alpha_i^{(n)}$ to $\mathring{\alpha}_{i+1}^{(n)}$.

Remark 6.1 When dealing with the arcs $\alpha_i^{(n)}$, the floor function in $\mathring{\alpha}_i^{(n)}$ is actually unnecessary, as $\ell_c(\alpha_i^{(n)})$ will always be even with the central two characters being 0_o0_u. However, at this point in the paper we have not proven this fact, which is a consequence of Proposition 6.2, so we will not use it in what follows.

Proposition 6.2 *For any n and any i, we have $g_n(\mathring{\alpha}_i^{(n)}0_u) = \mathring{\alpha}_{i+1}^{(n)}0_u$ in a reduced code.*

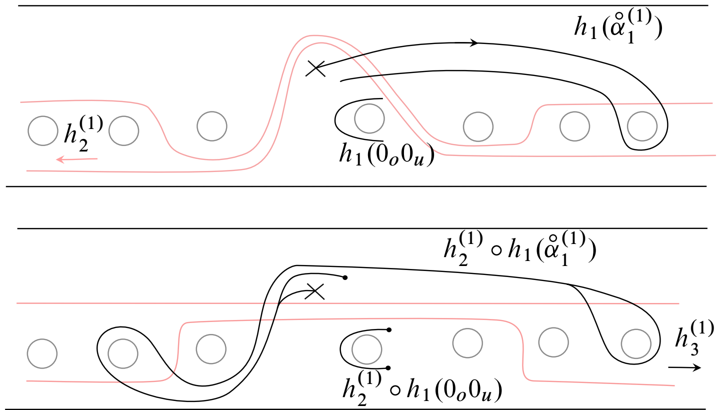


Figure 31: Here we consider $h_1(\mathring{\alpha}_1 0_u) = h_1(\mathring{\alpha}_1) + h_1(0_o0_u)$ and take the images of each subsegment separately. The images are drawn as train tracks for clarity, but we omit the weights as they are immaterial. When we put the images $h_2^{(n)}(h_1(\mathring{\alpha}_1))$ and $h_2^{(n)}(h_1(0_o0_u)) = 0_o0_u$ in standard position individually with respect to $h_3^{(n)}$ (the second picture), there is a gap in the segment between p and B_1, which would have been a full crossing. This results in a loss of information in the image under $h_3^{(n)}$.

In order to prove Proposition 6.2, we need to understand how to control cascading cancellation for our homeomorphisms g_n (see Section 4.1 for the definition and a discussion of cascading cancellation). The following lemma is almost a direct consequence of Lemma 4.10, the difference being that g_n is not itself a shift.

Lemma 6.3 *For all $k \geq n+1$, if $k_{o/u}k_{u/o}$ appears in a reduced code for $\alpha_i^{(n)}$, then there is no cancellation involving $g_n(k_{o/u}k_{u/o})$ in an unreduced code for $\alpha_{i+1}^{(n)}$. In particular, such a pair $k_{o/u}k_{u/o}$ will yield $(k+1)_{o/u}(k+1)_{u/o}$ in a reduced code for $\alpha_{i+1}^{(n)}$.*

Proof The union of the turbulent regions for h_1, $h_2^{(n)}$, and $h_3^{(n)}$ is $[-n-1, n+1]$. Therefore we may apply Lemma 4.10 three times once we remark that $h_2^{(n)}$ shifts left, $h_1(k_{o/u}k_{u/o}) = (k+1)_{o/u}(k+1)_{u/o}$, and $h_2^{(n)}$ fits the hypothesis of Lemma 4.10, as $k+1 > n+1$. □

We can now use this control on cascading cancellation to prove Proposition 6.2.

Proof of Proposition 6.2 As $\alpha_i^{(n)}$ is symmetric, the only way this proposition could be false is if there is cancellation in $g_n(\alpha_i^{(n)})$ involving the image of the central 0_o0_u (which is also 0_o0_u) in $\alpha_i^{(n)}$. Equivalently, we need to show that there is no cancellation involving the final 0_o0_u in $g_n(\mathring{\alpha}_i^{(n)}0_u)$. We begin by considering the case $i = 1$, where

$$\mathring{\alpha}_1^{(n)}0_u = P_s 0_o 1_o 2_o \dots (n+1)_o (n+1)_u n_o (n-1)_o \dots 2_o 1_o 0_o 0_u, \tag{4}$$

and all of the characters that are not displayed are k_o for the appropriate k. By Lemma 6.3, there is no cancellation in $g_n(\alpha_1^{(n)})$ involving $g_n((n+1)_o(n+1)_u)$, and so there can be no cascading cancellation involving both the final $0_o 0_u$ and the image of any character before $(n+1)_o(n+1)_u$ in (4). Thus it suffices to consider $g_n(\mu 0_u)$, where

$$\mu = (n+1)_o(n+1)_u n_o(n-1)_o \dots 2_o 1_o 0_o.$$

A direct computation yields, in a reduced code,

$$g_n(\mu 0_u) = (n+2)_o(n+2)_u(n+1)_o n_o \dots (n-1)_o n_o \mu 0_u, \tag{5}$$

and from this computation we can see that there is no cancellation in $g_n(\mathring{\alpha}_1^{(n)})$ involving the final characters, $0_o 0_u$. Thus $g_n(\mathring{\alpha}_1^{(n)} 0_u) = \mathring{\alpha}_2^{(n)} 0_u$ in a reduced code.

By (5), we also see that $g_n(\mu 0_u)$ also ends in $\mu 0_u$. Hence, to show that the result holds for any index i, we simply repeatedly apply g_n and Lemma 6.3 as in the previous paragraph. □

We can deduce from Proposition 6.2 that if an arc γ starts like $\alpha_i^{(n)}$, then an *unreduced code* for its image $g_n(\gamma)$ will start like $\alpha_{i+1}^{(n)}$. We will use the following technical theorem to show that this is true for a *reduced* code for $g_n(\gamma)$, as well.

Theorem 6.4 *Let γ be a simple arc of the form $\gamma = P_s \eta_1 0_o \eta_2 P_s$ in standard position. Assume the following two conditions hold*:

(1) *The numerical value of η_2^i is at most 0, ie η_2^i is either 0_u or $P_{o/u}$.*

(2) *Either the first two characters of η_2 are not a loop around P or*, *if they are*, *the initial segment of η_2 is given by $P_u P_o 0_o 1_o$.*

Then there is no cancellation with the initial 0_o in a reduced code for $g_n(0_o \eta_2 P_s)$.

We note that the first condition means that the segment 0_o is oriented to the left.

Remark 6.5 Theorem 6.4 is written with a particular orientation in mind, but such an orientation is arbitrary. That is to say, the exact same statement is true applied to the image of $\overline{\gamma}$ under g_n. We will use both the original and this "reverse" version of Theorem 6.4 later in the paper, so we make its statement precise: Suppose, as in Theorem 6.4, that $\gamma = P_s \eta_1 0_o \eta_2 P_s$ and

(1) the numerical value of η_1^t is at most 0;

(2) either the final two characters of η_1 are not a loop around P or the final segment of η_1 is given by $1_o 0_o P_o P_u$.

Then there is no cancellation with the terminal 0_o in a reduced code for $g_n(P_s \eta_1 0_o)$.

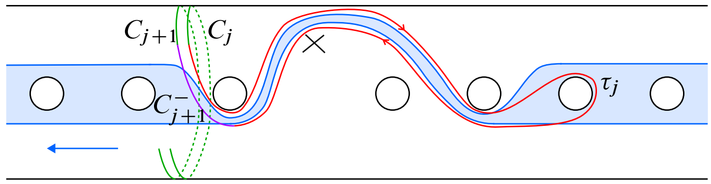


Figure 32: Situation in the proof of Lemma 6.6.

We will prove Theorem 6.4 through a series of lemmas. Fix an arc γ which satisfies the hypothesis of Theorem 6.4. We must show that there is no cancellation with the initial 0_o in a reduced code for $g_n(0_o\eta_2 P_s)$. For contradiction, suppose there is cancellation with the initial 0_o in a reduced code for $g_n(0_o\eta_2 P_s)$. Consider the subsegment of γ given by $0_o\delta$, where $0_o\delta$ is the minimal subsegment such that $g_n(0_o\delta)$ has cancellation with 0_o.

Lemma 6.6 *The segment δ does not contain any back loops.*

Proof Suppose towards a contradiction that δ contains at least one back loop and write δ in standard form as

$$\delta = \tau_1 C_1^- C_1 C_1^+ \tau_2 \dots \tau_s C_s^- C_s C_s^+ \tau_{s+1},$$

where the τ_i are possibly empty for $i \le s+1$. Then

$$g_n(\delta) = g_n(\tau_1 C_1^-) C_1 g_n(C_1^+ \tau_2 C_2^-) C_2 \dots g_n(C_{s-1}^+ \tau_s C_s^-) C_s g_n(C_s^+ \tau_{s+1}).$$

In order to have cascading cancellation involving 0_o, there must be an instance of 0_o in $g_n(C_s^+\tau_{s+1})$ and the initial subsegment of $g_n(\delta)$ which ends immediately before that 0_o must have trivial image. By the minimality of our choice of δ, this initial subsegment must contain $g_n(C_i) = C_i$ for all $i = 1, \dots, s$. Since its image is trivial, there must be cancellation involving each C_i. The only way that this can occur is if there is some j such that $C_j = \overline{C}_{j+1}$ and $C_j g_n(C_j^+ \tau_j C_{j+1}^-)\overline{C}_j = \varnothing$. Consequently, we must have $g_n(C_j^+\tau_j C_{j+1}^-) = \varnothing$. This implies that exactly one of the following holds:

(1) $h_1(C_j^+\tau_j C_{j+1}^-) = \varnothing$.

(2) $h_1(C_j^+\tau_j C_{j+1}^-) \ne \varnothing$ in a reduced code while $h_2^{(n)}(h_1(C_j^+\tau_k C_{j+1}^-)) = \varnothing$.

(3) $h_2^{(n)}(h_1(C_j^+\tau_j C_{j+1}^-)) \ne \varnothing$ in a reduced code while $h_3^{(n)}\big(h_2^{(n)}(h_1(C_j^+\tau_j C_{j+1}^-))\big) = \varnothing$.

Let i correspond to which of the above cases we are in and let $h = h_i^{(n)}$. Define

$$\tau_j' = \begin{cases} C_j^+\tau_j C_{j+1}^-, & h = h_1, \\ h_1(C_j^+\tau_j C_{j+1}^-), & h = h_2^{(n)}, \\ h_2^{(n)}(h_1(C_j^+\tau_j C_{j+1}^-)), & h = h_3^{(n)}. \end{cases}$$

By assumption, $\tau_j' \ne \varnothing$ and $h(\tau_j') = \varnothing$.

By Theorem 4.8, this implies that τ_j' is a loop which begins and ends in the turbulent region and does not fully cross D in (n_1, n_2). Moreover, in δ, the path τ_j' is preceded by C_j and followed by $C_{j+1} = \overline{C}_j$ and so one of the back loop connectors C_j^+ or C_{j+1}^- must fully cross D. However this is a contradiction, since these back loop connectors must occur in (n_1, n_2). See Figure 32. □

Notice that 0_o does not appear in the image under g_n of any segment supported on

$$(-\infty, -n-1) \cup [n+1, \infty).$$

Therefore, if δ contains any subsegment supported in this region then, by minimality, it must also contain a subsegment of the form $q_{o/u} q_{u/o}$ for $q \geq n+1$ or $q < -n-1$; in other words, the segment δ must "turn around" in the shift region. Moreover, by Lemma 6.6, δ can have no back loops, and hence Lemma 4.10 implies that such a $q_{o/u} q_{u/o}$ blocks cascading cancellation, so this cannot occur. Thus, we conclude that a reduced code for $0_o \delta$ is supported on $[-n-1, n+1)$.

We will now show that there is no cancellation involving 0_o under each of $h_1, h_2^{(n)}$, and $h_3^{(n)}$. By the assumptions of the theorem, the 0_o in $0_o\delta$ is oriented to the left. If there is no cancellation involving 0_o under a shift, then 0_o persists in the image and is still oriented to the left. This implies that the character after 0_o in the image still satisfies condition (1) of the theorem, that is, it is either 0_u or $P_{o/u}$.

A straightforward calculation shows that $h_1(0_o\delta)$ cannot cancel 0_o for any δ. This follows from the fact that ∂D_1 contains 0_u, not 0_o, where D_1 is the domain for h_1. Write $0_o\delta'$ for the reduced code of $h_1(0_o\delta)$, which is another segment completely contained in the region $(-n-1, n+1]$ with no back loops. The first character of δ' is still either 0_u or $P_{o/u}$.

It remains to rule out any cancellation involving 0_o under $h_2^{(n)}$ and $h_3^{(n)}$. We do this in the following two lemmas.

Lemma 6.7 *There is no cancellation involving 0_o under $h_2^{(n)}$.*

Proof For contradiction, assume that there is cancellation involving 0_o, and let ζ be the minimal subsegment of $0_o\delta'$ which has cancellation with 0_o under $h_2^{(n)}$.

If $\zeta^t = 0_o$, then ζ is a loop whose image is trivial. By Theorem 4.8, ζ must be of the form

$$\partial D_2|_{[0,n_1)} (n_1)_{o/u} (n_1)_{u/o} \overline{\partial D_2|_{[0,n_1)}},$$

where D_2 is the domain of $h_2^{(n)}$. In particular, $\zeta^i = 0_o$ and this 0_o is oriented to the right. However, as noted above, the initial 0_o of $0_o\delta$ is oriented to the left, which contradicts that ζ is an initial subsegment of $0_o\delta'$.

Therefore, $\zeta = 0_o \zeta' a_1 a_2$, where the 0_o which cancels with $\zeta^i = 0_o$ appears in $h_2^{(n)}(a_1 a_2)$ and is not the terminal character. This is the case exactly when $a_1 a_2$ is either a full or half crossing. Moreover, in order for 0_o to appear in the image of $a_1 a_2$, we must have that the numerical values of a_1 and a_2 are greater than or equal to zero, since $h_2^{(n)}$ shifts to the left.

There are two cases to consider: either $(\zeta')^i = 0_u$ or $(\zeta')^i = P_{o/u}$. If $(\zeta')^i = 0_u$, then ζ' contains a character which disagrees with ∂D_2, namely 0_u. On the other hand, if $(\zeta')^i = P_{o/u}$, then it must be oriented to the left. Since the numerical values of a_1 and a_2 are at least 0 and $P < 0$, ζ' must "turn around" and contain an over-under loop in turbulent region. In particular, it must again contain a character which disagrees with ∂D_2. In either case, call the character which disagrees with the boundary of the domain b.

By Lemma 4.11, b persists in a reduced code for $h_2^{(n)}(\zeta')$ and such a b must block any cascading cancellation. However, by assumption, there is cancellation between two instances of 0_o in an *unreduced code* for $h_2^{(n)}(\zeta')$, one which precedes b and one which succeeds b. This is a contradiction. □

It remains only to check for cancellation under $h_3^{(n)}$.

Lemma 6.8 *There is no cancellation involving 0_o under $h_3^{(n)}$.*

Proof The argument is similar to the proof of Lemma 6.7. For contradiction assume that there is cancellation involving 0_o and write $\beta = 0_o\delta''$ for the minimal subsegment of a reduced code for $h_2^{(n)}(0_o\delta')$ which has cancellation with 0_o under $h_3^{(n)}$. Note that $(\delta'')^i$ is still either 0_u or $P_{o/u}$. By the same reasoning as in the proof of Lemma 6.7, $\beta = 0_o\beta' c_1 c_2$, and c_1c_2 is a full or half crossing. Recall that for the shift $h_3^{(n)}$, $n_1 = -n-1$, $n_2 = n+1$, and $h_3^{(n)}$ shifts to the right.

If $\beta' c_1$ contains a character which does not agree with ∂D_3, then the contradiction follows by applying Lemma 4.11 as in the proof of Lemma 6.7. However, it is possible that $\beta' c_1$ does not contain such a character. This occurs exactly when $\beta = \overline{\partial D_3|_{(n_1,0]}}(n_1)_{o/u}$. This possibility did not arise when considering h_2 because h_2 is a left shift, and so $n_1 > 0$.

In this case, β must begin with $0_o P_u (-1)_o$. We will now show that β could not be the image of $0_o\delta$ under $h_2^{(n)} \circ h_1$. In the notation above, $(-1)_o$ is in the image of $\xi = 0_o\delta'$ under $h_2^{(n)}$. Since $(-1)_o$ disagrees with ∂D_2, there must be an instance of $(-1)_o$ in δ' by Claim 4.12. Let $\xi' = 0_o\sigma(-1)_o$ be the subsegment of ξ so that $h_2^{(n)}(\xi') = 0_o P_u (-1)_o$. We can find ξ directly by computing $(h_2^{(n)})^{-1}(0_o P_u(-1)_o)$. Recall that we cannot always take the preimage of a segment under a permissible shift (see the discussion in Section 3.4). However, in this case we are able to take the preimage precisely because Claim 4.12 ensures that $0_o P_u (-1)_o$ is the image of a subsegment of $h_2^{(n)}(\xi)$. Computing $(h_2^{(n)})^{-1}(0_o P_u(-1)_o)$ results in

$$\xi' = 0_o P_u \partial D_2|_{[P,n+1)}(n+1)_u(n+1)_o\overline{\partial D_2|_{[P,n+1)}}(-1)_o.$$

In particular, ξ' starts with $0_o P_u P_o 0_o 1_u$. To conclude the proof, we will show that ξ' cannot occur as the image under h_1 of the segment $0_o\delta$.

Assume for contradiction that $h_1(0_o\delta) = \xi'$. Since P_o does not agree with ∂D_1, there must be an instance of P_o in δ by Claim 4.12. Let $0_o\tau P_o$ be the initial subsegment of $0_o\delta$ so that $h_1(0_o\tau P_o) = 0_o P_u P_o$. As above, we may directly compute the preimage of $0_o P_u P_o$ under h_1. This shows that $0_o\tau P_o = 0_o P_u P_o$. The assumptions of the theorem then imply that $0_o\delta$ starts with $0_o P_u P_o 0_o 1_o$, where 1_o is oriented to the right. Since ξ' starts with $0_o P_u P_o 1_u$, the only way $h_1(0_o\delta) = \xi'$ is if there is cancellation with the terminal 1_o in $0_o P_u P_o 0_o 1_o$ when we apply h_1 to $0_o\delta$. Since h_1 shifts to the right and ξ' contains $(-1)_o$, it must be the case that $0_o\delta$ contains a loop $k_{o/u}k_{u/o}$ with $k \geq 1$. In particular, $0_o\delta$ starts with either $0_o P_u P_o 0_o 1_o 1_u$ (if $k = 1$) or $0_o P_u P_o 0_o 1_o \eta k_{o/u}k_{u/o}$ where η is strictly monotone increasing (if $k > 1$). Since $n_2 = 1$ for h_1, Lemma 4.10 then implies that this $k_{o/u}k_{u/o}$ blocks cancellation between the terminal 1_o in $0_o P_u P_o 0_o 1_o$ and any characters in $0_o\delta$ that appear *after* $k_o k_u$. Since the segment η of $0_o\delta$ between 1_o and $k_o k_u$ is either empty (if $k = 1$) or a strictly monotone increasing segment (if $k > 1$), using the fact that h_1 shifts to the right, it is straightforward to check that there can be no cancellation with the terminal 1_o in $0_o P_u P_o 1_o$ when we apply h_1, which is a contradiction. This completes the proof. □

Lemmas 6.6, 6.7, and 6.8 show that $g_n(0_o\delta)$ has no cancellation with 0_o, which completes the proof of Theorem 6.4.

As promised at the beginning of this section, we have the following corollary.

Corollary 6.9 *For any fixed n and any $i \geq 0$, $\alpha_{i+1}^{(n)}$ starts like $\alpha_i^{(n)}$. More generally, for any $i \geq 1$, if γ starts like $\alpha_i^{(n)}$ then $g_n(\gamma)$ starts like $\alpha_{i+1}^{(n)}$.*

Proof First, a direct computation shows that for any fixed n, $\alpha_1^{(n)}$ starts like $\alpha_0^{(n)}$. For all $i \geq 1$, we will prove both statements simultaneously, using strong induction. We will show that for each n and any $i \geq 1$, if γ is a simple arc that starts like α_i, then $g_n(\gamma) = \mathring{\alpha}_{i+1}^{(n)}\eta'$ for some reduced η' which has the form of η_2 in the hypotheses of Theorem 6.4.

For the base case $i = 1$, suppose γ starts like $\alpha_1^{(n)}$ so that we may write

$$\gamma = \mathring{\alpha}_1^{(n)}\eta,$$

in a reduced code. Applying g_n to both sides of this equation, we have, in an *unreduced* code,

$$g_n(\gamma) = g_n(\mathring{\alpha}_1^{(n)})\eta',$$

for some reduced η'. By Proposition 6.2, $g_n(\mathring{\alpha}_1^{(n)}) = \mathring{\alpha}_2^{(n)}$ in a reduced code. Thus to conclude that $g_n(\gamma)$ starts like $\alpha_2^{(n)}$, we need to show that $g_n(\mathring{\alpha}_1^{(n)}) = \mathring{\alpha}_2^{(n)}$ persists in a reduced code for $g_n(\gamma)$. Since $\mathring{\alpha}_1^{(n)}$ ends with 0_o, it suffices to show that η has the form of η_2 in the hypotheses of Theorem 6.4 so that there is no cancellation with this 0_o. We have $\mathring{\alpha}_1^{(n)} = P_s 0_o 1_o 2_o \ldots (n+1)_o (n+1)_u n_o (n-1)_o \ldots 2_o 1_o 0_o$, and so the terminal 0_o is oriented to the left. It follows that η satisfies Theorem 6.4(1). Moreover, since γ is simple and starts like $\alpha_1^{(n)}$, if η begins with a loop around p, it must begin with $P_u P_o 0_o 1_o$ (see Figure 33). Thus η satisfies Theorem 6.4(2) and so Theorem 6.4 shows that there is no cancellation in $g_n(\gamma)$ between the final 0_o in $g_n(\mathring{\alpha}_1^{(n)}) = \mathring{\alpha}_2^{(n)}$ and η'. Therefore, $g_n(\gamma)$ starts like $\alpha_2^{(n)}$. In particular, a direct computation shows that $\alpha_2^{(n)}$ starts like $\alpha_1^{(n)}$ and so by setting $\gamma = \alpha_2^{(n)}$, we have shown that $g_n(\alpha_2^{(n)}) = \alpha_3^{(n)}$ starts like $\alpha_2^{(n)}$.

To finish the base case, we will now show that η' has the form of η_2 in the hypotheses of Theorem 6.4. A direct computation shows that $\mathring{\alpha}_2^{(n)}$ ends with $(n+1)_o(n+1)_u n_o (n-1)_o \ldots 2_o 1_o 0_o$, and so that η'

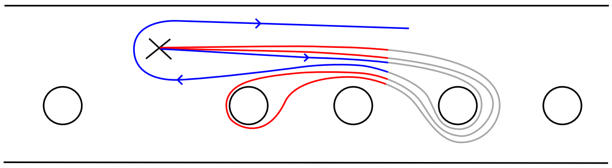


Figure 33: The arc $\alpha_1^{(n)}$ is in red; the gray portions may vary depending on n. One possibility for γ and the subsegment η is shown in blue; the gray portions may vary depending on n.

satisfies Theorem 6.4(1). As above, since $g_n(\gamma) = \mathring{\alpha}_2^{(n)}\eta'$ is simple and $\alpha_2^{(n)}$ (and therefore $\mathring{\alpha}_2^{(n)}$) starts like $\alpha_1^{(n)}$, it follows that if η' starts with a loop around p, it must begin with $P_u P_o 0_o 1_o$, so η' satisfies Theorem 6.4(2). This completes the base case.

Now assume the results hold for all $j < i$. For the induction step, suppose that $\gamma = \mathring{\alpha}_i^{(n)}\eta$ for some η which has the form of η_2 in the hypotheses of Theorem 6.4. The argument for this case goes through exactly as in the base case except we need one additional argument to show that η' has the form of η_2 in the hypotheses of Theorem 6.4, where $g_n(\gamma) = \mathring{\alpha}_{i+1}^{(n)}\eta'$ in a reduced code. By the proof of Proposition 6.2, $\mathring{\alpha}_{i+1}^{(n)}$ ends with $\mu = (n+1)_o(n+1)_u n_o(n-1)_o \dots 2_o 1_o 0_o$ and therefore η' satisfies Theorem 6.4(1). To see that η' satisfies Theorem 6.4(2), notice that $\alpha_{i+1}^{(n)}$ starts like $\alpha_i^{(n)}$, which starts like $\alpha_{i-1}^{(n)}$, etc, so that $\alpha_{i+1}^{(n)}$ (and therefore $\mathring{\alpha}_{i+1}^{(n)}$) starts like $\alpha_1^{(n)}$. In particular, γ must start like $\alpha_1^{(n)}$. Since γ is simple, if η' starts with a loop around p, it must begin with $P_u P_o 0_o 1_o$ and Theorem 6.4(2) is satisfied. This completes the induction step, and the result is proved. □

7 Highways in arcs

In this section, we introduce and examine the prevalence of certain segments of the code for $\alpha_i^{(n)}$ that we call *highways*. The presence of highways forces arcs disjoint from $\alpha_i^{(n)}$ to have very specific initial and terminal subsegments. This will be instrumental in proving Theorem 7.10, which shows that if δ is any arc which starts like $\alpha_i^{(n)}$ and γ is an arc disjoint from δ, then γ starts like $\alpha_{i-1}^{(n)}$, provided i is large enough. In Section 8, we will use Theorem 7.10 to show that the arcs $\alpha_i^{(n)}$ lie on a quasigeodesic in the modified arc graph.

In Section 7.1, we will give general preliminary definitions and results for general arcs, and then in Section 7.2 we will analyze highways in the arcs $\alpha_i^{(n)}$.

7.1 Preliminaries on highways

Recall our convention that all arcs are assumed to be simple and start and end at p.

Definition 7.1 Given an arc $\delta = P_s q_1 \eta q_2 P_s$, where q_1, q_2 are single characters which are not C and η is a segment, we say that δ *has highways* if either $q_1 P_{o/u} P_{u/o} q_1$ or $q_2 P_{o/u} P_{u/o} q_2$ is a subsegment of δ.

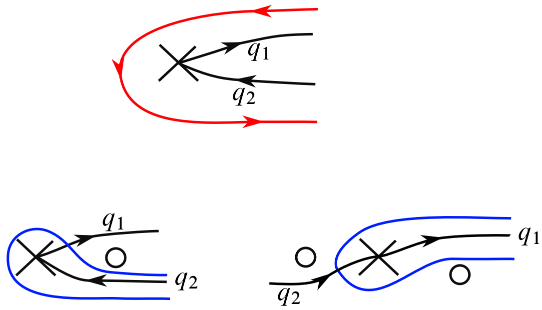


Figure 34: At the top is an arc where $q_1 = q_2$ and the red segment shows that this arc has highways. Below are two arcs where $q_1 \neq q_2$ and the blue segments must appear if the arc has highways. However, this arrangement contradicts the simplicity of each arc (see Lemma 7.2) so neither arc has highways.

The following lemma is an almost immediate corollary of the definition and the fact that δ is simple.

Lemma 7.2 *If δ is an arc that has highways, then*

$$\delta = P_s \tau \dots \overline{\tau} P_s,$$

for some segment τ with $\ell_c(\tau) > 0$. That is to say, the first part of the code for δ always overlaps with the reverse of the last part of the code in at least two characters, one of which is P_s.

Proof It suffices to consider the case when τ is a single character, so that $\tau = a = \overline{\tau}$ for some a. Since $\delta = P_s q_1 \eta q_2 P_s$, if the conclusion does not hold, then $q_1 \neq q_2$. In this case, either the subsegment $P_s q_1$ intersects $q_2 P_o P_u q_2$ or the subsegment $q_2 P_s$ intersects $q_1 P_o P_u q_1$, contradicting the fact that δ is simple. See Figure 34. □

Recall that in the code for an arc, the character P_s does not correspond to any subsegment. Since the first and last characters of every arc are always P_s, we use the first *two* characters of δ in the statement of the above lemma to ensure that there is an initial subsegment of δ which fellow travels a terminal subsegment.

In the future, we will need to use a refined notion of highways to constrain the beginnings of certain arcs. For this we define a notion of right lane and left lane. See Figure 35 for examples and nonexamples.

Definition 7.3 Given an arc δ that has highways, a subsegment λ of δ is called a *left lane* if one of the following holds:

(1) $\lambda = P_o P_u \gamma$, where γ does not contain C and γ is maximal with respect to the property that the code for $P_s \gamma$ coincides with the initial $\ell_c(P_s \gamma)$ many characters in δ and $\overline{\gamma} P_s$ coincides with the terminal $\ell_c(P_s \gamma)$ many characters in δ; or

(2) $\lambda = \overline{\gamma} P_u P_o$ is the reverse of the segment of the form in (1).

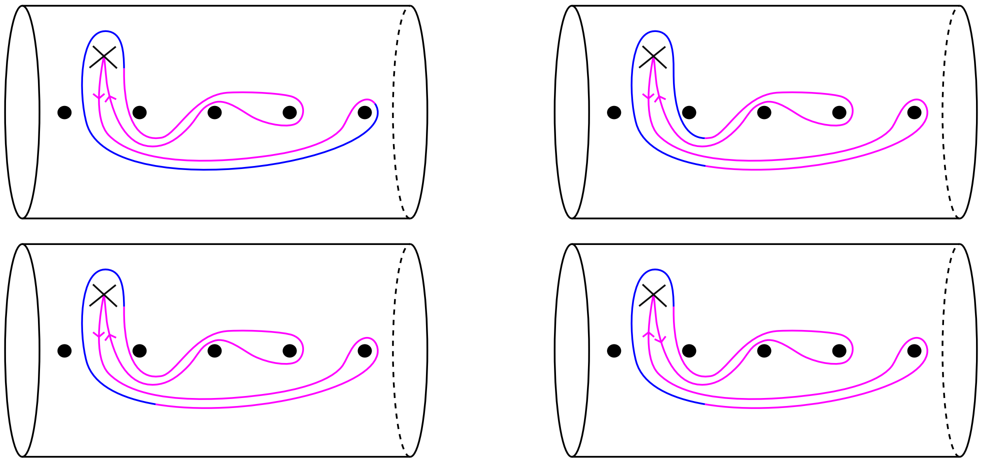


Figure 35: Some examples and nonexamples of left and right lanes are shown in blue. Top left: the blue segment is *not* a left lane since it does not coincide with the terminal subsegment of the arc. Top right: the blue segment contains both left and right lanes. There is a left lane via case (2) of Definition 7.3 and a right lane via case (1) of Definition 7.4. Bottom left: the blue segment contains two left lanes via case (2) of Definition 7.3. Bottom right: the blue segment contains two left lanes via case (1) of Definition 7.3 (notice the orientation on the arc has been changed).

We note that the reason that there are two possibilities for a left lane is that we want the definition to be independent of the orientation of δ.

Definition 7.4 Given an arc δ that has highways, a segment ρ of the code is called a *right lane* if one of the following holds:

(1) $\rho = P_u P_o \gamma$, where γ does not contain C and γ is maximal with respect to the property that the code $P_s\gamma$ coincides with the initial $\ell_c(P_s\gamma)$ many characters in δ and $\overline{\gamma} P_s$ coincides with the terminal $\ell_c(P_s\gamma)$ many characters in δ; or

(2) $\rho = \overline{\gamma} P_o P_u$ is the reverse of the segment from (1).

If δ is a symmetric arc (see Definition 4.4), then in Definitions 7.3 and 7.4 it suffices to check the overlap on just the initial part of the code for δ. However, in the general case, checking the overlap with both the initial and terminal parts of the code for δ is necessary.

Definition 7.5 Let δ be an arc with highways. The *lane length* $L(\xi)$ of a left or right lane ξ of δ is defined to be

$$L(\xi) = \ell_c(\xi) - 1.$$

We denote the collection of all left lanes of δ by $\mathcal{L}$ and similarly of all right lanes of δ by $\mathcal{R}$.

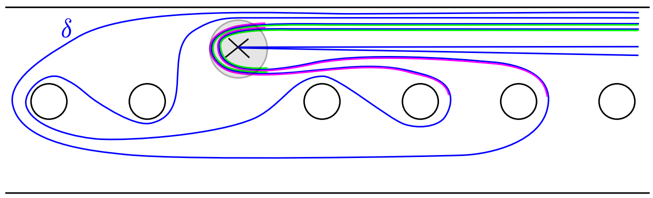


Figure 36: In blue is (a portion) of an arc δ. The shaded disk around the puncture is D_p. This example has two right lanes (in green), a portion of which are shown, and two left lanes (in pink), which are shown in their entirety. The $P_{o/u} P_{u/o}$ portion of these lanes are contained in D_p while all other strands of δ are disjoint from D_p.

Through the rest of the section, we fix a closed topological disc D_p of sufficiently small radius with center at the puncture p such that D_p has empty intersection with each B_i where $i \in \mathbb{Z}$, contains B_P, and has empty intersection with each separating curve $\{S_i\}$ from Section 3.2.1. Moreover, we will only work with homotopy representatives of δ which have reduced code and the property that any pair $P_{o/u} P_{u/o}$ lies inside of D_p, while any other segments remain outside, except for the two that come from the initial and terminal two characters of δ (see Figure 36). Throughout the section, when we further homotope δ we will only do so relative to D_p, so one can assume that the set $\delta \cap D_p$ is pointwise fixed.

A left or right lane ξ is called *innermost* if every oriented straight line segment with initial point at the puncture and terminal point on the boundary circle of D_p intersects the $P_{o/u} P_{u/o}$ at the beginning (resp. end) of ξ before it intersects any other lane of the same type (left or right). If the oriented line segment does not intersect any lane, then this condition is vacuously satisfied. See Figure 37. In particular, innermost left and right lanes are the lanes which are closest to an initial and terminal subsegments of δ.

We then have the following lemma.

Lemma 7.6 *Let δ be an arc that has highways. Let $\lambda \in \mathcal{L}$ and $\rho \in \mathcal{R}$ denote the innermost left and right lanes of δ, respectively. Then $L(\lambda) \geq L(\lambda')$ for all $\lambda' \in \mathcal{L}$ and $L(\rho) \geq L(\rho')$ for all $\rho' \in \mathcal{R}$. Moreover, writing $\lambda = P_u P_o \beta_l$ or its reverse and $\rho = P_o P_u \beta_r$ or its reverse, then if γ is an arc disjoint from δ, then one of the following holds*:

(1) *γ has initial code $P_s \beta_l$ and terminal code $\overline{P_s \beta_l}$,*

(2) *γ has initial code $P_s \beta_r$ and terminal code $\overline{P_s \beta_r}$.*

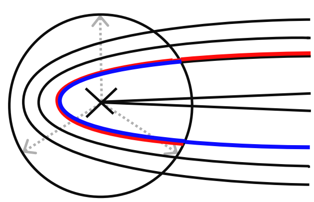


Figure 37: The blue and red segments are the initial portions of innermost left and right lanes, respectively. The dotted gray segments intersect these segments before intersecting the initial or terminal portion of any other lane.

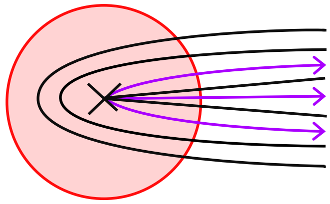


Figure 38: In black is (a portion) of an arc δ. The red disk is D_p. Any arc γ which is disjoint from δ must have initial and terminal segment which follows one of the three purple strands emanating from the puncture and continues to fellow travel δ.

Intuitively, this lemma states that the arc γ must "stay in a lane" of the arc δ; see Figure 38. Note that we do not require γ to be distinct from δ, so this statement applies to δ as well. Indeed, the arc δ can always be homotoped to be disjoint from itself while retaining the same code.

Proof The hypothesis that δ has highways implies that $\mathcal{L}, \mathcal{R} \neq \varnothing$, and hence an innermost left and right lane exist. We prove the first statement for the innermost left lane λ; an identical proof works for ρ.

Since λ is innermost, if there exists a left lane $\lambda' \in \mathcal{L}$ such that $L(\lambda') > L(\lambda)$, then we get a contradiction to the simplicity of δ, as λ must intersect either λ' or an initial segment of δ. See Figure 39.

For the final statement, fix an arc γ that is disjoint from δ. It must be the case that the initial and terminal subsegments of γ are each contained in a single connected component of $D_p \setminus \delta$ (not necessarily the same component). Moreover these initial and terminal subsegments must begin at the puncture p and therefore must also fellow travel λ, ρ, and the initial/terminal parts of δ (see Figure 38). Therefore, by the same reasoning as in the previous paragraph, we see that γ satisfies conclusion (1) of the lemma if $\ell_c(\beta_l) \leq \ell_c(\beta_r)$ and satisfies conclusion (2) of the lemma if $\ell_c(\beta_r) \leq \ell_c(\beta_l)$. □

7.2 Highways for the arcs $\alpha_i^{(n)}$

The main goal of this section is to apply the technology of the previous section to show that any arc disjoint from an arc which starts like $\alpha_i^{(n)}$ starts like $\alpha_{i-1}^{(n)}$, provided i is large enough (see Theorem 7.10).

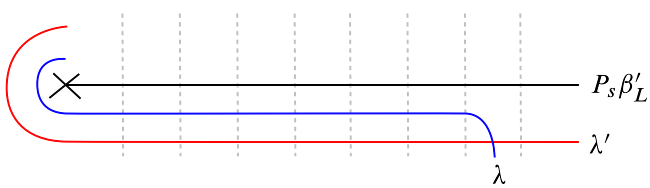


Figure 39: A schematic of the two left lanes: λ, which is innermost, and λ', which is not. If λ' fellow travels an initial segment of δ for longer than λ does, then λ must intersect either λ' (pictured) or the initial segment of δ.

To do this, we will show that $\alpha_i^{(n)}$ has highways for all $i \geq 2$ and all n, beginning by proving the result for $i = 2$. For fixed i and n, define $\chi_i^{(n)}$ to be all of $\mathring{\alpha}_i^{(n)}$ except the initial P_s, so that

$$\mathring{\alpha}_i^{(n)} = P_s \chi_i^{(n)}. \tag{6}$$

Proposition 7.7 *The segment $\alpha_2^{(n)}$ contains a subsegment of the form $\overline{\chi}_1^{(n)} P_u P_o \chi_1^{(n)}$ that appears as the final subsegment of $\mathring{\alpha}_2^{(n)}$. In particular, $\alpha_2^{(n)}$ has highways whose lanes contain $\chi_1^{(n)}$.*

Proof We prove this by carefully examining the code for $\alpha_1^{(n)}$, which we assume is in standard position relative to the disk D_p. As in Section 6, we will retain the 0_u which immediately follows $\mathring{\alpha}_1^{(n)}$ when we compute images. We will show that $\overline{\chi}_1^{(n)} P_u P_o \chi_1^{(n)} 0_u$ appears as the final subsegment of $\mathring{\alpha}_2^{(n)} 0_u$, which then implies the statement of the proposition.

By Proposition 6.2, $\mathring{\alpha}_2^{(n)} 0_u = g_n(\mathring{\alpha}_1^{(n)} 0_u)$. Recall from (4) that

$$\mathring{\alpha}_1^{(n)} 0_u = P_s 0_o 1_o \dots (n+1)_o (n+1)_u n_o (n-1)_o \dots 2_o 1_o 0_o 0_u. \tag{7}$$

Defining $\sigma := \varnothing$ if $n = 1$ or $\sigma := n_o(n-1)_o \dots 2_o$ if $n > 1$, a computation shows that

$$\begin{aligned} h_1(\mathring{\alpha}_1^{(n)} 0_u) &= P_s 0_o 1_o 2_o \dots (n+2)_o (n+2)_u (n+1)_o n_o \dots 2_o 1_o 0_o 0_u \\ &= P_s 0_o 1_o 2_o \dots (n+2)_o (n+2)_u (n+1)_o \sigma 1_o 0_o 0_u. \end{aligned}$$

See also Figure 40, top pair. We will show that the image of the $1_o 0_o 0_u$ at the end of this last equation under $h_3^{(n)} \circ h_2^{(n)}$ produces the requisite segment in $\alpha_2^{(n)}$.

A direct computation shows that

$$h_2^{(n)}((n+2)_o (n+2)_u (n+1)_o \sigma 1_o 0_o 0_u) = (n+1)_o (n+1)_u \sigma h_2^{(n)}(1_o 0_o 0_u),$$

since all of σ is disjoint from D_2. In standard position, $1_o 0_o 0_u$ contains a full crossing, and we compute that

$$h_2^{(n)}(1_o 0_o 0_u) = 1_o \overline{\partial D_2|_{(-n-1,0]}} (-n-1)_o (-n-1)_u \partial D_2|_{(-n-1,0]} 0_o 0_u,$$

in an unreduced code. We thus have the decomposition

$$\begin{aligned} & h_2^{(n)}((n+2)_o (n+2)_u \sigma 1_o 0_o 0_u) \\ &\qquad = (n+1)_o (n+1)_u \sigma 1_o \overline{\partial D_2|_{(-n-1,0]}} (-n-1)_o (-n-1)_u \partial D_2|_{(-n-1,0]} 0_o 0_u \\ &\qquad = \tau_1 P_o (-1)_u \tau_2 (-n-1)_o (-n-1)_u \tau_3 (-1)_u P_o 0_u, \end{aligned} \tag{8}$$

where each τ_i is defined by the second equality. See Figure 40, middle pair. None of τ_1, τ_2, τ_3 fully cross D_3 in standard position. As $(-1)_u P_o 0_u$ fully crosses D_3 twice, we compute that

$$\begin{aligned} h_3^{(n)}((-1)_u P_o 0_u) &= (-1)_u P_u 0_o \partial D_3|_{(0,n]} (n+1)_u (n+1)_o \overline{\partial D_3|_{(0,n]}} 0_o P_u P_o 0_o \partial D_3|_{(0,n]} \\ &\qquad\qquad (n+1)_o (n+1)_u \overline{\partial D_3|_{(0,n]}} 0_o 0_u \\ &= (-1)_u P_u \overline{\chi}_1^{(n)} P_u P_o \chi_1^{(n)} 0_u. \end{aligned} \tag{9}$$

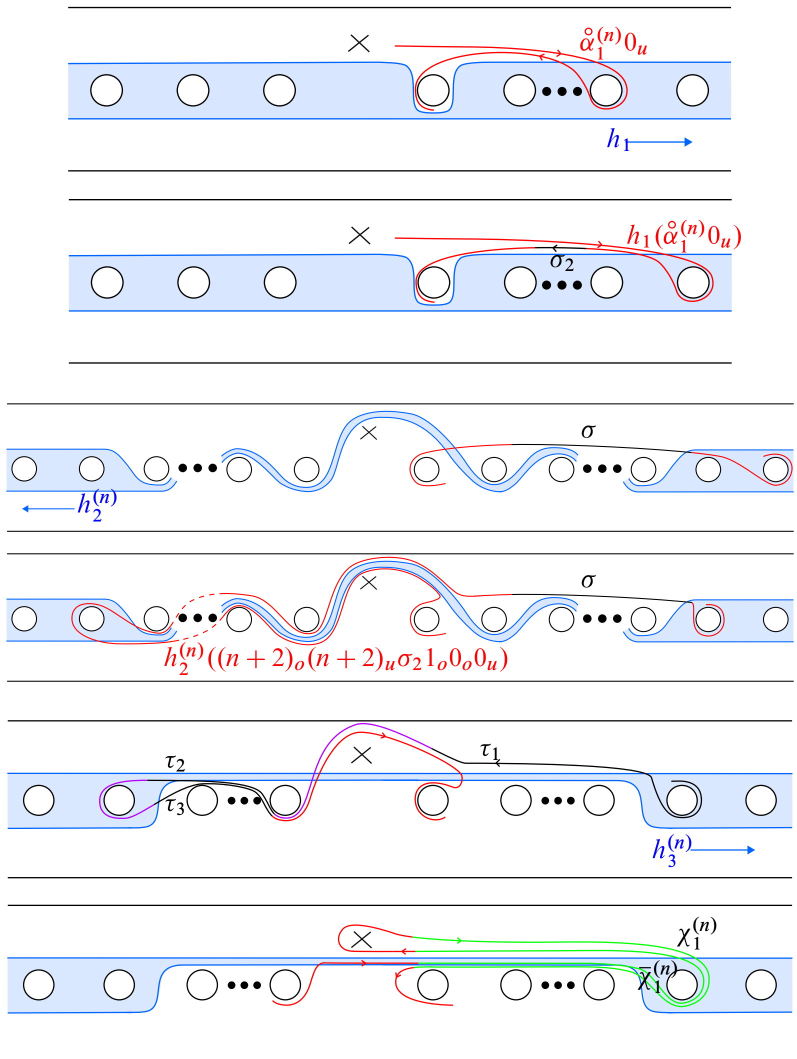


Figure 40: The direct computations done in the proof of Proposition 7.7. Top pair: the image of $\alpha_1^{(n)}0_u$ under h_1. Middle pair: the image under $h_2^{(n)}$ of the segment σ, shown in red and black. Bottom pair: the two full crossings in red give rise to $\chi_1^{(n)}$ and $\bar{\chi}_1^{(n)}$, which are shown in green, in the proof of Proposition 7.7 upon applying $h_3^{(n)}$.

Hence we see that $\bar{\chi}_1^{(n)} P_u P_o \chi_1^{(n)} 0_u$ is contained in an *unreduced* code for $g_n(\mathring{\alpha}_1^{(n)}0_u)$. It remains to show that this segment persists in a *reduced* code for $g_n(\mathring{\alpha}_1^{(n)}0_u) = \mathring{\alpha}_2^{(n)}0_u$.

Proposition 6.2 shows that if $\bar{\chi}_1^{(n)} P_u P_o \chi_1^{(n)} 0_u$ persists in a reduced code for $\mathring{\alpha}_2 0_u$ then it persists in a reduced code for α_2. To check if $\bar{\chi}_1^{(n)} P_u P_o \chi_1^{(n)} 0_u$ persists in a reduced code for $\mathring{\alpha}_2 0_u$, we need only consider a reduced code for the image under g_n of the characters following the $(n+1)_o(n+1)_u$ in (7), because Lemma 6.3 shows that the pair $(n+1)_o(n+1)_u$ blocks cascading cancellation between characters on either side. This is precisely the segment whose image under $h_2^{(n)}$ is given by (8).

We will use direct computation to show that the segment $\overline{\chi}_1^{(n)} P_u P_o \chi_1^{(n)} 0_u$ is the final segment of a reduced code for

$$h_3^{(n)}\big(\tau_1 P_o(-1)_u \tau_2(-n-1)_o(-n-1)_u \tau_3(-1)_u P_o 0_u\big). \tag{10}$$

Before making this computation, we point out the relevant pieces. Each $(-1)_u P_o$ (or its reverse) in (10) fully crosses D_3 and has image given by the first half of (9), up to reversing the orientation. The terminal $P_o 0_u$ also fully crosses D_3, and so we may similarly compute its image. The segments τ_2, τ_3 are invariant under $h_3^{(n)}$. The image of τ_1 is $h_3^{(n)}(\tau_1) = (n+2)_o(n+2)_u(n+1)_o \sigma 1_o 0_o$. Moreover, the four characters $\tau_2^t(-n-1)_o(-n-1)_u \tau_3^i$ form a loop which fully crosses D_3 exactly once. (Note that $\tau_2^t = \tau_3^i = (-n)_{o/u}$, where the choice of o/u depends on the parity of n.) Combining all of these remarks, we find the following *reduced* code (see Figure 40, bottom pair):

$$\begin{aligned}
&h_3^{(n)}\big(\tau_1 P_o(-1)_u \tau_2(-n-1)_o(-n-1)_u \tau_3(-1)_u P_o 0_u\big)\\
&= (n+2)_o(n+2)_u(n+1)_o \sigma 1_o 0_o P_o P_u 0_o \dots (n+1)_o(n+1)_u n_o \dots 0_o\\
&\qquad P_u \overline{\partial D_2|_{(-n,-1]}}(-n)_u(-n)_o \partial D_3|_{(-n,n+1)}(n+1)_o(n+1)_u \overline{\partial D_2|_{(-n,n]}}(-n)_o(-n)_u \partial D_2|_{(-n,-1]} P_u\\
&\qquad\qquad \partial D_3|_{[0,n+1)}(n+1)_u(n+1)_o \overline{\partial D_3|_{(P,n+1)}} P_u P_o \partial D_3|_{[0,n+1)}(n+1)_o(n+1)_u \overline{\partial D_3|_{[1,n+1)}} 0_o 0_u.
\end{aligned}$$

In particular, we see the requisite segment as the last line of this string. Precisely, we have

$$\chi_1^{(n)} = \partial D_3|_{[0,n+1)}(n+1)_o(n+1)_u \overline{\partial D_3|_{[1,n+1)}} 0_o.$$

This completes the proof of Proposition 7.7. □

The next corollary shows that $\alpha_i^{(n)}$ has highways whose lanes contain $\chi_{i-1}^{(n)}$. Notice that $g_n(\chi_{i-1}^{(n)}) = \chi_i^{(n)}$ by Proposition 6.2.

Corollary 7.8 *When $i \geq 2$, $\alpha_i^{(n)}$ contains a subsegment of the form $\overline{\chi}_{i-1}^{(n)} P_u P_o \chi_{i-1}^{(n)}$ that appears as the final subsegment of $\mathring{\alpha}_i^{(n)}$. In particular, $\alpha_i^{(n)}$ has highways for all $i \geq 2$ with lanes containing $\chi_{i-1}^{(n)}$.*

Proof As above, we will retain the 0_u which immediately follows $\mathring{\alpha}_i^{(n)}$ when we compute images. We will show by induction that $g_n(\mathring{\alpha}_{i-1}^{(n)} 0_u)$ ends with $\overline{\chi}_{i-1}^{(n)} P_u P_o \chi_{i-1}^{(n)} 0_u$ in a reduced code, which will show that $\overline{\chi}_{i-1}^{(n)} P_u P_o \chi_{i-1}^{(n)}$ appears as the final subsegment of $\mathring{\alpha}_i^{(n)}$. Since $g_n(\mathring{\alpha}_{i-1}^{(n)} 0_u) = \mathring{\alpha}_i^{(n)} 0_u$ by Proposition 6.2, this will imply that such a segment persists in $\alpha_i^{(n)}$.

The base case $i = 2$ was shown in Proposition 7.7, so we proceed to the induction step. Using the induction hypothesis, write

$$\mathring{\alpha}_{i-1}^{(n)} 0_u = P_s \eta_1 \overline{\chi}_{i-2}^{(n)} P_u P_o \chi_{i-2}^{(n)} 0_u,$$

for some subsegment η_1. A calculation shows that in an unreduced code

$$\begin{aligned}
g_n(\overline{\chi}_{i-2}^{(n)} P_u P_o \chi_{i-2}^{(n)} 0_u) &= g_n(\overline{\chi}_{i-2}^{(n)} + 0_o P_u P_o 0_o + \chi_{i-2}^{(n)} 0_u)\\
&= g_n(\overline{\chi}_{i-2}^{(n)}) + g_n(0_o P_u P_o 0_o) + g_n(\chi_{i-2}^{(n)} 0_u)\\
&= \overline{\chi}_{i-1}^{(n)} P_u P_o \chi_{i-1}^{(n)} 0_u,
\end{aligned}$$

where one verifies that the last line is in fact reduced. We therefore reduce to showing that $\bar{\chi}^{(n)}_{i-1} P_u P_o \chi^{(n)}_{i-1} 0_u$ persists as the terminal substring of $\mathring{\alpha}^{(n)}_i 0_u$. For this, we must prove that no characters in an unreduced code for $g_n(P_s \eta_1 0_o)$ cancel with this terminal 0_o, which is $(\bar{\chi}^{(n)}_{i-1})^i$.

Recall from (6) that $\chi^{(n)}_{i-1}$ is by definition all of $\mathring{\alpha}^{(n)}_{i-1}$ except the initial P_s. In particular, its terminal character $0_o = (\chi^{(n)}_{i-1})^t$ is oriented to the left. Thus the initial 0_o of the reverse segment $\bar{\chi}^{(n)}_{i-1}$ is oriented to the right. Since we are considering the segment $P_s \eta_1 0_o = P_s \eta_1 (\bar{\chi}^{(n)}_{i-1})^i$, this implies that $(\eta_1)^t$ has numerical value at most 0. Moreover, simplicity of $\mathring{\alpha}^{(n)}_{i-1}$ shows that if η_1 ends by looping around p, then it must have $1_o 0_o P_o P_u$ as its final segment. Theorem 6.4 combined with Remark 6.5 then shows that there is no cancellation with $(\bar{\chi}^{(n)}_{i-1})^i = 0_o$ in a reduced code for $g_n(P_s \eta_1 0_o)$. □

The following corollary follows immediately from Corollaries 6.9 and 7.8, our definition of $\chi^{(n)}_i$ in (6), and the definition of lane length.

Corollary 7.9 *Fix $i \geq 2$ and let $\mathcal{L}$ and $\mathcal{R}$ denote the collections of left and right lanes in $\alpha^{(n)}_i$, respectively. Then*

$$n_L = \max_{\lambda \in \mathcal{L}} L(\lambda) \geq \ell_c(\mathring{\alpha}^{(n)}_{i-1}) \quad \textit{and} \quad n_R = \max_{\rho \in \mathcal{R}} L(\rho) \geq \ell_c(\mathring{\alpha}^{(n)}_{i-1}).$$

We are now in a position to prove the main result of the section.

Theorem 7.10 *If δ is an arc which starts like $\alpha^{(n)}_i$ for some $i \geq 2$ and γ is any arc disjoint from δ, then γ starts like $\alpha^{(n)}_{i-1}$.*

Proof Proposition 7.7 and Corollary 7.8 show that $\mathring{\alpha}^{(n)}_i$ contains the segment $\bar{\chi}^{(n)}_{i-1} P_u P_o \chi^{(n)}_{i-1}$ and hence $\mathring{\alpha}^{(n)}_i$ has highways for all $i \geq 2$.

As δ starts like $\alpha^{(n)}_i$, the first part of its code is identical to that of $\mathring{\alpha}^{(n)}_i$ and therefore δ also has highways and also contains the segment $\bar{\chi}^{(n)}_{i-1} P_u P_o \chi^{(n)}_{i-1}$. By Lemma 7.6, the innermost left lane $P_u P_o \alpha_l$ (or its reverse) and the innermost right lane $P_o P_u \alpha_r$ (or its reverse) in δ each have lane length at least $\ell_c(\mathring{\alpha}^{(n)}_{i-1})$. Hence the first $\ell_c(\mathring{\alpha}^{(n)}_{i-1}) - 1$ characters of these lanes immediately following $P_{o/u} P_{u/o}$ must agree with $\chi^{(n)}_{i-1}$ or its reverse. As γ is disjoint from δ, the moreover statement of Lemma 7.6 gives that the code for γ must begin with $P_s \alpha_l$ or $P_s \alpha_r$. Consequently it must begin with $\mathring{\alpha}^{(n)}_{i-1} = P_s \chi^{(n)}_{i-1}$, as required. □

8 Loxodromic elements for $\mathcal{A}(\Sigma, p)$

In this section, we will first conclude the proof of Theorem 1.1 and then go on to explore some properties of our loxodromic elements g_n, including identifying an explicit geodesic axis for g_n, as well as describing the limit points of g_n on the boundary of the modified arc complex $\mathcal{A}(\Sigma, p)$ for an admissible surface Σ.

8.1 The proof of Theorem 1.1

Let S be the biinfinite flute surface we defined in Section 2 with isolated puncture p. Recall that $\mathcal{A}(S,p)$ denotes the modified arc graph of S, as in Definition 2.3. Let $\alpha_0 = P_s 0_o 0_u P_s \in \mathcal{A}(S,p)$ be the arc from Definition 5.2.

For each n, consider the map

$$\varphi_n \colon \mathcal{A}(S,p) \to \mathbb{Z}_{\geq 0}, \tag{11}$$

defined by setting $\varphi_n(\delta)$ equal to the largest $i \geq 0$ such that δ starts like $\alpha_i^{(n)}$. If there is no such i then set $\varphi_n(\delta) = 0$.

Lemma 8.1 *For any $\gamma, \delta \in \mathcal{A}(S,p)$, we have $d_{\mathcal{A}(S,p)}(\gamma,\delta) \geq |\varphi_n(\gamma) - \varphi_n(\delta)|$.*

Proof It follows from Theorem 7.10 that if $d_{\mathcal{A}(S,p)}(\gamma,\delta) = 1$, then $|\varphi_n(\gamma) - \varphi_n(\delta)| \leq 1$ for any n. In particular, if $\varphi_n(\delta) = \varphi_n(\gamma)$, then we are done. If not, then without loss of generality, assume that $\varphi_n(\delta) < \varphi_n(\gamma)$ and $\varphi_n(\gamma) = j$. Then by Theorem 7.10, $\varphi_n(\delta) = j-1$ so that $|\varphi_n(\gamma) - \varphi_n(\delta)| \leq 1$. The result then follows from the subadditivity of the absolute value. □

Let $\{g_n\}_{n\in\mathbb{N}}$ be as in Definition 5.1, and let $\{\alpha_i^{(n)}\}_{i\in\mathbb{Z}}$ be as in Definition 5.2. We first show that the elements $g_n \in \mathrm{Map}(S,p)$ are loxodromic with respect to the action on $\mathcal{A}(S,p)$.

Proposition 8.2 *For each $n \in \mathbb{N}$, the homeomorphism $g_n \in \mathrm{Map}(S,p)$ is a loxodromic isometry of $\mathcal{A}(S,p)$ with a $(2,0)$-quasigeodesic axis $\{\alpha_i^{(n)}\}_{i\in\mathbb{Z}}$.*

Proof We first show that the map $\mathbb{Z}_{\geq 0} \to \langle g_n\rangle\alpha_0 \subset \mathcal{A}(S,p)$ defined by $i \mapsto g_n^i(\alpha_0)$ is a $(2,0)$-quasi-isometry. In other words, we will show that $\{\alpha_i^{(n)}\}_{i\geq 0}$ is a quasigeodesic half-axis for g_n along which g_n acts as translation.

Let φ_n be the map defined in (11). By Lemma 8.1, we have that $d_{\mathcal{A}(S,p)}(\gamma,\delta) \geq |\varphi_n(\gamma) - \varphi_n(\delta)|$ for any $\gamma,\delta \in \mathcal{A}(S,p)$. Since $\varphi_n(\alpha_i^{(n)}) = i$ for all $i \geq 0$, this implies that

$$d_{\mathcal{A}(S,p)}(\alpha_0, \alpha_i^{(n)}) \geq i. \tag{12}$$

Consider the arc $\beta = P_s(-1)_o(-1)_u P_s$. Then β is disjoint from both $\alpha_0^{(n)}$ and $\alpha_1^{(n)}$. Since g_n is a homeomorphism, $g_n^i(\beta)$ is disjoint from both $\alpha_i^{(n)}$ and $\alpha_{i+1}^{(n)}$, and thus $d_{\mathcal{A}(S,p)}(\alpha_i^{(n)}, \alpha_{i+1}^{(n)}) \leq 2$ for all i. Therefore, for all i,

$$d_{\mathcal{A}(S,p)}(\alpha_0, \alpha_i^{(n)}) \leq 2i. \tag{13}$$

Together, (12) and (13) show that $\{\alpha_i^{(n)}\}_{i\geq 0}$ is a $(2,0)$-quasigeodesic half-axis for g_n.

Since $\{\alpha_i^{(n)}\}_{i\geq 0}$ is an unbounded orbit of g_n, we can see that g_n is not elliptic, and since g_n acts as translation along this quasigeodesic half-axis, g_n cannot be parabolic. Thus we may conclude that g_n is a loxodromic isometry of $\mathcal{A}(S,p)$. □

Recall from Definition 2.6 that a surface Σ with an isolated puncture p is admissible if there exists a proper embedding $S \hookrightarrow \Sigma$ where S contains p, the two nonisolated ends of S correspond to distinct ends of Σ, and with the property that either there are countably many connected components of $\Sigma \setminus S$ of the same (nontrivial) topological type or countably many isolated punctures of S remain isolated punctures when embedded in Σ. We denote this special class of connected components by $\mathcal{U}$, so that the elements of $\mathcal{U}$ are all homeomorphic to a fixed surface Σ_0. Recall that this definition ensures that the shift maps $h_1, h_2^{(n)}$, and $h_3^{(n)}$ we are interested in on S extend to shift maps on Σ. In particular, $g_n = h_3^{(n)} \circ h_2^{(n)} \circ h_1$ extends to a homeomorphism of Σ. When we reference g_n below, we will try to be specific about when we are considering g_n as an element of $\mathrm{Map}(S, p)$ versus an element of $\mathrm{Map}(\Sigma, p)$. Recall from Definition 2.2 that a homeomorphism $f \in \mathrm{Map}(\Sigma, p)$ is intrinsically infinite-type if $f \notin \overline{\mathrm{Map}_c(\Sigma)}$.

Theorem 1.1 is then a direct consequence of the following theorem.

Theorem 8.3 *Let Σ be an admissible surface. For each $n \in \mathbb{N}$, the homeomorphism $g_n \in \mathrm{Map}(\Sigma, p)$ is a loxodromic isometry of $\mathcal{A}(\Sigma, p)$ with a $(4,0)$-quasigeodesic axis $\{\alpha_i^{(n)}\}_{i \in \mathbb{Z}}$. Moreover, g_n is intrinsically infinite-type.*

Proof Fix $n \in \mathbb{N}$. By Proposition 8.2, g_n is loxodromic with respect to the action of $\mathrm{Map}(S, p)$ on $\mathcal{A}(S, p)$. Moreover, g_n extends to an element of $\mathrm{Map}(\Sigma, p)$ and so g_n acts on by isometries $\mathcal{A}(\Sigma, p)$ as well. By Lemma 2.10, there is a $(2,0)$-quasi-isometric embedding $\mathcal{A}(S, p) \hookrightarrow \mathcal{A}(\Sigma, p)$. Therefore the image of the $(2,0)$-quasigeodesic half-axis for g_n constructed in Proposition 8.2 is a $(4,0)$-quasigeodesic half-axis in $\mathcal{A}(\Sigma, p)$. It is clear that g_n stabilizes this $(4,0)$-quasigeodesic half-axis and so the arcs $\{\alpha_i^{(n)}\}_{i \in \mathbb{Z}_{\geq 0}}$ form a quasigeodesic half-axis for g_n in $\mathcal{A}(\Sigma, p)$. Therefore, g_n is loxodromic with respect to the action of $\mathrm{Map}(\Sigma, p)$ on $\mathcal{A}(\Sigma, p)$.

We now show that $g_n \notin \overline{\mathrm{Map}_c(\Sigma)}$. If Σ_0 has a nontrivial end space, then $g_n \notin \mathrm{PMap}(\Sigma)$ since g_n translates the elements of $\mathcal{U}$. Note that $\overline{\mathrm{Map}_c(\Sigma)} < \mathrm{PMap}(\Sigma)$ so that g_n is of intrinsically infinite-type in this case.

Now suppose that the end space of Σ_0 is trivial, so that Σ_0 is a finite-genus surface with one boundary component. Therefore, $S \cup \mathcal{U} = \Sigma'$ is homeomorphic to an infinite-genus surface with two nonplanar ends and a countable number of planar ends. Note that $\Sigma \setminus \Sigma'$ consists of all of the additional topology of Σ that is irrelevant to our construction of shift maps. In this way, the planar ends of Σ' cut away the irrelevant topology of Σ.

By [3, Corollary 6], $\mathrm{PMap}(\Sigma') = \overline{\mathrm{Map}_c(\Sigma')} \rtimes \mathbb{Z}(h')$, where h' is the standard handleshift on Σ', which also corresponds to a handleshift on Σ by extending h' via the identity on $\Sigma \setminus \Sigma'$. For simplicity of notation, we drop the subscript on g_n in what follows since the argument does not depend on n. Theorem 8.5 below tells us that when g is considered as an element of $\mathrm{Map}(S, p)$, there is a compactly supported mapping class φ such that $g = \varphi h$, where h is the right shift on S shifting the punctures corresponding to the elements of $\mathcal{U}$. As an element of $\mathrm{PMap}(\Sigma', p)$, and therefore of $\mathrm{Map}(\Sigma, p)$, g is thus equal to $\varphi \cdot (h')^m$, where $\varphi \in \mathrm{Map}_c(\Sigma') < \mathrm{Map}_c(\Sigma)$ and m is the genus of Σ_0. By the proof of [22, Proposition 6.3], $(h')^m \notin \overline{\mathrm{Map}_c(\Sigma)}$, so that $g \notin \overline{\mathrm{Map}_c(\Sigma)}$ as well. □

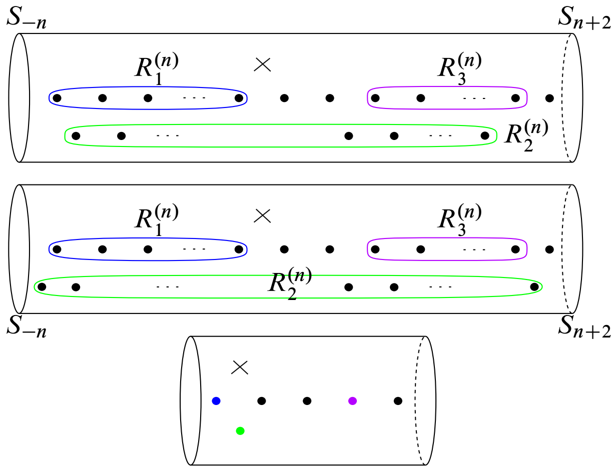


Figure 41: The curves $R_i^{(n)}$ when n is odd and even (top and middle, respectively), and the surface $\Pi = \Pi_n$ when $n \geq 3$ (bottom). In Π, the blue, green, and purple punctures correspond to where Π_n' was cut along the curves $R_1^{(n)}$, $R_2^{(n)}$, $R_3^{(n)}$, respectively.

8.2 Alternative description of g_n

In this section, we show that our mapping classes g_n can be written as the composition of a pseudo-Anosov on a finite-type subsurface and a standard shift. Moreover, we will show that g_n is the composition of *the same* pseudo-Anosov on a *fixed* finite-type subsurface and a standard shift whenever $n > 3$; however, this subsurface is embedded in Σ in different ways for different n, yielding distinct elements of $\mathrm{Map}(\Sigma, p)$. Though the g_n can be expressed as a pseudo-Anosov composed with a shift map, they are not end-periodic maps as defined in the work of Cantwell, Conlon, and Fenley [13]. In particular, there are isolated planar ends, or punctures, that are fixed by our homeomorphisms (for example, the puncture p). However, it is possible that some of the methods used to study end periodic maps could also apply to the homeomorphisms we have constructed.

Let h be the shift map on our subsurface S that translates the punctures that correspond to elements of $\mathcal{U}$, that is, the right shift whose domain contains the simple closed curves B_i for all $i \in \mathbb{Z}$ and maps B_i to B_{i+1}.

Without loss of generality, for each n we may modify our separating curves S_{-n} and S_{n+1} so that a connected component Π_n' of $S \setminus (S_{-n} \cup S_{n+2})$ is a sphere with 2 boundary components and $n+4$ punctures, one of which is p. This amounts to pushing any extra topology on the back of S outside of this subsurface. We now further modify Π_n' to form a subsurface Π_n for each n as follows. Let $R_1^{(n)}$, $R_2^{(n)}$, $R_3^{(n)}$ be the simple closed curves as shown in Figure 41. In particular,

- $R_1^{(n)}$ encloses all B_i with $-n \leq i \leq -1$ and i odd;
- $R_2^{(n)}$ encloses all B_i with $-n \leq i \leq n$ and i even; and
- $R_3^{(n)}$ encloses all B_i with $3 \leq i \leq n$ and i odd.

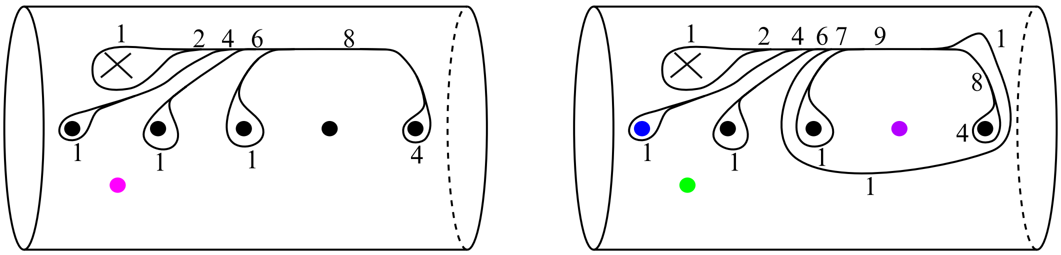


Figure 42: The train tracks τ_n on the finite-type subsurface Π_n. Left: $n = 1, 2$. The pink puncture only appears when $n = 2$. Right: $n \geq 3$.

Note that $R_i^{(1)}$ is empty for all $i = 1, 2, 3$, $R_i^{(2)}$ is empty when $i = 1, 3$, and $R_i^{(3)}$ is empty when $i = 3$. For all $n \geq 4$, $R_i^{(n)}$ is nonempty for each $i = 1, 2, 3$.

Definition 8.4 Let Π_n be the component of $\Pi'_n \setminus (R_1^{(n)} \cup R_2^{(n)} \cup R_3^{(n)})$ which contains the puncture p.

The surface Π_n is a sphere with 2 boundary components and some number of punctures: five punctures if $n = 1$, six punctures if $n = 2$, and seven punctures if $n \geq 3$. Notice that the Π_n are homeomorphic for all $n \geq 3$. However, the embedding $\iota_n \colon \Pi_n \to \Sigma$ are different for distinct n. For any $f \in \mathrm{PMap}(\Pi_n, p)$, let $\bar{f} := \iota_n \circ f$.

Theorem 8.5 *For each $n \geq 1$, there is a pseudo-Anosov $\varphi_n \in \mathrm{PMap}(\Pi_n)$ so that $g_n = \overline{\varphi}_n h$. Moreover, for all $n, n' \geq 3$, $\Pi_n = \Pi_{n'}$ and $\varphi_n = \varphi_{n'}$ as elements of $\mathrm{PMap}(\Pi_n)$. However, $\overline{\varphi}_n$ and $\overline{\varphi}_{n'}$ are distinct elements of $\mathrm{Map}(\Sigma, p)$ since the embeddings ι_n are distinct.*

Proof We define $\overline{\varphi}_n := g_n h^{-1}$ for all $n \geq 1$. It is clear that $\overline{\varphi}_n$ stabilizes the subsurface Π_n and is the identity on $\Sigma \setminus \Pi_n$. Let φ_n be the restriction of $\overline{\varphi}_n$ to $\mathrm{PMap}(\Pi_n)$, so that $\iota \circ \varphi_n = \overline{\varphi}_n$. We will show that φ_n is pseudo-Anosov. To do this, we will apply [28, Lemma 3.1], which states that a mapping class is pseudo-Anosov if it preserves a large, generic, birecurrent train track whose associated transition matrix is Perron–Frobenius. We will construct such a train track τ_n for each n.

The cases $n = 1, 2$ are slightly different and we will deal with them separately. For all $n \geq 3$, the surfaces Π_n are homeomorphic and we will build a single train track which will satisfy the above conditions. Notice that τ_n is a train track on the surface Π_n, so to show that τ_n is large, generic, and birecurrent it suffices to consider Π_n. However, since φ_n is defined as a restriction of $\overline{\varphi}_n$, which is a product of elements of $\mathrm{Map}(\Sigma, p)$ which are not supported on Π_n, we must consider the different embeddings of Π_n into Σ in order to show that τ_n is preserved by φ_n and to calculate the transition matrix of τ_n.

The train tracks τ_n for $n = 1, 2$ and for $n \geq 3$ are shown on Π_n in Figure 42. For all n, the following hold. All switches are trivalent and so τ_n is generic. Each complementary region is a once-punctured disk or a polygon and so τ_n is large. The weights on each branch of τ_n are positive and so τ_n is recurrent. Moreover,

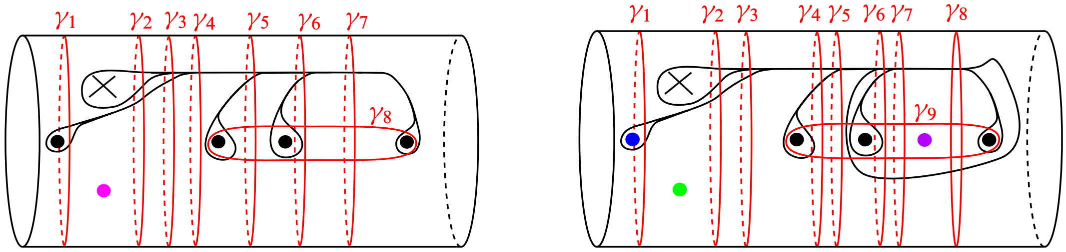


Figure 43: A collection of simple closed curves $\{\gamma_i^{(n)}\}$ such that each branch of τ_n is intersected transversely and efficiently by some $\gamma_i^{(n)}$. As in Figure 42, the pink puncture only appears when $n = 2$. Left: $n = 1, 2$. Right: $n \geq 3$.

the finite collections of simple closed curves $\{\gamma_i^{(n)}\}$ in Figure 43 is such that each branch of τ_n is intersected transversely and *efficiently* by some $\gamma_i^{(n)}$, ie $\gamma_i^{(n)} \cup \tau_n$ has no complementary bigon regions for any i. Therefore τ_n is transversely recurrent. Since τ_n is recurrent and transversely recurrent, it is birecurrent.

Figures 44 and 45 show that τ_n is preserved by φ_n for $n = 1, 2$ and $n \geq 3$, respectively. It is immediate from these figures that the matrix A_n associated to τ_n is

$$A_1 = A_2 = \begin{pmatrix} 5 & 6 & 0 & 2 \\ 6 & 9 & 0 & 2 \\ 10 & 10 & 2 & 3 \\ 6 & 6 & 1 & 2 \end{pmatrix} \quad \text{and} \quad A_n = \begin{pmatrix} 5 & 6 & 0 & 2 & 0 \\ 6 & 9 & 0 & 2 & 0 \\ 10 & 10 & 2 & 2 & 1 \\ 6 & 6 & 1 & 1 & 1 \\ 6 & 6 & 1 & 2 & 0 \end{pmatrix},$$

when $n \geq 3$.

For each n, $(A_n)^2$ has all positive entries, hence A_n is Perron–Frobenius. We conclude that φ_n is pseudo-Anosov for all n by [28, Lemma 3.1]. □

While it is not necessary for this paper, it is interesting to note that for all $n \in \mathbb{N}$, the top eigenvalue of A_n is $\frac{9}{2} + \frac{1}{2}\sqrt{41} + \sqrt{\frac{1}{2}(59 + 9\sqrt{41})}$, which is associated to a unique irrational lamination on Π_n that is carried by τ_n and fixed by φ_n.

We say that a homeomorphism of $\mathrm{Map}(\Sigma, p)$ is a *pseudo-Anosov shift* if it can be written as the composition of a pseudo-Anosov on a finite-type subsurface containing p and a standard shift. The results of this section inspire the following questions.

Question 8.6 *When is the composition of shift maps a pseudo-Anosov shift?*

Question 8.7 *Does every pseudo-Anosov shift act loxodromically on $\mathcal{A}(\Sigma, p)$?*

8.3 Geodesic axes

The proof of Theorem 1.1 shows that, for each n, the sequence $(\alpha_i^{(n)})_{i \in \mathbb{Z}}$ is a $(4, 0)$-quasigeodesic axis for g_n in $\mathcal{A}(\Sigma, p)$. In this section, we find a geodesic axis for g_n in $\mathcal{A}(\Sigma, p)$.

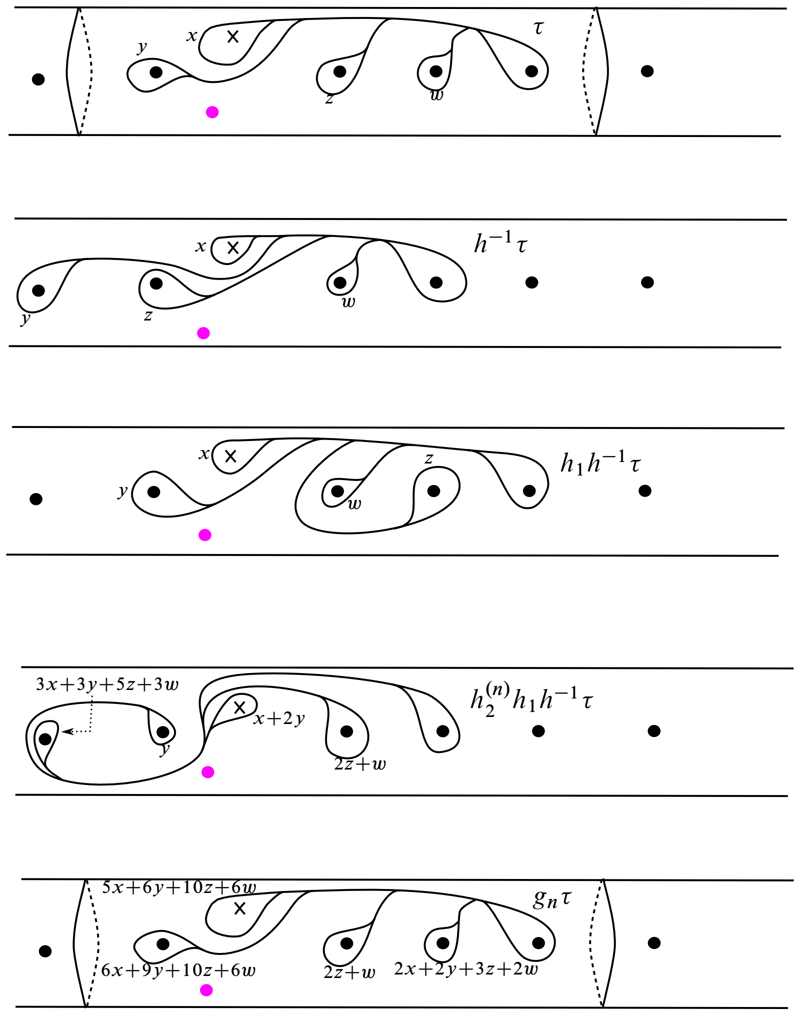


Figure 44: The train track τ_n is preserved by $\varphi_n = g_n h^{-1}$ for $n = 1, 2$; the pink puncture only appears when $n = 2$. The weights in the picture are used to calculate the matrix A_n associated to τ_n.

Theorem 8.8 *For each $n \in \mathbb{N}$, there is a geodesic axis for g_n in $\mathcal{A}(\Sigma, p)$. Furthermore, g_n has translation length* 1 *on this axis.*

Proof As $d_{\mathcal{A}(\Sigma,p)}(\gamma,\delta) \le d_{\mathcal{A}(S,p)}(\gamma,\delta)$ for any arcs $\gamma,\delta \in \mathcal{A}(S,p)$, the image of a geodesic under the inclusion $\mathcal{A}(S,p) \hookrightarrow \mathcal{A}(\Sigma,p)$ is still a geodesic. Thus it suffices to construct a geodesic axis for g_n in $\mathcal{A}(S,p)$. Toward this goal, define

$$\beta_0^{(n)} = P_s(-1)_o(-2)_o \dots (-n-1)_o(-n-1)_u(-n)_o \dots (-2)_o(-1)_o P_s \in \mathcal{A}(S,p),$$

and let

$$\beta_i^{(n)} = g_n^i(\beta_0^{(n)}).$$

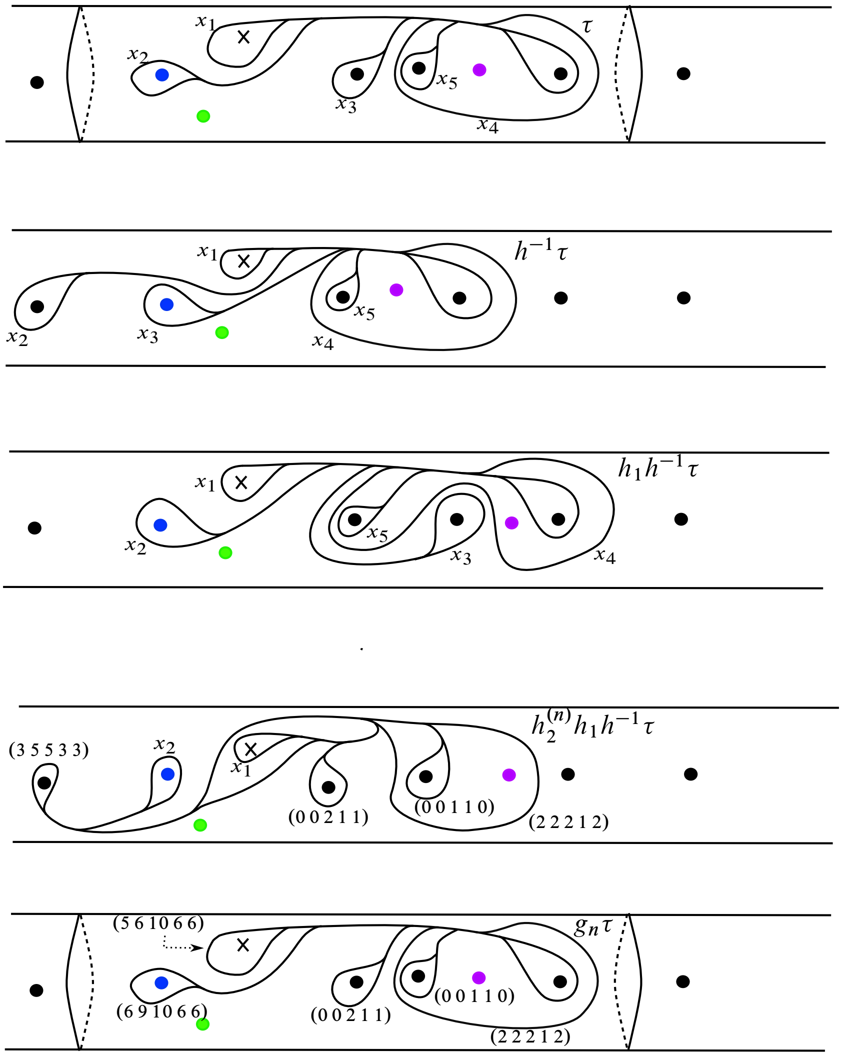


Figure 45: The train track τ_n is preserved by $\varphi_n = g_n h^{-1}$, when $n \geq 3$. The weights in the picture show how to calculate the matrix A_n associated to τ_n. For ease of notation, we often write the weights of each branch as a vector in the variables $x_1, \dots, x_5$. For example, the label (2 1 2 4 5) corresponds to the weight $2x_1 + x_2 + 2x_3 + 4x_4 + 5x_5$.

Since the arcs $\beta_i^{(n)}$ are the orbit of $\beta_0^{(n)}$ under $\langle g_n \rangle$ and g_n is a loxodromic isometry, it follows that they form a quasigeodesic axis in $\mathcal{A}(\Sigma, p)$ for g_n. We will show that $\beta_0^{(n)}$ is disjoint from $\beta_1^{(n)}$, which will prove that $(\beta_i^{(n)})_{i \in \mathbb{Z}}$ is a geodesic axis for g_n and that g_n has translation length one on this axis. In fact, it suffices to show that $\beta_1^{(n)}$ does not contain P_o or $k_{o/u}$ for any $k \leq -n-1$, as it then follows that $\beta_0^{(n)}$ is disjoint from $\beta_1^{(n)}$; see Figure 46.

Applying h_1 to $\beta_0^{(n)}$ yields

$$h_1(\beta_0^{(n)}) = P_s(-1)_o(-2)_o \dots (-n)_o(-n)_u(-n+1)_o \dots (-2)_o(-1)_o P_s. \tag{14}$$

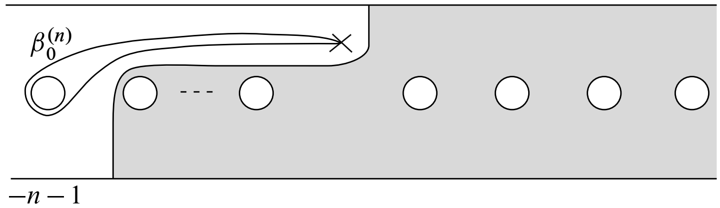


Figure 46: If an arc γ does not contain P_o or $k_{o/u}$ for any $k \leq -n-1$, then it must lie in the shaded region of S. In particular, γ must be disjoint from $\beta_0^{(n)}$.

Since all of $h_1(\beta_0^{(n)})$ is to the left of the puncture, the image under $h_2^{(n)}$ (which shifts to the left) will not contain P_o. Moreover, since P_o disagrees with the code for the domain of $h_3^{(n)}$, neither will $\beta_1^{(n)}$. Thus it remains to show that $\beta_1^{(n)}$ does not contain $k_{o/u}$ for any $k \leq -n-1$.

Recall that $h_3^{(n)}$ shifts to the right and has shift region $(-\infty, -n-1] \cup [n+1, \infty)$. Thus, any instance of $k_{o/u}$ with $k \leq -n-1$ in $\beta_1^{(n)}$ must be the image of $(k-1)_{o/u}$ in $h_2^{(n)}(h_1(\beta_0^{(n)}))$. Similarly, such a $(k-1)_{o/u}$ must be the image of $k_{o/u}$ in $h_1(\beta_0^{(n)})$, since $h_2^{(n)}$ shifts to the left. However, by (14), $h_1(\beta_0^{(n)})$ does not contain $j_{o/u}$ for $j \leq -n-1$, and we conclude that $\beta_1^{(n)}$ does not contain $k_{o/u}$ for any $k \leq -n-1$. □

8.4 Limit points of the g_n

Since the relative arc graph $\mathcal{A}(\Sigma, p)$ is a hyperbolic metric space, it has a well-defined Gromov boundary. This boundary was described by Bavard and Walker [7; 8]. In this section, we describe the limit set of g_n on $\partial\mathcal{A}(\Sigma, p)$ in terms of Bavard and Walker's characterization of the boundary.

8.4.1 The Gromov boundary of $\mathcal{A}(\Sigma, p)$ We begin by recalling some definitions from [7; 8]. It is important to mention that in [7; 8], the word *loop* is used for what we call an arc in this paper. Any time we mention a result from one of these two papers, we will convert it to our terminology. Let $E(\Sigma)$ denote the space of ends of Σ, which necessarily contains our preferred puncture p.

Fix any hyperbolic metric (of the first kind) on Σ, as in [8, Theorem 3.0.1]. For a fixed lift of p to the universal cover $\mathbb{H}^2$, which necessarily lies on $\partial\mathbb{H}^2$, there exists a parabolic subgroup $G < \pi_1(\Sigma)$ stabilizing this lift. Define $\widehat{\Sigma} = \mathbb{H}^2/G$ to be the intermediate cover associated to this parabolic subgroup. The space $\widehat{\Sigma}$ is called the *conical cover* of Σ. This cover has boundary $\mathbb{S}^1$ and contains a preferred lift $\widehat{p}$ of p which comes from our fixed choice in the universal cover. Let $\pi \colon \widehat{\Sigma} \to \Sigma$ be the natural quotient map, let $\widehat{\beta}$ be any geodesic ray from $\widehat{p}$ to $\partial\widehat{\Sigma}$, and let $\beta = \pi(\widehat{\beta})$. Thus $\widehat{\beta}$ has one endpoint on $\widehat{p}$ and the other endpoint somewhere in $\partial\widehat{\Sigma} \simeq \mathbb{S}^1$. The other endpoint may be a lift of p that is not our chosen $\widehat{p}$, (a lift of) a point in $E(\Sigma) \setminus \{p\}$, or a point which is neither. If β is simple, then in the first case β is an arc,[2] in the second case β is a *short ray*, and in the last case β is a *long ray*. Equivalently, an embedding

[2]As noted above, Bavard and Walker call this a *loop*.

$\beta\colon (0,1) \to \Sigma$ is a short ray if it can be continuously extended to a map $\widehat{\beta}\colon [0,1] \to \Sigma \cup E(\Sigma)$ such that $\widehat{\beta}|_{(0,1)} = \beta$, $\widehat{\beta}(0) = p$, and $\widehat{\beta}(1) \in E(\Sigma) \setminus \{p\}$ (see [8, Section 5.1.1]), and β is a long ray if it is neither an arc nor a short ray.

Bavard and Walker construct a graph involving all three kinds of rays, which they use to describe the Gromov boundary $\partial\mathcal{A}(\Sigma, p)$ of the relative arc graph.

Definition 8.9 The *completed ray graph* $\mathcal{R}(\Sigma, p)$ is the graph whose vertices are isotopy classes of simple arcs, short rays, and long rays and whose edges correspond to homotopically disjoint isotopy classes.

By definition, $\mathcal{A}(\Sigma, p)$ embeds in $\mathcal{R}(\Sigma, p)$, but the following theorem shows that something stronger is true. Recall that a clique is a complete graph.

Theorem 8.10 [8, Theorem 5.7.1] *The connected component of $\mathcal{R}(\Sigma, p)$ containing all arcs is quasi-isometric to $\mathcal{A}(\Sigma, p)$. All other connected components are cliques.*

A particular type of long ray will be important in the description of the Gromov boundary of $\mathcal{A}(\Sigma, p)$. A long ray β is *k-filling* if k is the minimum natural number such that there exists an arc β_0 and long rays $\beta_1, \dots, \beta_k = \beta$ such that $\beta_i \cap \beta_{i+1} = \varnothing$ for all $i \geq 0$. In other words, a long ray is k-filling if it is distance exactly k from an arc in $\mathcal{R}(\Sigma, p)$.

Definition 8.11 A long ray β is said to be *high-filling* if both of the following hold:

(1) β intersects every short ray; and

(2) β is not k-filling for any $k \in \mathbb{N}$.

All of the vertices of the connected components that form cliques are high-filling rays; accordingly, such cliques are called *high-filling cliques*.

Theorem 8.12 [8, Theorem 6.3.1] *The set of high-filling cliques is in bijection with the Gromov boundary $\partial\mathcal{A}(\Sigma, p)$ of the relative arc graph.*

8.4.2 The limit set of g_n In [8, Section 7.1], Bavard and Walker prove that to associated to any $f \in \mathrm{Map}(\Sigma, p)$ acting loxodromically on $\mathcal{A}(\Sigma, p)$, there exists an attracting clique of high-filling rays $C^+(f)$ and a repelling clique of high-filling rays $C^-(f)$ in $\mathcal{R}(\Sigma, p)$ that correspond to the attracting and repelling limit points of f in $\partial\mathcal{A}(\Sigma, p)$, respectively. The cliques $C^+(f)$ and $C^-(f)$ have the same (finite) number of vertices, called the *weight of f*, denoted by $w(f)$ (see [8, Definition 7.1.3]).

Following [7, Example 2.7.1], we have the following lemma.

Lemma 8.13 *For any $n \geq 1$, the homeomorphism g_n constructed in Theorem 1.1 satisfies $w(g_n) = 1$.*

Proof For notational simplicity, fix n and define $g = g_n$ and $\alpha_i = \alpha_i^{(n)}$. It suffices to prove that the attracting clique $C^+(g)$ consists of a single high-filling ray.

For each i, let $\hat{\alpha}_i$ be a lift of α_i to the conical cover $\hat{\Sigma}$. Then $\hat{\alpha}_i$ is a simple geodesic arc with one endpoint on $\hat{p}$ and the other on some $x_i \in \partial\hat{\Sigma}$. By [8, Lemma 5.2.2], the set of endpoints of arcs, short rays, and long rays is compact in $\partial\hat{\Sigma}$, and hence there exists a subsequence (x_{i_k}) of (x_i) which limits to a point $x_\infty \in \partial\hat{\Sigma}$. Let $\hat{\alpha}_\infty$ be the geodesic ray from $\hat{p}$ to x_∞ and $\alpha_\infty = \pi(\hat{\alpha}_\infty)$, where $\pi : \hat{\Sigma} \to \Sigma$ is the covering map. The construction of the conical cover $\hat{\Sigma}$ shows that α_∞ has an infinite-length code with initial segment $\mathring{\alpha}_{i_k}$ for any k. Since $\ell_c(\mathring{\alpha}_{i_k}) \to \infty$ as $k \to \infty$ and $\mathring{\alpha}_{i_k}$ has an initial segment equal to $\mathring{\alpha}_{i_{k-1}}$, this uniquely determines the entire infinite code. We claim that α_∞ is a high-filling ray and, moreover, that α_∞ is the sole vertex in $C^+(g)$.

That α_∞ is a ray follows from the fact that the set of endpoints of arcs in $\partial\hat{\Sigma}$ are isolated [8, Lemma 5.2.2]. To see that α_∞ is high-filling, it therefore suffices to show that it intersects every other ray (short or long). It follows from the same proof as Theorem 7.10 that any ray β which is disjoint from α_{i_k} must begin like α_{i_k-1}. In particular, we see that the first $\ell_c(\mathring{\alpha}_{i_{k-1}})$ characters in a code for β must agree with $\mathring{\alpha}_{i_{k-1}}$. Since $\ell_c(\mathring{\alpha}_{i_{k-1}}) < \ell_c(\mathring{\alpha}_{i_k})$ for all k, taking $k \to \infty$ shows that any ray β which is disjoint from α_∞ must have identical code and hence indeed must be exactly α_∞.

Finally, since α_∞ intersects every other ray, it must, in particular, intersect any other high-filling ray. By [7, Lemma 2.7.8], the connected component of any high-filling ray is a clique of high-filling rays, and thus α_∞ is its own connected component in $\mathcal{R}(\Sigma, p)$. Hence $w(g) = 1$, completing the proof. □

We close this section with the remark that if the weight of the limit points of g_n were not all the same, then Bavard and Walker give a method for constructing nontrivial quasimorphisms [8, Theorem 9.1.1]. However, since this is not the case, we must use a different method for showing that the space of quasimorphisms is infinite-dimensional, which is related to Bavard's original proof for the arc graph from [5]. We do so in Section 9.

9 Infinite-type quasimorphisms

A *quasimorphism* of a group G is a map $f : G \to \mathbb{R}$ such that there exists a real constant C for which $|f(ab) - f(a) - f(b)| \le C$ for all $a, b \in G$. The set of quasimorphisms forms a vector space $V(G)$ over $\mathbb{R}$, and, moreover, bounded functions and homomorphisms both form subspaces of $V(G)$. Let $\widetilde{\mathrm{QH}}(G)$ denote the quotient of $V(G)$ by the subspace spanned by bounded functions and homomorphisms. We call $\widetilde{\mathrm{QH}}(G)$ the *space of quasimorphisms* of G. The goal of this section is to use the elements constructed in Theorem 1.1 to prove Theorem 1.4, which we restate for the convenience of the reader.

Theorem 1.4 *Let Σ be an admissible surface. The space of nontrivial quasimorphisms on* $\mathrm{Map}(\Sigma, p)$ *is infinite-dimensional.*

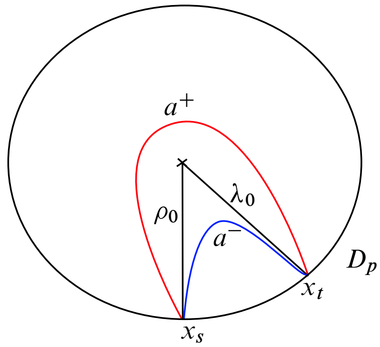


Figure 47: A schematic of D_p and the two arcs a^- and a^+.

Several of the ideas in this section closely follow the strategy and ideas of Bavard [5], though the production of the elements which give rise to our quasimorphisms differs. We begin by studying a specific subclass of arcs with the goal of describing a particular homotopy operation and intersection pairing on them. We then use this intersection pairing and a theorem of Bestvina and Fujiwara to prove the theorem.

9.1 An intersection pairing on symmetric arcs and first properties

Recall that our surface Σ has an embedding of the biinfinite flute surface S such that $\Sigma \setminus S$ has infinitely many connected components, a countable collection of which are homeomorphic to a fixed surface Σ_0. Moreover, the complement of each arc α_i separates Σ into two components, one of which is homeomorphic to $\operatorname{int}(\Sigma_0)$, the interior of the fixed surface Σ_0. Using Σ_0, we define $\mathcal{SA}$ to be the set of simple, symmetric arcs δ (see Definition 4.4) such that $\Sigma \setminus \delta$ has two connected components, one of which is homeomorphic to $\operatorname{int}(\Sigma_0)$. Notice that $\alpha_i^{(n)} \in \mathcal{SA}$ for all $i \in \mathbb{Z}$, $n \in \mathbb{N}$ and that $\mathcal{SA}$ is a $\operatorname{Map}(\Sigma, p)$-invariant subset of the set of all arcs.

Since p is isolated, we again fix the small once-punctured disk D_p containing p as in Section 7. This disk is homeomorphic to the closed unit disk minus an interior point. As in that section, given any element δ in $\mathcal{SA}$, we put δ in standard position so that its intersections with $\partial D_p \approx S^1$ are all transverse. Let x_s and x_t be the initial and terminal point where δ intersects ∂D_p. Let λ_0 and ρ_0 be the subsegments of δ which connect x_t and x_s to p, respectively.

We will modify δ to form a particular simple closed curve as follows. We can replace $\rho_0 \cup \lambda_0$ with either a^+ or a^-, as shown in Figure 47, forming two distinct simple closed curves, δ^+ and δ^-, respectively. One of these two curves bounds a surface homeomorphic to $\operatorname{int}(\Sigma_0)$; in Figure 47, this curve is δ^-. Fixing a hyperbolic metric on the surface, we let B_δ be the geodesic representative of this curve.

We now choose the homotopy representative of δ that will allow us to define an intersection pairing. As δ is symmetric, there exists an arc

$$\delta' = r_\delta B_\delta r_\delta^{-1}$$

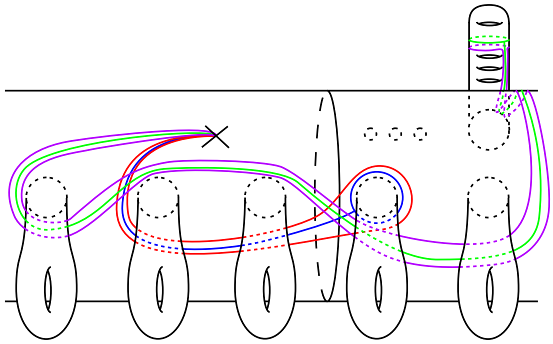


Figure 48: Two examples of arcs in $\mathcal{SA}$ (in the case that Σ_0 is a one-holed torus) and their corresponding zippings. The purple and red arcs are zipped to the green and blue arcs, respectively.

in the homotopy class of δ, where r_δ is a simple ray from the puncture to B_δ that intersects B_δ only at its endpoint. Intuitively, one can think of the arc δ' as being constructed from δ by "zipping" the initial and terminal portions of the arc together for as long as they fellow travel to form r_δ. Particular examples of r_δ and δ' are given in Figure 48.

We are now in a position to define the intersection pairings.

Definition 9.1 We define a map $I^{\pm}\colon \mathcal{SA} \times \mathcal{SA} \to \mathbb{Z}_{\geq 0}$ as follows. Let $I^+(\delta, \gamma)$ be the number of positively oriented intersections between minimal position representatives for the homotopy classes of r_δ and r_γ that do not occur on B_δ or B_γ. Here we require that the homotopy fixes the puncture and fixes each of B_δ and B_γ setwise (though not necessarily pointwise). We define $I^-(\delta, \gamma)$ similarly using negative intersections.

Notice that $I^+(-,-)$ and $I^-(-,-)$ are not necessarily symmetric in their arguments. However, it is straightforward to verify that the relationship

$$I^+(\delta, \gamma) = I^-(\gamma, \delta)$$

holds for any $\delta, \gamma \in \mathcal{SA}$.

For the remainder of the section we will fix an $n \in \mathbb{N}$ and use the notation that $g = g_n$, $\alpha_i = \alpha_i^{(n)}$, and $\varphi = \varphi_n$ is the "starts like" function from Section 8.

Example 9.2 One can readily compute from Figure 49 that we have the following relations:

$$5 = I^-(\alpha_0, \alpha_2) = I^+(\alpha_2, \alpha_0) \quad \text{and} \quad 6 = I^+(\alpha_0, \alpha_2) = I^-(\alpha_2, \alpha_0).$$

These calculations will be relevant later in the section.

We now collect some useful properties of the intersection pairing and its interaction with $\mathrm{Map}(\Sigma, p)$. These preliminary facts are inspired by Bavard [5], where similar statements are shown in Bavard's context.

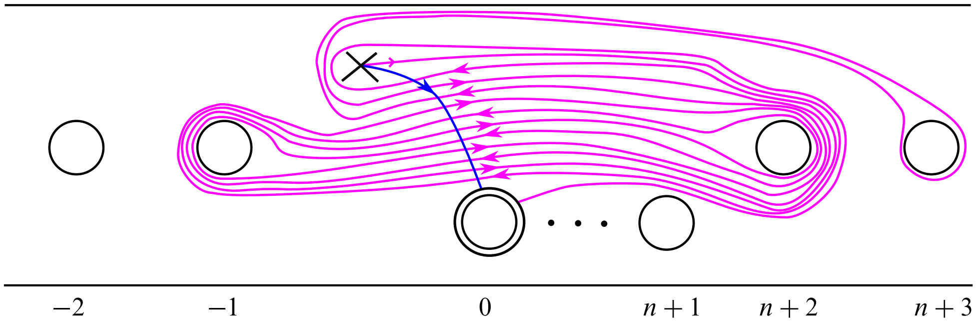


Figure 49: The oriented arc $r_{\alpha_0^{(n)}}$ is shown in blue and the oriented arc $r_{\alpha_2^{(n)}}$ is shown in pink. Each line of $r_{\alpha_0^{(n)}}$ and $r_{\alpha_2^{(n)}}$ represents two strands of $\alpha_0^{(n)}$ and $\alpha_2^{(n)}$, respectively.

Lemma 9.3 *The intersection pairing is mapping class group invariant. That is, for any $\delta, \gamma \in \mathcal{SA}$ and any $f \in \mathrm{Map}(\Sigma, p)$*

$$I^{\pm}(f(\delta), f(\gamma)) = I^{\pm}(\delta, \gamma).$$

Proof This is immediate from the fact that $\mathrm{Map}(\Sigma, p)$ is orientation preserving and preserves $\mathcal{SA}$. □

Recall that in Section 8, we defined the "starts like" function

$$\varphi\colon \mathcal{A}(S, p) \to \mathbb{Z}_{\geq 0},$$

by setting $\varphi(\delta)$ equal to the largest $i \geq 0$ such that δ starts like α_i. We now extend φ to all of $\mathcal{A}(\Sigma, p)$ by setting $\varphi(\delta) = 0$ if δ does not have a homotopy representative that is contained in S. We continue to call this extension φ.

Lemma 9.4 *If $\delta, \gamma \in \mathcal{SA}$ are arcs such that $2 + \varphi(\delta) \leq \varphi(\gamma)$, then*

$$6 \leq I^{-}(\gamma, \delta).$$

Proof By the mapping class group invariance of Lemma 9.3, the quantities $I^{\pm}(\alpha_i, \alpha_j)$ depend only on $j - i$. As α_{i+1} starts like α_i, the pairing $I^{-}(\alpha_i, \alpha_j)$ must be monotonically increasing in the difference $j - i$. In particular, if $2 + i \leq j$, then

$$6 = I^{-}(\alpha_2, \alpha_0) \leq I^{-}(\alpha_j, \alpha_i).$$

If $\varphi(\delta) = i$ and $\varphi(\gamma) = j$, then δ starts like α_i and γ starts like α_j, so the arcs γ and δ must have at least as many negatively oriented intersections as α_i and α_j. Thus we have

$$6 \leq I^{-}(\alpha_j, \alpha_i) \leq I^{-}(\gamma, \delta),$$

and the result follows. □

9.2 Production of "infinite-type" quasimorphisms

We now use the intersection pairing from the previous subsection to show that the elements g_n give rise to nontrivial quasimorphisms. For this, we need the following theorem of Bestvina and Fujiwara. We explain the two conditions on h_1, h_2 after the statement.

Theorem 9.5 [10, Theorem 1] *Suppose that a group G has a nonelementary action by isometries on a δ-hyperbolic graph X. If there exist independent loxodromic elements $h_1, h_2 \in G$ such that $h_1 \nsim h_2$, then the space of quasimorphisms is infinite-dimensional.*

Two loxodromic isometries h_1 and h_2 are *independent* if their limit sets in the boundary ∂X of X are disjoint. For the second condition, fix constants $K \geq 1$ and $K' \geq 0$ so that h_i has a (K, K')-quasigeodesic axis ℓ_i in X for $i \in \{1, 2\}$. A fundamental property of δ-hyperbolic spaces is that there exists $B = B(K, K', \delta)$ such that any two finite (K, K')-quasigeodesics with common endpoints are within distance B of each other. Define an equivalence relation on elements $h_1, h_2 \in G$ so that $h_1 \sim h_2$ if the following holds: for any arbitrarily long segment L of ℓ_1, there exists an $f \in G$ such that $f(L)$ is contained in the B-neighborhood of ℓ_2 and the map $f\colon L \to f(L)$ is orientation-preserving with respect to the h_i-orientation on ℓ_i for $i \in \{1, 2\}$. For the definition of the h_i-orientation on L and $f(L)$, see [10, page 72].

We now recall some arguments from [5, Section 4.3] which, when adapted into our language, show that $g \nsim g^{-1}$ for our loxodromic isometries $g = g_n$. Fix $B \geq 1$ to be the constant defined above for all $(4, 0)$-quasigeodesics in $\mathcal{A}(\Sigma, p)$. Let $\ell = \{g^i(\alpha_0)\}_{i \in \mathbb{Z}}$, so that ℓ is a $(4, 0)$-quasigeodesic axis of g by Theorem 8.3. We then have the following statements that are similar to [5, Lemmas 4.6, 4.7]. We supply the proofs for the reader's convenience.

Lemma 9.6 *Let L be a subpath of ℓ from α_i to α_j for $0 < i < j$. Let $f \in \mathrm{Map}(\Sigma, p)$ be such that $d_{\mathcal{A}(\Sigma,p)}(\alpha_i, f(\alpha_j)) \leq B$ and such that $f(L) \subset N_B(\ell)$ with the opposite orientation. If $j - i > 8B + 3$, then there exists some k such that $i \leq k < j$ and $\varphi(f(\alpha_{k+2})) \leq \varphi(f(\alpha_k)) - 2$.*

Proof Since $d_{\mathcal{A}(\Sigma,p)}(f(\alpha_j), \alpha_i) \leq B$, we conclude by Lemma 8.1 that $\varphi(f(\alpha_j)) \leq i + B$. Since $f(L) \subset N_B(\ell)$ with the opposite orientation, we may apply [5, Lemma 4.4] to L and the reverse of $f(L)$ to conclude that $d_{\mathcal{A}(\Sigma,p)}(\alpha_j, f(\alpha_i)) \leq 3B$. Note that [5, Lemma 4.4] is stated for geodesics, but the exact same proof goes through for quasigeodesics. Again applying Lemma 8.1, we see that $\varphi(f(\alpha_i)) \geq j - 3B$.

Suppose towards a contradiction that $\varphi(f(\alpha_{k+2})) > \varphi(f(\alpha_k)) - 2$ for all $1 \leq k < j$. Equivalently, $\varphi(f(\alpha_{k+2})) \geq \varphi(f(\alpha_k)) - 1$ for every k. Then, if $j - i$ is even,

$$i + B \geq \varphi(f(\alpha_j)) \geq \varphi(f(\alpha_i)) - \tfrac{1}{2}(j - i) \geq j - 3B - \tfrac{1}{2}(j - i),$$

where the second inequality is obtained by applying $\varphi(f(\alpha_{k+2})) \geq \varphi(f(\alpha_k)) - 1$ repeatedly starting with $j = k + 2$. If $j - i$ is odd, then by the same reasoning we have

$$i + B \geq \varphi(f(\alpha_j)) \geq \varphi(f(\alpha_{i+1})) - \tfrac{1}{2}(j - i - 1).$$

Since

$$|\varphi(f(\alpha_i)) - \varphi(f(\alpha_{i+1}))| \le d(f(\alpha_i), f(\alpha_{i+1})) = d(\alpha_i, \alpha_{i+1}) = 2,$$

it follows that

$$i + B \ge \varphi(f(\alpha_{i+1})) - \tfrac{1}{2}(j - i - 1) \ge \varphi(f(\alpha_i)) - \tfrac{3}{2} - \tfrac{1}{2}(j - i) \ge j - 3B - \tfrac{3}{2} - \tfrac{1}{2}(j - i).$$

Hence we conclude that, in either case,

$$4B + \tfrac{3}{2} \ge \tfrac{1}{2}(j - i),$$

which contradicts that $j - i > 8B + 3$. Thus there must be some k for which $\varphi(f(\alpha_{k+2})) \le \varphi(f(\alpha_k)) - 2$, as required. □

Proposition 9.7 *For any segment L of ℓ whose length is greater than $32B + 12$ and any $f \in \mathrm{Map}(\Sigma, p)$, if $f(L) \subset N_B(\ell)$ then $f(L)$ has the same orientation as ℓ.*

Proof After possibly increasing the length of L, we may assume that L is a subpath of ℓ from α_i to α_j for some $i < j$. Since the length of L is greater than $32B + 12$ and ℓ is a $(4, 0)$-quasigeodesic edge path by Theorem 8.3, we have that $j - i > 8B + 3$. By precomposing f with a suitable power of g, we can and do assume that $i, j > 0$.

Assume for contradiction that $f(L)$ has the opposite orientation as ℓ. By Example 9.2 and Lemma 9.3 we have that

$$I^-(f(\alpha_k), f(\alpha_{k+2})) = I^-(\alpha_k, \alpha_{k+2}) = I^-(\alpha_0, \alpha_2) = 5 \tag{15}$$

for all $k \in \mathbb{Z}$. On the other hand, by Lemma 9.6 there is some fixed index $i \le k < j$ for which

$$2 + \varphi(f(\alpha_{k+2})) \le \varphi(f(\alpha_k)). \tag{16}$$

Applying Lemma 9.4 to $f(\alpha_k)$ and $f(\alpha_{k+2})$ for this index k shows that

$$6 \le I^-(f(\alpha_k), f(\alpha_{k+2})) = I^-(\alpha_k, \alpha_{k+2}) = I^-(\alpha_0, \alpha_2).$$

However, this contradicts (15), and so we conclude that $f(L)$ has the same orientation as L. □

Proposition 9.7 implies that $g \not\sim g^{-1}$ since the axis ℓ has opposite orientations for g and g^{-1}. Additionally, since conjugate elements are equivalent under the relationship $\sim$, we have the following immediate corollary of Proposition 9.7.

Corollary 9.8 *For fixed $n \in \mathbb{N}$, the loxodromic elements $g = g_n$ have the property that $g \not\sim hg^{-1}h^{-1}$ for any $h \in \mathrm{Map}(\Sigma, p)$.*

With this in hand, we can prove the main result of this section.

Proof of Theorem 1.4 Fix $n \in \mathbb{N}$, and continue to use the notation that $g = g_n$. By Theorem 9.5 and Corollary 9.8 it suffices to show that there exists an $h \in \mathrm{Map}(\Sigma, p)$ such that g and $hg^{-1}h^{-1}$ are independent loxodromic elements. For this we can use any $h \in \mathrm{Map}(\Sigma, p)$ which does not fix the limit set of g (which is the same as the limit set of g^{-1}). For example, fix any finite-type subsurface Π_0 with boundary of sufficient complexity which contains the puncture p, and take any pseudo-Anosov $h \in \mathrm{Map}(\Pi_0, p)$. Extending h by the identity outside of Π_0, we may consider h as an element of $\mathrm{Map}(\Sigma, p)$. By Lemma 2.10, there is a $(2,0)$-quasi-isometric embedding $\iota\colon \mathcal{A}(\Pi_0, P) \hookrightarrow \mathcal{A}(\Sigma, p)$. Since h is loxodromic with respect to the action of $\mathrm{Map}(\Pi_0, p)$ on $\mathcal{A}(\Pi_0, p)$, it is therefore also loxodromic with respect to the action of $\mathrm{Map}(\Sigma, p)$ on $\mathcal{A}(\Sigma, p)$. In addition, out of all such pseudo-Anosovs, there is a choice of h whose limit points are different from the limit points $\xi_\pm$ of g. (Note that the quasi-isometric embedding ι ensures that two pseudo-Anosovs with distinct limit points in the boundary of $\mathcal{A}(\Pi_0, p)$ also have distinct limit points in the boundary of $\mathcal{A}(\Sigma, p)$.) Therefore, the limit points $h(\xi_\pm)$ of $hg^{-1}h^{-1}$ are distinct from those of g. In particular, g and $hg^{-1}h^{-1}$ are independent loxodromic elements, as required. □

10 Convergence to a geodesic lamination

The goal of this section is to prove Theorem 1.2, which we restate for the convenience of the reader. As in [26], we equip Σ with its unique conformal hyperbolic metric. Additionally, we require that with this metric, Σ is equal to its convex core, which is equivalent to eliminating hyperbolic funnels and half-planes in Σ. This is necessary in order to consider geodesic laminations on an infinite-type surface; see [26].

Theorem 1.2 *If Σ is an admissible surface equipped with its conformal hyperbolic metric that is equal to its convex core, then there exists a simple closed curve c_0 on Σ such that the sequence $(g_n^i(c_0))_{i \in \mathbb{N}}$ converges to a geodesic lamination on Σ.*

10.1 Geodesic laminations

We begin by reviewing some facts about geodesic laminations on infinite-type surfaces. For a complete treatment of the subject, we refer the reader to [26].

Definition 10.1 A *geodesic lamination* λ on Σ is a foliation of a closed subset of Σ by complete geodesics.

Fix a locally finite geodesic pants decomposition $\{P_n\}$ of Σ and a train track Θ on Σ constructed as in [26, Section 4]. Denote by $\widetilde{\Theta}$ the lift of Θ to $\widetilde{\Sigma}$, the universal cover of Σ. An *edge path* of $\widetilde{\Theta}$ is a finite, infinite, or biinfinite sequence of edges of $\widetilde{\Theta}$ such that consecutive edges meet smoothly at each vertex. Every biinfinite edge path has two distinct accumulation points on $\partial_\infty \widetilde{\Sigma}$ by [26, Proposition 4.5].

Given a biinfinite edge path $\widetilde{\gamma}$ of $\widetilde{\Theta}$, let $G(\widetilde{\gamma})$ be the geodesic of $\widetilde{\Sigma}$ whose endpoints on $\partial_\infty \widetilde{\Sigma}$ are the two distinct endpoints of $\widetilde{\gamma}$. A geodesic g of $\widetilde{\Sigma}$ is *weakly carried* by $\widetilde{\Theta}$ if there exists a biinfinite path $\widetilde{\gamma}$ in $\widetilde{\Theta}$ such that $G(\widetilde{\gamma}) = g$.

We now gather various results from [26] which will be useful in what follows. These are all analogous to the situation for finite-type surfaces (see, for example, [11; 23; 27]). The first gives a correspondence between geodesics weakly carried by $\widetilde{\Theta}$ and biinfinite edge paths.

Proposition 10.2 [26, Proposition 4.7] *There is a one-to-one correspondence between biinfinite edge paths of $\widetilde{\Theta}$ and geodesics weakly carried by $\widetilde{\Theta}$.*

Let γ be an edge path in Θ and $\tilde{\gamma}$ a single component of the lift of γ to $\widetilde{\Sigma}$. Then, as above, we construct a geodesic $G(\tilde{\gamma})$ with the same endpoints as $\tilde{\gamma}$ and denote by $G(\gamma)$ its projection to Σ. We then say that the geodesic $G(\gamma)$ is *weakly carried by* Θ if $G(\tilde{\gamma})$ is weakly carried by $\widetilde{\Theta}$. A geodesic lamination λ on Σ is *weakly carried by* Θ if every geodesic of λ is weakly carried by Θ. The next result gives a correspondence between geodesic laminations and certain families of biinfinite edge paths.

Proposition 10.3 [26, Proposition 4.11] *The set of geodesic laminations on Σ that are weakly carried by Θ is in one-to-one correspondence with the families Γ of biinfinite edge paths of $\widetilde{\Theta}$ that satisfy*

- *any two biinfinite edge paths γ and γ' in Γ do not cross*; *and*
- *if γ is a biinfinite edge path such that for any finite edge subpath there is a biinfinite edge path in Γ that contains it*, *then $\gamma \in \Gamma$.*

The following proposition describes when a sequence of geodesics carried by Θ converge. Let $G(\widetilde{\Sigma})$ be the set of unoriented geodesics on $\widetilde{\Sigma}$.

Proposition 10.4 [26, Proposition 4.9] *Let $f_n, f \in G(\widetilde{\Sigma})$ be weakly carried by a train track $\widetilde{\Theta}$, and denote by $\tilde{\gamma}_n, \tilde{\gamma}$ the corresponding biinfinite edge paths in $\widetilde{\Theta}$. Then f_n converges to f as $n \to \infty$ if and only if for each finite subpath $\tilde{\gamma}'$ of $\tilde{\gamma}$ there exists $n_0 \geq 0$ such that $\tilde{\gamma}'$ is contained in the path $\tilde{\gamma}_n$ for all $n \geq n_0$.*

The final proposition we will need shows that every geodesic lamination is weakly carried by a train track.

Proposition 10.5 [26, Proposition 4.12] *Every geodesic lamination λ on a hyperbolic surface X is weakly carried by a train track Θ that is constructed as above starting from a fixed locally finite geodesic pants decomposition.*

10.2 Construction of the train track Θ

We now construct a train track on our surface Σ. In the next section, we will define our simple closed curve c_0 and show that $g_n^i(c_0)$ converges as $i \to \infty$ to a geodesic lamination that is weakly carried by this train track.

First, fix a simple closed curve γ_p' on S so that the bounded component of $S \setminus \gamma_p$ is the pair of pants with cuffs γ_p', p_{-1}, and p (recall that punctures are allowed to be cuffs). Let γ_p be the image of γ_p' under the embedding $S \hookrightarrow \Sigma$. We construct our first pair of pants Q_p to have cuffs γ_p, B_{-1}, and p. Next

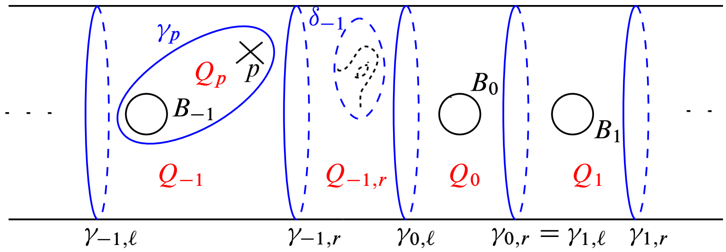


Figure 50: The cuffs of the pants decomposition are in blue and black, and the pairs of pants are labeled in red. Anything dotted occurs on the back of the embedded copy of S in Σ.

choose two simple closed separating curves $\gamma_{-1,l}$ and $\gamma_{-1,r}$ on the embedded copy of S in Σ so that one component of $\Sigma \setminus (\gamma_{-1,l} \cup \gamma_{-1,r})$ contains two pairs of pants: Q_p and a pair of pants Q_{-1} with cuffs $\gamma_{-1,l}, \gamma_{-1,r}$, and γ_p. Let $\gamma_{-1,l}$ be the curve which bounds the connected component of $\Sigma \setminus (\gamma_{-1,l} \cup \gamma_{-1,r})$ containing B_{-2}. For each $i \in \mathbb{Z} \setminus \{-1\}$, choose two simple closed separating curves $\gamma_{i,l}$ and $\gamma_{i,r}$ on the embedded copy of S in Σ so that one component of $\Sigma \setminus (\gamma_{i,l} \cup \gamma_{i,r})$ contains a pair of pants, Q_i, with cuffs $\gamma_{i,l}, \gamma_{i,r}$, and B_i. Similarly, let $\gamma_{i,l}$ be the curve which bounds the connected component of $\Sigma \setminus (\gamma_{-1,l} \cup \gamma_{-1,r})$ containing B_{i-1}.

Next, for all $i \in \mathbb{Z}$, consider the component C_i of $\Sigma \setminus (\gamma_{i,r} \cup \gamma_{i+1,l})$ which does not contain $\gamma_{i,l}$ for any i. If C_i is a cylinder, that is, if $\gamma_{i,r}$ is homotopic to $\gamma_{i+1,l}$, then modify $\gamma_{i,r}$ and Q_i so that $\gamma_{i,r} = \gamma_{i+1,l}$. If C_i is a pair of pants, let $Q_{i,r} = C_i$. If C_i is neither a pair of pants nor a cylinder, fix a simple closed curve δ_i so that one component of $C_i \setminus \delta_i$ is a pair of pants $Q_{i,r}$ with cuffs $\gamma_{i,r}, \gamma_{i+1,l}$, and δ_i. See Figure 50.

We replace each cuff with a geodesic representative of the same homotopy class; by a slight abuse of notation, we continue to call the resulting geodesic pairs of pants Q_i, Q_p, and $Q_{i,r}$. This is a locally finite geodesic pants decomposition of a subsurface of the embedded copy of S in Σ, and we extend it to a locally finite geodesic pants decomposition of Σ, which we denote $\mathcal{Q}$.

We next construct the train track Θ as follows. The specific connectors on the front of the surface for all Q_i, Q_p, and $Q_{i,r}$ are as in Figure 51. For all other pairs of pants, we choose any connectors that satisfy the conditions of [26].

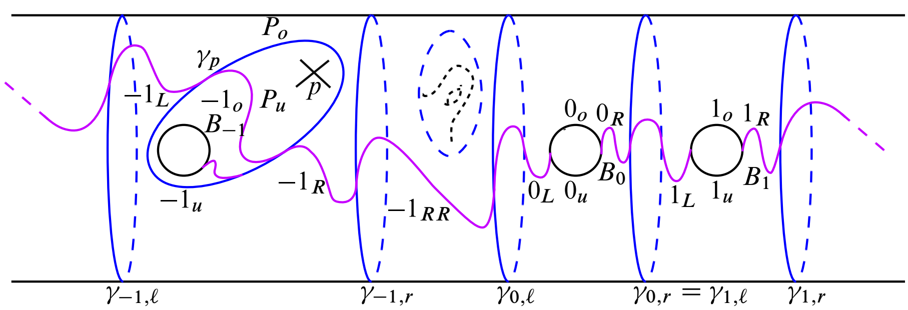


Figure 51: A portion of the train track Θ. The connectors are in purple, and the labels are the codes for certain connectors and cuffs.

We code a subset of the connectors and cuff segments of Θ which lie on the front of S in the following way (see Figure 51):

- In each pair of pants Q_i with $i \neq -1$, denote the connectors from $\gamma_{i,l}$ and $\gamma_{i,r}$ to B_i by i_L and i_R, respectively. The two segments of the cuff B_0 are denoted $0_o, 0_u$, as in our standard code.
- In each pair of pants $Q_{i,r}$, denote the single connector on the front of S by i_{RR}.
- In the pair of pants Q_p, denote the connector which has both endpoints on γ_p by P_u. This connector divides the cuff γ_p into two segments, denote these by P_o and $(-1)_u$, as in Figure 51. We will not code the second connector or the segments on the cuff B_{-1} in this pair of pants.
- In the pair of pants Q_{-1}, denote the connectors from $\gamma_{-1,l}$ and $\gamma_{-1,r}$ to γ_p by $(-1)_L$ and $(-1)_R$, respectively.

If a simple closed curve is carried by this train track and does not intersect any subpaths without a code, then we call this a Θ-*code* for the given simple closed curve. We say a Θ-code is *reduced* if no two adjacent characters are the same.

10.3 The simple closed curves c_i

For the remainder of this section, we fix $n \in \mathbb{N}$ and use the notation $g = g_n$ and $\alpha_i = \alpha_i^{(n)}$.

In [26], is it assumed that the relevant geodesic laminations do not contain geodesics that run out a cusp at one (or both) ends (see the discussion before [26, Proposition 4.12]). It is impossible for our sequence α_i to converge to a geodesic lamination of this type. In this section, we describe how to associate simple closed curves c_i to each α_i so that $g(c_{i-1}) = c_i$. In the following subsection, we prove that they converge to a geodesic lamination as in [26].

Let D_p be a small disk around the puncture p which is invariant under our homeomorphism g, as in Section 7.1. As before, we choose D_p small enough so that, for each i, $\alpha_i \cap D_p$ consists of two segments, one starting at p and one ending at p, with endpoints $z_{i,1}, z_{i,2}$ on ∂D_p. Up to homotopy, we may assume without loss of generality that $z_{i,1} = z_{j,1}$ and $z_{i,2} = z_{j,2}$ for all i, j. To form the simple closed curves c_i, we start with α_i and remove $\alpha_i \cap \operatorname{int}(D_p)$. We then add an arc of ∂D_p from $z_{i,1}$ to $z_{i,2}$; there two possible choices of arc, and we choose the one so that one connected component of $S \setminus c_0$ contains p and p_0. Note that this is the opposite choice than the one made in Section 9.1. See Figure 52.

For example, given the train track in Figure 51, a Θ-code of c_0 is

$$P_o(-1)_R(-1)_{RR}0_L0_o0_u0_L(-1)_{RR}(-1)_R P_u. \tag{17}$$

Note that since curves do not have a well-defined starting point, any cyclic permutation of this Θ-code for c_0 is also a Θ-code for c_0. For the rest of the section, we fix the starting point P_o for the Θ-code for c_0 as in (17).

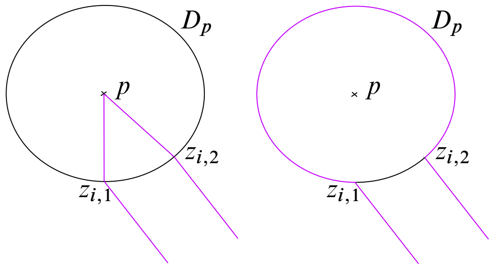


Figure 52: Forming c_i (right) from α_i (left). The initial and terminal segments of α_i are in purple (left), and the corresponding segment of c_i is in purple (right).

Our curves c_i and the arcs α_i agree outside of D_p, and since D_p is invariant under g, it follows that $g(c_i) = c_{i+1}$. For each i, we fix the starting point P_o for a Θ-code for c_i in such a way that applying g to the Θ-code for c_i yields the Θ-code for c_{i+1}.

10.4 Proof of Theorem 1.2

We will show that the simple closed curves c_i defined in the previous section converge to a geodesic lamination T on Σ. Our strategy is as follows. We will first fix a lift $\widetilde{\gamma}_i$ of each curve c_i in $\widetilde{\Sigma}$ and show that the sequence $\widetilde{\gamma}_i$ converges to some $\widetilde{\gamma}$. We then show that if $f_i = G(\widetilde{\gamma}_i)$ and $f = G(\widetilde{\gamma})$ are the corresponding geodesics that are weakly carried by $\widetilde{\Theta}$, then $\lim_{i\to\infty} f_i = f$. Finally, we take the geodesic lamination T to be the image of f in Σ.

For each i, let $\ell_\Theta(c_i)$ be the length of any reduced Θ-code for c_i. Recall that $\ell_c(\alpha_i)$ is the code length of α_i, as in Definition 3.6.

Lemma 10.6 *For each i, $\ell_c(\alpha_i) \le \ell_\Theta(c_i) \le 5\ell_c(\alpha_i)$.*

Proof It is clear that the Θ-code for c_i is at least as long as the code for α_i. For the second inequality, notice that each instance of $k_{o/u}$ with $k \ne -1, P$ in α_i is replaced with one of the following strings, depending on what precedes/follows the character $k_{o/u}$ and on the chosen pants decomposition,

$$k_L k_{o/u} k_R, \quad k_L k_{o/u}, \quad k_R k_{o/u}, \quad (k-1)_{RR} k_L k_{o/u} k_R, \quad (k-1)_{RR} k_L k_{o/u} k_R k_{RR}, \quad \text{or} \quad k_L k_{o/u} k_R k_{RR}.$$

If $k = -1$, then $(-1)_{o/u}$ is replaced with

$$(-1)_L(-1)_{o/u}, \quad (-1)_{o/u}(-1)_R, \quad (-2)_{RR}(-1)_L(-1)_{o/u},$$
$$(-1)_{o/u}(-1)_R(-1)_{RR}, \quad \text{or} \quad (-2)_{RR}(-1)_L(-1)_{o/u}(-1)_{RR}.$$

Finally, if $k = P$, then each $P_{o/u}$ remains the same and P_s is replaced with $P_{o/u}$. Therefore, each character in the code for α_i is replaced with at most 5 characters in the Θ-code for c_i, which gives the upper bound. $\square$

For each i, let $\ell_i = \ell_\Theta(c_i)$, and consider the Θ-code for c_i

$$c_i = c_1^i c_2^i \dots c_{\ell_i}^i. \tag{18}$$

For each i, fix the lift $\widetilde{\gamma}_i$ of c_i that is the periodic biinfinite edge path

$$\widetilde{\gamma}_i = (\dots, b^i_{-1}, b^i_0, b^i_1, b^i_2, \dots),$$

with period ℓ_i, where $b^i_j = c^i_j$ for each $1 \le j \le \ell_i$, and $b^i_j = b^i_{j-\ell_i}$ for all j. The codes for α_i and α_{i+1} agree on the first $\frac{1}{2}\ell_c(\alpha_i)$ characters. Thus it follows from Lemma 10.6 that the Θ-codes of c_i and c_{i+1} defined by (18) agree on at least the first $L_i = \lfloor \frac{1}{10}\ell_i \rfloor$ characters. Therefore, $b^i_j = b^{i+1}_j$ for all $1 \le j \le L_i$.

We define a biinfinite path

$$\widetilde{\gamma} = (\dots, d_{-1}, d_0, d_1, d_2, \dots),$$

as follows. Intuitively, our goal is to define $\widetilde{\gamma}$ so that it agrees with each $\widetilde{\gamma}_i$ from d_{-L_i} to d_{L_i}. In the first step, we define the characters d_{-L_0} to d_{L_0} of $\widetilde{\gamma}$ so that they agree with $\widetilde{\gamma}_0$. In the second step, we define the characters d_{-L_1} to d_{L_1} of $\widetilde{\gamma}$ so that they agree with $\widetilde{\gamma}_1$. The key point here is that $\widetilde{\gamma}_0$ and $\widetilde{\gamma}_1$ agree on the characters of $\widetilde{\gamma}$ that we have already defined in the first step. Thus we are not redefining d_i if $-L_0 \le i \le L_0$. Rather, these characters remain, and the additional information from the second step is the definition of d_i if $-L_1 \le i < -L_0$ or $L_0 < i \le L_1$. We then continue this process.

Formally, this is equivalent to the following definition. For each $i \ge 0$ and each $1 \le j \le L_i$, define

$$d_1 = b^i_1, \quad \dots, \quad d_j = b^i_j, \quad \dots, \quad d_{L_i} = b^i_{L_i},$$

and define

$$d_0 = b^i_1, \quad \dots, \quad d_{-j+1} = b^i_j, \quad \dots, \quad d_{-L_i+1} = b^i_{L_i}.$$

For each i and all $1 \le j \le L_i$, since $b^i_j = b^{i+1}_j$, there is no conflict with the previous defined edges as i increases.

By construction, $\widetilde{\gamma}$ is a biinfinite path in $\widetilde{\Theta}$. Let $f_i = G(\widetilde{\gamma}_i)$ and $f = G(\widetilde{\gamma})$ be the corresponding geodesics which are weakly carried by $\widetilde{\Theta}$, the existence of which is guaranteed by Proposition 10.2.

Lemma 10.7 $$\lim_{i\to\infty} f_i = f.$$

Proof This is almost immediate from the construction of $\widetilde{\gamma}$. Fix any finite subpath $T \subset \widetilde{\gamma}$. Then T is supported on $[-k, l]$ for some $k, l \ge 1$. Let $L = \max\{k, l\}$, and fix N such that $L_N \ge L$. Such an N exists since $\ell_c(\alpha_i) \to \infty$ implies that $\lim_{i\to\infty} L_i = \infty$. Then by construction T appears in all $\widetilde{\gamma}_i$ with $i \ge N$. Convergence follows by Proposition 10.4. □

Let T be the image of f in Σ. Figure 53 shows the train track which weakly carries the lamination, that is, the image of $\widetilde{\gamma}$ in Σ.

Lemma 10.8 *T is a geodesic lamination on Σ.*

Proof This follows immediately from Proposition 10.3 applied to $\Gamma = \{\widetilde{\gamma}\}$. □

Finally, Theorem 1.2 follows from Lemma 10.8.

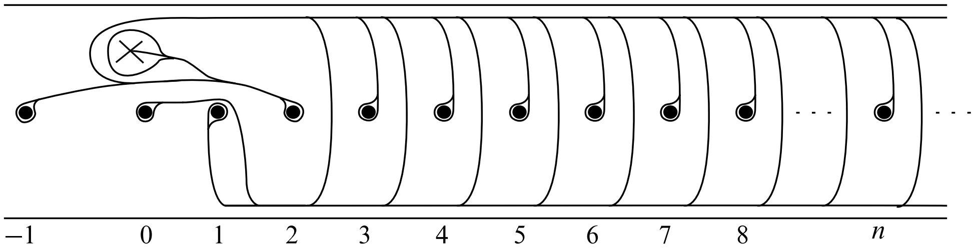


Figure 53: The train track on Σ which weakly carries the lamination T when $n = 1$. Note that the train track is, in fact, contained on the front of the embedded copy of S in Σ.

References

[1] **American Institute of Mathematics**, *Surfaces of infinite type*, problem lists (2019) Available at `http://aimpl.org/genusinfinity/`

[2] **J Aramayona**, **A Fossas**, **H Parlier**, *Arc and curve graphs for infinite-type surfaces*, Proc. Amer. Math. Soc. 145 (2017) 4995–5006 MR Zbl

[3] **J Aramayona**, **P Patel**, **N G Vlamis**, *The first integral cohomology of pure mapping class groups*, Int. Math. Res. Not. 2020 (2020) 8973–8996 MR Zbl

[4] **J Aramayona**, **N G Vlamis**, *Big mapping class groups: an overview*, from "In the tradition of Thurston: geometry and topology" (K Ohshika, A Papadopoulos, editors), Springer (2020) 459–496 MR Zbl

[5] **J Bavard**, *Hyperbolicité du graphe des rayons et quasi-morphismes sur un gros groupe modulaire*, Geom. Topol. 20 (2016) 491–535 MR Zbl

[6] **J Bavard**, **A Genevois**, *Big mapping class groups are not acylindrically hyperbolic*, Math. Slovaca 68 (2018) 71–76 MR Zbl

[7] **J Bavard**, **A Walker**, *The Gromov boundary of the ray graph*, Trans. Amer. Math. Soc. 370 (2018) 7647–7678 MR Zbl

[8] **J Bavard**, **A Walker**, *Two simultaneous actions of big mapping class groups*, Trans. Amer. Math. Soc. 376 (2023) 7603–7650 MR Zbl

[9] **M Bestvina**, **K Bromberg**, **K Fujiwara**, *Constructing group actions on quasi-trees and applications to mapping class groups*, Publ. Math. Inst. Hautes Études Sci. 122 (2015) 1–64 MR Zbl

[10] **M Bestvina**, **K Fujiwara**, *Bounded cohomology of subgroups of mapping class groups*, Geom. Topol. 6 (2002) 69–89 MR Zbl

[11] **F Bonahon**, *Closed curves on surfaces*, preprint (2000) Available at `https://dornsife.usc.edu/francis-bonahon/wp-content/uploads/sites/205/2023/06/Bouquin.pdf`

[12] **D Calegari**, *Big mapping class groups and dynamics*, blog post (2009) Zbl Available at `https://lamington.wordpress.com/2009/06/22/big-mapping-class-groups-and-dynamics`

[13] **J Cantwell**, **L Conlon**, **S R Fenley**, *Endperiodic automorphisms of surfaces and foliations*, Ergodic Theory Dynam. Systems 41 (2021) 66–212 MR Zbl

[14] **F Fournier-Facio**, **Y Lodha**, **M C B Zaremsky**, *Braided Thompson groups with and without quasimorphisms*, Algebr. Geom. Topol. 24 (2024) 1601–1622 MR Zbl

[15] **S Hensel**, **P Przytycki**, **R C H Webb**, 1-*slim triangles and uniform hyperbolicity for arc graphs and curve graphs*, J. Eur. Math. Soc. 17 (2015) 755–762 MR Zbl

[16] **W P Hooper**, *The invariant measures of some infinite interval exchange maps*, Geom. Topol. 19 (2015) 1895–2038 MR Zbl

[17] **B von Kerékjártó**, *Vorlesungen über Topologie, I: Flächentoplogie*, Springer (1923) Zbl

[18] **J Lanier**, **M Loving**, *Centers of subgroups of big mapping class groups and the Tits alternative*, Glas. Mat. Ser. III 55(75) (2020) 85–91 MR Zbl

[19] **W B R Lickorish**, *A representation of orientable combinatorial* 3*-manifolds*, Ann. of Math. 76 (1962) 531–540 MR Zbl

[20] **I Morales**, **F Valdez**, *Loxodromic elements in big mapping class groups via the Hooper–Thurston–Veech construction*, Algebr. Geom. Topol. 22 (2022) 3809–3854 MR Zbl

[21] **J Nielsen**, *Surface transformation classes of algebraically finite type*, Danske Vid. Selsk. Mat.-Fys. Medd. 2, Munksgaard, Copenhagen (1944) MR Zbl

[22] **P Patel**, **N G Vlamis**, *Algebraic and topological properties of big mapping class groups*, Algebr. Geom. Topol. 18 (2018) 4109–4142 MR Zbl

[23] **R C Penner**, **J L Harer**, *Combinatorics of train tracks*, Ann. of Math. Stud. 125, Princeton Univ. Press (1992) MR Zbl

[24] **A J Rasmussen**, *WWPD elements of big mapping class groups*, Groups Geom. Dyn. 15 (2021) 825–848 MR Zbl

[25] **I Richards**, *On the classification of noncompact surfaces*, Trans. Amer. Math. Soc. 106 (1963) 259–269 MR Zbl

[26] **D Šarić**, *Train tracks and measured laminations on infinite surfaces*, Trans. Amer. Math. Soc. 374 (2021) 8903–8947 MR Zbl

[27] **W P Thurston**, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Amer. Math. Soc. 19 (1988) 417–431 MR Zbl

[28] **Y Verberne**, *A construction of pseudo-Anosov homeomorphisms using positive twists*, Algebr. Geom. Topol. 23 (2023) 1601–1639 MR Zbl

Department of Mathematics, Brandeis University
Waltham, MA, United States

Department of Mathematics, University of Oklahoma
Norman, OK, United States

Department of Mathematics, University of Utah
Salt Lake City, UT, United States

carolynabbott@brandeis.edu, nickmbmiller@ou.edu, patelp@math.utah.edu

Received: 9 August 2023

Guidelines for Authors

Submitting a paper to Algebraic & Geometric Topology

Papers must be submitted using the upload page at the AGT website. You will need to choose a suitable editor from the list of editors' interests and to supply MSC codes.

The normal language used by the journal is English. Articles written in other languages are acceptable, provided your chosen editor is comfortable with the language and you supply an additional English version of the abstract.

Preparing your article for Algebraic & Geometric Topology

At the time of submission you need only supply a PDF file. Once accepted for publication, the paper must be supplied in LaTeX, preferably using the journal's class file. More information on preparing articles in LaTeX for publication in AGT is available on the AGT website.

`arXiv` papers

If your paper has previously been deposited on the `arXiv`, we will need its `arXiv` number at acceptance time. This allows us to deposit the DOI of the published version on the paper's `arXiv` page.

References

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited at least once in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used, but will be converted to the house style (see a current issue for examples).

Figures

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Fuzzy or sloppily drawn figures will not be accepted. For labeling figure elements consider the pinlabel LaTeX package, but other methods are fine if the result is editable. If you're not sure whether your figures are acceptable, check with production by sending an email to graphics@msp.org.

Proofs

Page proofs will be made available to authors (or to the designated corresponding author) in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

ALGEBRAIC & GEOMETRIC TOPOLOGY

Volume 25 Issue 1 (pages 1–644) 2025