AGT

Algebraic & Geometric Topology

Volume 26 (2026)

Issue 3 (pages 825–1227)

msp

ALGEBRAIC & GEOMETRIC TOPOLOGY

msp.org/agt

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

The subscription price for 2026 is US $795/year for the electronic version, and $1170/year (+$80, 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, 2000 Allston Way # 59, Berkeley, CA 94701-4004. Periodical rate postage paid at Oakland, CA 94615-9651, and additional mailing offices. POSTMASTER: send address changes to Mathematical Sciences Publishers, 2000 Allston Way # 59, Berkeley, CA 94701-4004.

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

https://msp.org/

msp

Algebraic & Geometric Topology **26**:3 (2026) 825–862
DOI: 10.2140/agt.2026.26.825
Published: April 1, 2026

Standard position for surfaces in link complements in arbitrary 3-manifolds

JESSICA S. PURCELL AND ANASTASIIA TSVIETKOVA

Since the 1980s, it has been known that essential surfaces in alternating link complements can be isotoped to be transverse to the link diagram almost everywhere, with the exception of some well-understood intersections, and described combinatorially as a result. This was called standard position for surfaces and has had numerous applications. However, the original techniques only apply to classical alternating links projected onto the 2-sphere inside the 3-sphere. In this paper, we prove that standard position for surfaces can be extended to a broader class, namely weakly generalized alternating links. Such links include all classical prime nonsplit alternating links in the 3-sphere, and also many links that are alternating on higher-genus surfaces, or lie in manifolds besides the 3-sphere. As an application, we show that all such links are prime, and that under mild restrictions, essential Conway spheres for such links interact with the diagram exactly as in the classical alternating setting.

1 Introduction

Essential embedded surfaces in a 3-manifold can provide useful information about the topology of that manifold. For example, Haken manifolds, which contain an orientable closed essential surface, have been an object of significant research since they were introduced in the 1960s and have many important properties [6; 29]. To work with essential embedded surfaces, it is often useful to put them in a certain topological form that relates to properties of the ambient 3-manifold. When the ambient 3-manifold is an alternating link complement, Menasco [20] and Menasco and Thistlethwaite [22] showed that an essential surface could be isotoped into *standard position* with respect to the link projection surface.

For more general 3-manifolds, an essential surface may be isotoped into *normal form* with respect to a triangulation of a 3-manifold [6], or similarly normal form with respect to polyhedral or other decompositions [5; 16]. In [14], Howie and Purcell extended normal form to broad generalizations of alternating links, called weakly generalized alternating links, and used this to determine geometric information, generalizing several of the above results.

In this paper, we take this further, generalizing various results on standard position in [20; 22] to weakly generalized alternating links.

As a corollary, we show that all weakly generalized alternating links are prime, extending a result of Howie and Purcell [14, Corollary 4.7], who showed primeness of such links with some restrictions.

We also classify essential Conway spheres for weakly generalized alternating links satisfying mild restrictions, showing that they either intersect the diagram in a single curve meeting four punctures (a "visible" sphere), or intersect it in two well-behaved curves (a "hidden" sphere). The visible-hidden dichotomy is analogous to a similar result for classical alternating links, discovered by Menasco [20],

MSC2020: 57K10, 57K12, 57K30.

named and refined by Thistlethwaite [27], and used to study alternating tangles by, for example, Hass, Thompson, and Tsvietkova [9], and Champanerkar, Kofman, and Purcell [4].

1.1 Alternating links in the 3-sphere and beyond

Normal and standard form for surfaces in the complements of alternating links in the 3-sphere have had many applications. Historically, Menasco used it to prove a number of results for alternating link complements in the 3-sphere, such as classifying when the link is prime, split and hyperbolic [20], and determining when a surface is incompressible [21]. Menasco and Thistlethwaite used it to show that the only reducible Dehn fillings are the expected ones on $(2, q)$-torus knots [22], and thus to prove the cabling conjecture for all alternating knots. Lackenby used normal form to describe exceptional Dehn fillings [16]. Howie used it to show that essential embedded surfaces with boundary can be used to detect classical alternating links [13]. Hass, Thompson and Tsvietkova used it to give universal polynomial bounds for the number of embedded surfaces in alternating link complements [7; 8]. These are just some examples of numerous applications to alternating links in the 3-sphere.

In recent years, there has been significant interest in the study of alternating projections onto surfaces besides the 2-sphere, for example due to Adams [1], Hayashi [11], Ozawa [23], Howie [12], and others. There has also been interest in alternating knots within manifolds besides the 3-sphere, such as virtual knots, for example by Adams [2], Champanerkar, Kofman, and Purcell [3], and Howie and Purcell [14]. The weakly generalized alternating links of this paper match those of Howie and Purcell. They include classical alternating knots on a 2-sphere in the 3-sphere, and also include many of the other generalizations above, including families of toroidally alternating knots and alternating knots on other Heegaard surfaces, alternating knots in thickened surfaces, i.e., virtual alternating knots, and further examples still. The full definition of such links is given in Section 2.

1.2 Surfaces containing meridians

A key result in this paper is to show that an essential surface meeting a knot in meridians can be isotoped into normal form without distorting the boundary curves. We call this *meridianal form*, defined in Section 5.

Menasco showed that any closed essential surface embedded in a nonsplit prime alternating link complement in S^3 contains a closed curve isotopic to the link meridian [20]. This sometimes is referred to as the meridian lemma. A meridian lemma holds for other knots, such as algebraic knots [24]. A meridian lemma also holds for certain weakly generalized alternating links [14, Lemma 4.9], and has been announced for weakly generalized alternating links that are also virtual alternating links by Wei Lin [17].

By compressing along a meridianal annulus, one can reduce the study of closed essential surfaces to the study of essential surfaces with meridianal boundary within the complement of a regular neighborhood of the link, i.e., the link exterior. For such surfaces, it is important that isotopies retain the meridianal form of the boundary. This has been used, for example, in the proof of a polynomial upper bound on the number of closed surfaces in alternating links by Hass, Thompson and Tsvietkova [7], and in Lozano and Przytycki's work on 3-braid links [18].

Studying surfaces that contain meridians has also led to results on prime diagrams. For example, a classical alternating link is prime if and only if its diagram is prime, due to Menasco [20].

In this paper, using normal and meridianal form of surfaces containing meridians, we show that all weakly generalized alternating links must be prime.

Theorem 10.3 *A weakly generalized alternating link is prime.*

Theorem 10.3 is analogous to Menasco's result, since weakly generalized alternating links have a diagrammatic primeness condition built into the definition. Howie and Purcell proved a similar result for weakly generalized alternating links with an additional condition (namely $\hat{r} > 4$; see Section 2) [14, Corollary 4.7]. Howie and Purcell's result already implies that all virtual alternating links that are checkerboard colorable and have a weakly prime diagram must also be prime. For example this is true for virtual alternating links whose diagrams have all complementary regions disks, as studied by Adams et al. [2]. However, the results in this paper extend primeness to broader families of weakly generalized alternating links, for example on projection surfaces that are compressible, which do not satisfy Howie and Purcell's stronger requirement.

Note that in the virtual setting, other notions of primeness also appear in the literature, for example recent work of Kindred [15].

Finally, surfaces in a knot exterior meeting the boundary in meridians include 4-punctured spheres, which separate a knot into tangles. Examining such surfaces has allowed the detection and study of prime tangles, for example by Menasco and Thistlethwaite [20; 27], and more recently by Hass, Thompson and Tsvietkova [9], and Champanerkar, Kofman, and Purcell [4]. The tools of this paper also allow us to extend such results to larger families of links. We consider essential 4-punctured spheres in the complements of weakly generalized alternating links satisfying an extra condition (namely $\hat{r}(\pi(L), \Pi) > 4$), and we show that they satisfy the same constraints as in the classical alternating case.

Theorem 12.1 *Let L be a weakly generalized alternating link on a connected projection surface Π with a cellular diagram and hat-representativity $\hat{r}(\pi(L), \Pi) > 4$. Then any essential Conway sphere has one of two forms: visible or hidden.*

A "visible" tangle has one *PPPP* curve in standard position in Menasco's language; a "hidden" tangle has two *PSPS* curves. See Section 12 for definitions. We expect this to lead to further study of tangles for knots projected onto surfaces beyond the 2-sphere.

1.3 Cyclic words associated to a surface

We extend standard position to closed surfaces and meridianal surfaces, but also to surfaces with non-meridianal boundary.

One of the benefits of standard position for classical alternating links is that it allows us to translate topological properties of surfaces into combinatorial properties of their curves of intersection with the (slightly modified) projection surface for a link. Such intersections were encoded in [20] by letters P (for link intesections) and S (for crossing ball intersections). Every curve of intersection was put in

correspondence with a cyclic word in such letters. Then possible words were investigated. In [7], this was extended further to surfaces with boundary, now with letters B (standing for arriving at the boundary of the surface) and S.

This combinatorial approach played an important role in several results for alternating links in the 3-sphere, including many noted above. In addition, it has been used to classify genus-two surfaces [27] and to find incompressible surfaces in alternating link complements in [21].

We generalize this to weakly generalized alternating links in this paper. The following theorem summarizes some of the main results.

Theorem 1.1 *Let L be a weakly generalized alternating link on a projection surface Π in a 3-manifold Y, and $(Z, \partial Z)$ an essential surface embedded in $(Y - N(L), \partial N(L))$. (In particular, this means L, Y, Π, and Z satisfy Assumption 6.1, and $Y - N(L)$ decomposes into two chunks as in Section 3.1.) Then the following hold.*

(1) *The surface Z can be isotoped into normal form with respect to the chunk decomposition of $Y - N(L)$, with any meridianal boundary components of ∂Z put into meridianal form (Theorem 6.2). The normal components of Z are denoted by Z_i, where ∂Z_i give words in the letters S (saddle intersections), B (boundary intersections), and P (meridianal form, where applicable).*

(2) *For Z meridianally incompressible, any SSSS disk is an essential compression disk for Π meeting the diagram of L exactly four times. Thus if the representativity satisfies $r(\pi(L), \Pi) > 4$, there are no SSSS disks (Theorem 9.1).*

(3) *No Z_i is a disk with boundary one of SS, PP, or PS (Theorem 10.1). No Z_i has a boundary component meeting an odd number of letters S and P (Theorem 10.2).*

(4) *If the hat-representativity satisfies $\hat{r}(\pi(L), \Pi) > 4$, and the diagram of L is not a string of bigons on Π, then a connected collection of BBBB disks is a disk, as is a connected collection of BBSS disks (Lemma 11.2). In particular, all subsurfaces Z_i that are neither BBBB disks nor BBSS disks, together with L, determine the surface Z up to isotopy (Theorem 11.4).*

The lemmas in Section 12 give additional restrictions on possible words when Z is an essential Conway sphere, when the link satisfies the hypotheses of Theorem 12.1.

In a subsequent paper, we apply this work to the problem of bounding the total number of embedded essential surfaces in these link complements [26]. We apply these results together with combinatorial and geometric arguments to give a polynomial upper bound on the number of embedded surfaces in a wide class of cusped 3-manifolds, namely weakly generalized alternating link complements and their Dehn fillings. Our upper bound only depends on the crossing number of the diagram, and is independent of the 3-manifold otherwise. These are the first bounds that explicitly capture the dependence on the 3-manifold besides those for classical alternating links in S^3 in [7; 8]; in other existing work, the way in which the bounds depend on the 3-manifold is not clear.

1.4 Organization

In Section 2 we recall the definition of weakly generalized alternating links. Their primary feature is recalled in Section 3, namely, they have a decomposition into checkerboard *chunks* that extend much of the useful decomposition of alternating links into topological polyhedra via checkerboard decomposition, which Thurston likened to gears of a machine [28]. We recall the definition of normal surfaces in chunks in Section 4. In Section 5 we describe the labeling of intersections of such surfaces with the chunk boundary by letters P, B, and S.

Our first main result, on meridianal form for normal surfaces, is proved in Section 6. In Section 7, we compile a few results on surfaces in standard form, including a result that shows that such a surface cannot meet the same saddle twice in a simple way.

In Section 8, we recall the notion of combinatorial area, introduced in this setting by Howie and Purcell [14]. Combining combinatorial area results with the labeling of boundary curves by letters P, B, and S, we find that the only zero area surfaces in a weakly generalized alternating link complement (with some restrictions) are disks with certain words on their boundaries. Sections 9, 10, and 11 are devoted to ruling out instances of zero area disks, or controlling when they occur. As a corollary of the work of Section 10, we prove the result on prime links.

In Section 12, we apply our results to essential Conway spheres to reproduce a result of Menasco [20], that such spheres are either "visible" or "hidden", with terminology due to Thistlethwaite [27].

2 Weakly generalized alternating link complements

In this section, we follow Howie and Purcell [14] to introduce a broad class of links that each have a diagram that is alternating on a closed surface Π embedded in a 3-manifold Y, possibly with boundary. As noted in the introduction, this is a very broad class that includes classical alternating knots, but also alternating knots on a Heegaard torus or more general Heegaard surface, originally studied by Adams [1] and Hayashi [11], virtual alternating knots [2; 3], and knots with alternating diagrams on any embedded surface in any compact orientable 3-manifold.

Assumption 2.1 Throughout this paper, we will always require the 3-manifold Y to be compact, orientable, and irreducible. The projection surface Π is required to be a closed, orientable surface. If Y has boundary, we will require ∂Y to be incompressible in $Y - N(\Pi)$, where $N(\cdot)$ always denotes a regular neighborhood. We further require $Y - N(\Pi)$ to be irreducible.

Given a link L in Y, the *link complement* is the manifold $Y - L$. The *link exterior* is the compact manifold $Y - N(L)$. If Y is closed, then the link complement is homeomorphic to the interior of the link exterior.

2.1 Generalized projection

A *generalized projection surface* Π is a (possibly disconnected) oriented surface embedded in Y so that $Y - \Pi$ is irreducible. The connected components of Π, denoted by $\Pi_1, \ldots, \Pi_p$, are closed two-sided

(orientable) surfaces. Let $N(\Pi) = \Pi \times (-1, 1)$ denote a regular neighborhood. For each component Π_i of Π, define $\Pi_i^{\pm}$ to be $\Pi_i \times \{\pm 1\} \subset N(\Pi)$. Denote $\bigcup \Pi_i^{+}$ by Π^{+} and similarly for Π^{-}.

Since $Y - \Pi$ is irreducible, if some Π_i is a 2-sphere, then Π is homeomorphic to S^2, and Y is homeomorphic to S^3. Let L be a link that can be projected onto Π in general position. That is, L can be isotoped through Y to lie in $N(\Pi)$ so that the image of the projection $\pi(L)$ consists of crossings and arcs between them on the surface Π. We call $\pi(L)$ a *generalized diagram*, or simply a *diagram*.

Whenever in the paper we mention Π or $\pi(L)$, we mean a generalized projection surface and the generalized diagram of L on Π. We will also use the following terms.

- An arc of $\pi(L)$ (on Π) between two crossings is called an *edge* of the diagram.
- A *crossing arc* is a simple arc in the complement of L running from an overpass to an underpass of a crossing.
- A *region* of the diagram is a complementary region of the projection of $\pi(L)$ to Π. It is bounded by edges of the diagram.

Every knot has a very simple generalized diagram on the torus boundary of a regular neighborhood of the knot. To ensure our diagrams are sufficiently complicated, in a way that depends on Y and Π, we introduce the notion of representativity. Define $r^{\pm}(\pi(L), \Pi_i)$ to be the minimum number of intersections between the projection of $\pi(L)$ onto $\Pi_i^{\pm}$ and the boundary of any essential compressing disk for $\Pi_i^{\pm}$ in $Y - \Pi$. If there are no essential compressing disks for $\Pi_i^{\pm}$ in $Y - \Pi$, then set $r^{\pm}(\pi(L), \Pi_i) = \infty$. The *representativity* $r(\pi(L), \Pi)$ is the minimum of all values of $r^{-}(\pi(L), \Pi_i)$ and $r^{+}(\pi(L), \Pi_i)$, over all i. By definition, a sphere S^2 embedded in S^3 admits no essential compressing disk, so a usual alternating diagram has representativity ∞.

Example 2.2 Figure 1, left, which is modified from a figure in [14], shows a diagram $\pi(L)$ on a torus Π. The representativity of the diagram depends on the manifold Y and the embedding of Π into Y. For example, first let $Y = S^3$, and embed the torus Π as the standard Heegaard torus for S^3, with the vertical red curve shown mapping to a meridian of one solid torus in the Heegaard splitting of S^3 and the horizontal red curve mapping to a meridian of the other solid torus. Then $r^{+}(\pi(L), \Pi) = 3$ and $r^{-}(\pi(L), \Pi) = 0$, so $r(\pi(L), \Pi) = 0$.

If instead we let $Y = T^2 \times [-1, 1]$, the thickened torus, and we embed the torus Π as the surface $T^2 \times \{0\}$, then Π admits no essential compressing disk, so $r^{+}(\pi(L), \Pi) = r^{-}(\pi(L), \Pi) = \infty$.

Finally, define the hat-representativity, $\hat{r}(\pi(L), \Pi)$, to be the minimum of

$$\bigcup_i \max\{r^{-}(\pi(L), \Pi_i), r^{+}(\pi(L), \Pi_i)\}.$$

Thus for example a connected surface Π with no compressing disks on one side will give infinite hat-representativity.

A generalized diagram $\pi(L)$ is said to be *alternating* if for each region of $\Pi - \pi(L)$, each boundary component of the region is alternating, i.e., it can be given an orientation such that crossings run from

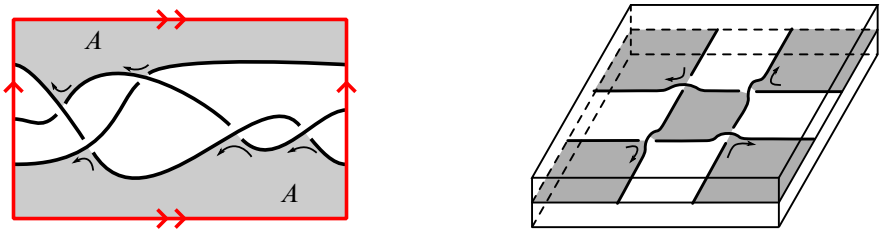


Figure 1: Left: an example of an alternating diagram on a torus. The representativity will depend on the embedding of the torus into Y. In any case, the diagram is not checkerboard colorable. Right: a checkerboard colorable diagram.

under to over in the direction of orientation. An alternating generalized diagram $\pi(L)$ is said to be *checkerboard colorable* if each region of $\Pi - \pi(L)$ can be oriented so that the induced orientation on each region's boundary is alternating: crossings run from under to over in the direction of orientation. Given a checkerboard colorable diagram, regions on opposite sides of an edge of $\pi(L)$ will have opposite orientations. We can color all regions with one orientation white, and all regions with the opposite orientation shaded.

Example 2.3 Figure 1, left, shows a diagram that is alternating on a torus. For the outer annular region A of the diagram, the figure shows an orientation assigned to its boundary. The crossings run from under to over with this orientation. Hence the region is alternating. However, this orientation is not induced by an orientation on A. If we choose an orientation on A, and then take the induced orientation on ∂A, one boundary component will be oriented so that crossings run under to over, and the other will be oriented so that crossings run over to under. Hence this example is not checkerboard colorable. Note it will not be checkerboard colorable regardless of the manifold Y it is embedded within.

However, Figure 1, right, shows another diagram of an alternating link on a torus, embedded in $Y = T^2 \times I$. This link diagram is checkerboard colorable.

Many results in the literature on alternating links require a reduced or simplified diagram; for example [7; 8; 20]. For a classical alternating link on $S^2 \subset S^3$, the condition we need is diagrammatic primeness: if an essential curve intersects the diagram exactly twice, then it bounds a region of the diagram containing a single embedded arc. Diagrammatic primeness rules out connected sums of knots and nugatory crossings. Analogously, a generalized diagram $\pi(L)$ on generalized projection surface $\Pi = \bigcup \Pi_i$ is *weakly prime* if whenever $D \subset \Pi_i$ is a disk with ∂D intersecting $\pi(L)$ transversely exactly twice, either the disk D contains a single embedded arc, or Π_i is a 2-sphere, $\pi(L)$ has at least two crossings on Π_i, and there is a single embedded arc of $\pi(L)$ on $\Pi_i - D$.

2.2 Weakly generalized alternating link

The diagram $\pi(L)$ on Π is said to be a *weakly generalized alternating link diagram* if

(1) $\pi(L)$ is alternating on Π,

(2) $\pi(L)$ is weakly prime,

(3) $\pi(L) \cap \Pi_i \neq \varnothing$ for each component Π_i of Π,

(4) each component of L projects to at least one crossing in $\pi(L)$,

(5) $\pi(L)$ is checkerboard colorable, and

(6) the representativity $r(\pi(L), \Pi)$ is at least 4.

We say that a link L is *weakly generalized alternating* if it has a weakly generalized alternating diagram.

Note these conditions were originally enumerated by Howie [12], and in some sense are as general as possible to obtain prime alternating diagrams on embedded surfaces in S^3 in Howie's setting. Perhaps the most mysterious condition is item (6), the representativity condition. The restriction on representativity means that small surfaces such as compressing disks and meridianal annuli must be split by the projection surface Π into disks parallel to Π, and thus they can be isotoped to lie entirely in $N(\Pi)$. The representativity condition is used throughout Howie–Purcell [14], and consequently used throughout this paper. Recall as well that classical alternating knots and virtual alternating knots have infinite representativity; the representativity condition only needs to be checked when Π is compressible, for example if it is a Heegaard surface.

From now on, every link we consider will be weakly generalized alternating. Note that a classical reduced, prime, alternating diagram of a link L on $\Pi = S^2$ in S^3 is an example of a weakly generalized alternating link diagram. The example of Figure 1, right, is also a weakly generalized alternating link diagram on $\Pi = T^2 \times \{0\}$ in $Y = T^2 \times [-1, 1]$.

These conditions are enough to guarantee that the link exteriors are irreducible and boundary irreducible [14, Corollary 3.16], which we will use below.

3 Decomposition of a link complement into chunks

Knot and link complements alternating on the projection sphere in S^3 have a well-known decomposition into topological polyhedra, suggested by W. Thurston and described by Menasco [19]; see also [16; 25, Section 11.1.1]. A more general decomposition into angled blocks was defined by Futer and Guéritaud [5]. This was generalized further by Howie and Purcell in [14] for weakly generalized alternating links. We review this generalization in this section.

3.1 A decomposition of weakly generalized alternating link complements

A checkerboard colorable diagram admits two checkerboard surfaces, one white and one shaded. After choosing a checkerboard coloring of the diagram, the white checkerboard colored surface is obtained by taking a surface corresponding to each white region of the diagram, and connecting these surfaces by twisted bands at crossings. Thus it embeds in the (open) link complement. Its intersection with the (compact) link exterior is a properly embedded surface with boundary on $\partial N(L)$. Similarly for the shaded surface. The decomposition of weakly generalized alternating link complements is obtained by cutting along the white and shaded checkerboard surfaces. Precisely, remove an open regular neighborhood

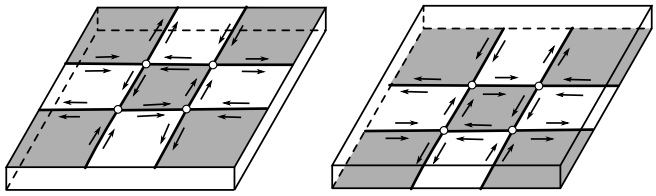


Figure 2: An example of a chunk decomposition. On the left is a manifold homeomorphic to $T^2 \times [-1, -\epsilon]$ with faces, edges, and ideal vertices marked on $T^2 \times \{-\epsilon\}$. On the right is a manifold homeomorphic to $T^2 \times [\epsilon, 1]$ with faces, edges, and ideal vertices marked on $T^2 \times \{+\epsilon\}$.

of the checkerboard surfaces from the link exterior. Since the regions of these surfaces tile Π, this cuts $Y - N(L)$ into components corresponding to $Y - N(\Pi)$, with $\partial N(\Pi)$ decorated with portions of checkerboard surfaces and remnants of $\partial N(L)$. The checkerboard surfaces intersect exactly at crossing arcs. After cutting, each crossing arc gives rise to four ideal edges lying on $\partial(N(\Pi))$, two in each of $\Pi^{\pm}$. Strands of the knot corresponding to overcrossings (resp. undercrossings) become ideal vertices in Π^+ (resp. in Π^-); we contract each of these to lie at a crossing of the diagram. More details are in [14].

We record here the results of Propositions 3.1 and 3.3 of [14], which prove that the decomposition has the following properties. The components of the decomposition, which are called *chunks*, are homeomorphic to connected components of $Y - N(\Pi)$. These are 3-manifolds with boundary, where the boundary components are components of ∂Y, along with Π_i^- and Π_i^+. Note that each Π_i lies on the boundary of one or two chunks, appearing as Π_i^+ and Π_i^-. When Π is connected, we have one or two chunks in total, depending on whether Π is separating or not. For example if Π is the usual projection sphere S^2 for a classical alternating link diagram, there are two chunks, each a ball component of $S^3 - N(S^2)$.

Decorate the surfaces Π_i^- and Π_i^+ with:

(1) A copy of the edges of the link diagram on Π_i. These will be called *interior edges* of a chunk, and correspond to crossing arcs in $Y - N(L)$. There will be four such ideal edges in $\partial N(\Pi)$ per crossing arc, two in each of $\Pi^{\pm}$.

(2) Interior edges meet at crossings of the diagram, which become *ideal vertices*. That is, crossings become vertices, which are removed (ideal).

(3) Regions bounded by edges and vertices will be called *faces* of the chunk. They are not necessarily simply connected. The faces correspond to regions of the diagram, and will be checkerboard colored. For example, the outer region of Figure 1 is an annulus, so gives rise to a face that is an annulus.

An example of the chunk decomposition for the link from Figure 1, right, is shown in Figure 2. The two chunks are the components of the complement of $T^2 \times (-\epsilon, \epsilon)$ in $T^2 \times [-1, 1]$ for some $\epsilon > 0$. Hence they are homeomorphic to $T^2 \times [-1, -\epsilon]$ and $T^2 \times [\epsilon, 1]$. Edges, faces, and ideal vertices are marked on $T^2 \times \{-\epsilon\}$ and $T^2 \times \{\epsilon\}$, respectively, with faces shown checkerboard colored.

Associated to a weakly generalized alternating link is a chunk decomposition corresponding to a weakly generalized alternating diagram constructed as above. We will assume throughout that this is the chunk decomposition used in this paper.

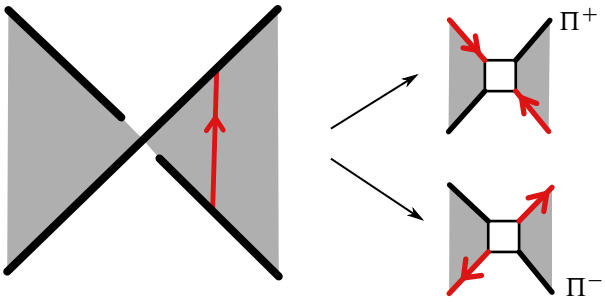


Figure 3: Four edges are identified to a crossing arc, with two on each of Π^- and Π^+. The edges meet as opposite edges at a vertex.

3.2 Gluing

To obtain $Y - L$ from the chunks, each face of Π_i^- should be glued to the corresponding face of Π_i^+. A face F^- of Π_i^- corresponds to a face F^+ of Π_i^+ if the same region of $\pi(L)$ gave rise to F^- and F^+. The gluing is by the identity away from boundary components of the faces, i.e., away from the edges of the chunk(s). In a neighborhood of a boundary component of the face, the gluing rotates the ideal edges by one notch either clockwise or counterclockwise, depending on whether the face is white or shaded. Namely, when viewing $\Pi_i^\pm$ from $Y - N(\Pi)$ near Π_i^+, an ideal edge in a white face will rotate to the next ideal edge of the same face in the clockwise direction for the gluing, and similarly for shaded faces in the counterclockwise direction. The arrows on the faces of the chunk in Figure 2 indicate the gluing.

Under the gluing, four interior edges glue to a single *crossing arc* in $Y - L$. The crossing arc is identified to two edges each on Π^- and Π^+, with the edges meeting as opposite edges at a vertex. This is illustrated in Figure 3.

Now truncate the ideal vertices of chunks: this replaces an ideal vertex with a quadrilateral *truncation face*. We call an edge bordering a truncation face a *truncation edge*. The truncation faces correspond to crossings of the link diagram, and tile the boundary torus $\partial N(L)$ with a *harlequin tiling*. See Figure 4. Note the slightly different terminology here: such faces and edges are called boundary faces and edges in [14].

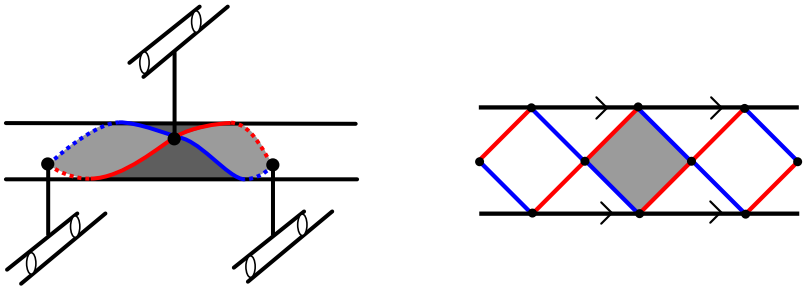


Figure 4: On the left is shown a single truncation face (shaded) as it appears embedded in the link complement. On the right, the boundary torus of the link has been unrolled into an annulus; truncation faces are shown.

A chunk with truncated ideal vertices is called a *truncated chunk*. Below, if we refer to faces of a chunk, we mean both truncation faces and other faces. Similarly, if we refer to an edge of a chunk, we mean either a truncation edge or an interior edge. We note that under the gluing described above, the truncation faces that are adjacent to an interior face I are also rotated one notch around I, either clockwise or counterclockwise, together with the boundary of ∂I.

4 Normal surfaces

Both normal and standard position begin with a surface being transverse: normal surfaces are transverse to faces, edges, and vertices of tetrahedra in a triangulation of a 3-manifold, and surfaces in standard position are transverse to a projection surface of a link away from crossings. Here we review surfaces that are normal with respect to a chunk decomposition, introduced in [14], which generalizes both of the above.

Assumption 4.1 We use the topological definitions of incompressible, boundary incompressible, and essential surfaces:

- A surface Z that is neither a disk nor a 2-sphere is incompressible in a 3-manifold M if, for any embedded disk $D \subset M$ with $D \cap Z = \partial D$, ∂D bounds a disk in Z. A surface that is not incompressible admits an essential compression disk, namely an essential disk $D \subset M$ with $D \cap Z = \partial Z$, for which ∂D is an essential simple closed curve on Z.
- A boundary compression disk for a properly embedded nondisk surface Z in M is an embedded disk $D \subset M$ with ∂D consisting of two arcs: $\alpha = D \cap Z$ and $\beta = D \cap \partial M$, with $\alpha \cap \beta = \partial\alpha = \partial\beta$. The boundary compression disk is essential if α does not cobound a disk in Z with another arc in ∂Z. If there exists an essential boundary compression disk for Z, then Z is boundary compressible. Otherwise, it is boundary incompressible.
- A 2-sphere is incompressible if and only if it does not bound a 3-ball.
- By convention, disks will be neither incompressible nor compressible, and neither boundary incompressible nor boundary compressible in this paper.
- A surface is essential if it is incompressible, boundary incompressible, and not boundary parallel.
- Note we allow Z to be orientable or nonorientable, with or without boundary.

4.1 Normal surface in a chunk

Consider a surface Z', possibly with boundary, properly embedded in a truncated chunk C, with $\partial Z' \subset \partial C$.

Definition 4.2 The surface Z' is *normal* with respect to the chunk C if it satisfies the following.

(0) Each nondisk component of Z' is incompressible in C.

(1) Z' and $\partial Z'$ are transverse to all faces, edges, and vertices of C.

(2) If a component of $\partial Z'$ lies entirely in a face of C, then it does not bound a disk in that face.

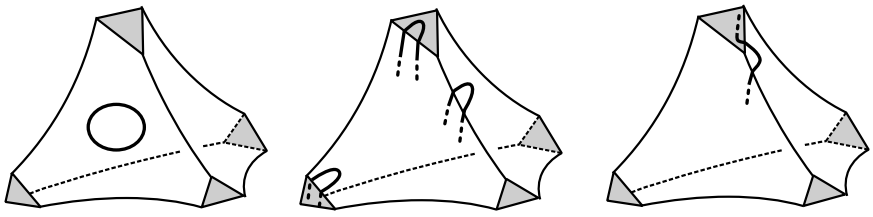


Figure 5: Shown are arcs of $\partial Z' \cap \partial C$ that make the surface Z' fail to be normal. On the left, the surface fails (2), in the middle it fails (3), on the right it fails (4).

(3) If an arc γ of $\partial Z'$ in a face of C has both endpoints on the same edge, then the arc γ along with an arc of the edge cannot bound a disk in that face.

(4) If an arc γ of $\partial Z'$ in an interior face of C has one endpoint on a truncation edge and the other on an adjacent interior edge, then the union of γ as well as adjacent arcs of the two edges cannot bound a disk in the interior face.

Example of arcs that make a surface fail to be normal are shown in Figure 5.

4.2 Normal surface with respect to a chunk decomposition

Given a chunk decomposition of $Y - L$, a surface Z embedded in $Y - L$ is *normal* with respect to the chunk decomposition if for every chunk C, the intersection $Z \cap C$ is a (possibly disconnected) normal surface in C.

Here and further we subdivide a surface Z in subsurfaces Z_i cut out by faces of the chunks. We assume that each Z_i is connected, closed or with boundary, and possibly with multiple boundary components. All our surfaces and subsurfaces from now on will be embedded. Moreover, we will assume that if the surface Z has boundary and Y has boundary, then the boundary of Z lies on $N(L)$ and is disjoint from ∂Y. This includes surfaces of any slope on L, as well as closed surfaces.

It was shown in [14, Theorem 3.8] that any essential surface in a 3-manifold with a chunk decomposition can be put into normal form with respect to the chunk decomposition; see also [5, Theorem 2.8]. We will revisit this proof below in Theorem 6.2 to show that the process of putting surfaces into normal form preserves other desirable features of the surface.

We note that item (0) of Definition 4.2 is slightly stronger than the definition in [14, Definition 3.7], in that we require all nondisk components Z_i, and not just closed components, to be incompressible. However, we will see in Theorem 6.2 below that any surface that is normal with respect to the definition of [14] can be isotoped to be normal with respect to Definition 4.2.

5 Saddles, meridian punctures, and boundary of the surface

For normal subsurfaces Z_j of the surface Z, each boundary component of ∂Z_j runs over truncation edges and interior edges of the respective chunk. In this section we will describe a labeling of the components of ∂Z_j by letters P, S, and B, analogous to similar labelings in [7; 8; 20].

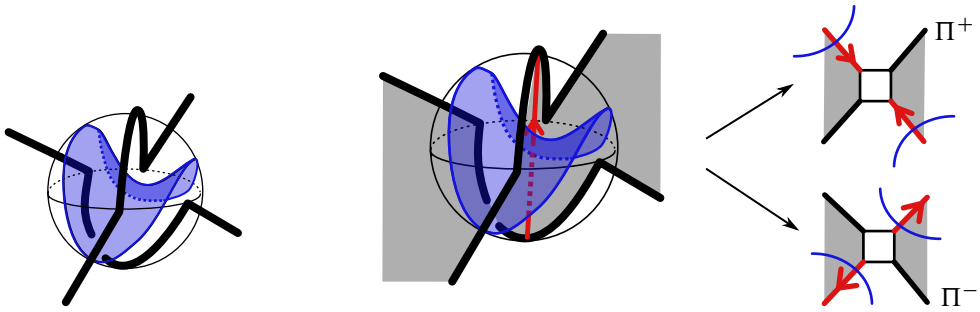


Figure 6: Left: a saddle runs between arcs of the diagram at a crossing. Right: this leads to Z intersecting interior edges of the chunk diagram. Shown are intersections of interior edges on Π^+ and Π^-.

In these other papers, an essential surface Z in an alternating link complement in S^3 is studied by considering its curves of intersection with the projection sphere and with small balls around crossings called *crossing balls*. The surface in standard position from [7; 8; 20] intersects the projection sphere where Z has *saddles* inside crossing balls. See Figure 6, left. In [7; 8; 20], each saddle is labeled with the letter S.

For a weakly generalized alternating link, if a surface Z has a saddle, that saddle intersects a crossing arc exactly once. Recall from Section 3.2 that each interior edge of the chunk decomposition is identified to exactly one crossing arc, with four interior edges in total identified to a single crossing arc. Thus saddles meet interior edges as in Figure 6, right. Shown in that figure are the four edges that glue up to the crossing arc (red), and the way the saddle surface meets them. For a surface in normal form, we consider how the surface intersects interior edges and truncation edges. Because interior edges are identified to crossing arcs, and a surface meeting a crossing arc runs through a saddle there, analogous to the notation of [20], we label each intersection of ∂Z_j and an interior edge with an S.

For a surface Z in a link exterior $Y - N(L)$, a *meridianal compression disk for* Z is a disk D embedded in Y such that $D \cap Z = \partial D$, and the interior of D intersects L exactly once transversely. The meridianal compression disk is *essential* if ∂D is not parallel through (an annulus in) Z to a meridian of $\partial N(L)$; otherwise it is *inessential*. The surface Z is *meridianally compressible* if it admits an essential meridianal compression disk. For such a surface, performing surgery along $D - N(L)$ yields a new surface whose boundary is the union of ∂Z and two meridians on $\partial N(L)$; this is called a *meridianal compression* of Z.

Both for classical alternating links in S^3 [20] and for weakly generalized alternating links under certain conditions [14, Lemma 4.9], every closed surface Z' is meridianally compressible. After meridianal compressions the resulting surface Z meets the diagram in meridians. In [7; 20] for links in S^3, each meridian on ∂Z is labeled with a P, which stands for a meridianal *puncture*. For weakly generalized alternating links, when ∂Z again consists only of meridians, each meridian will intersect two truncation faces. We also wish to label intersections with a P, but the setup will be slightly different. To describe our labeling, we first need the following definition.

Definition 5.1 Isotope Z to meet the diagram $\pi(L)$ transversely away from crossings. After this isotopy, in the chunk decomposition each meridianal curve of ∂Z meets exactly two truncation faces, which are

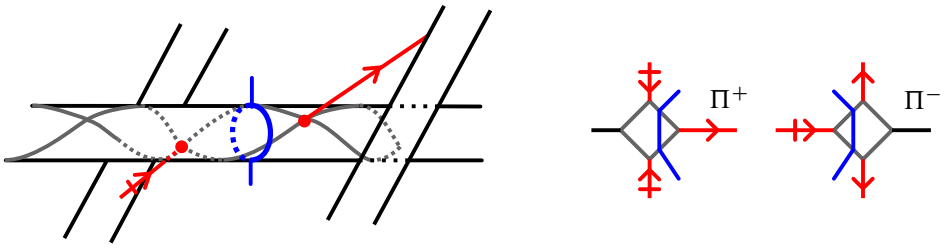


Figure 7: Left: isotope Z with meridianal boundary to meet $N(L)$ transversely away from crossings. Right: in the chunk decomposition, such a curve meets exactly two truncation faces, and cuts off a single corner of each truncation face. Note that the corner cut off by a meridian is not one of the corners that are identified.

quadrilaterals. It runs through adjacent truncation edges on each face, cutting off a single corner of each of two quads in the harlequin tiling of the boundary; see Figure 7. Moreover, note that two opposite vertices of a truncation face are always identified (they lie on the same endpoint of a crossing arc). The meridian does not cut off such a vertex, but rather cuts off one of the other two vertices of a truncation face. We say that a meridianal component of ∂Z is in *meridianal form* if it

(1) meets exactly two truncation faces, one on each side of the projection surface;

(2) runs between adjacent edges in each such truncation face, cutting off a vertex that is not identified to another corner on the same truncation face.

See Figure 7. We say that Z is in meridianal form if each meridianal component of ∂Z is in meridianal form.

Suppose Z is a surface in meridianal form and all components of ∂Z are meridians. Assign a label P to each intersection of ∂Z with a truncation face. Note just as for a classical alternating link in S^3, a meridianal puncture in Z corresponds to a single P above the diagram on Π^+, and a single P below on Π^-.

For a spanning surface Z of an alternating link in S^3, Hass, Thompson, and Tsvietkova [8] introduce the letters B. A letter B indicates where Z intersects the link transversally on the projection sphere. In the case of weakly generalized alternating links, a surface Z with boundary will have normal subsurfaces Z_j with ∂Z_j meeting truncation faces. When the corresponding curve on ∂Z is not necessarily a meridianal curve in meridianal form, we label each intersection of ∂Z_j with a truncation edge by B. Note that above, we labeled intersections of ∂Z_j and truncation faces by P. For surfaces with nonmeridianal boundary, we consider intersections with truncation edges rather than faces. The letter B is a reminder that the surface might have boundary components that are not meridianal.

The labeling of components of ∂Z_i (which are curves) by letters S, B, or P associates a cyclic word to every such curve. An example is shown in Figure 8.

Remark 5.2 The arguments below for surfaces with labels B on truncation edges work equally well when we replace an instance of P with two instances of B, and so we allow curves with B labels also

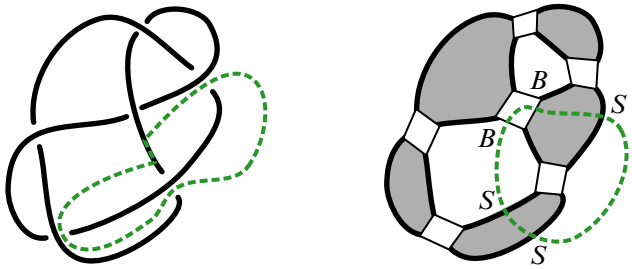


Figure 8: On the left is shown an example of a *BBSSS* curve from [8]. On the right is shown the corresponding *BBSSS* curve in a chunk of the decomposition of the same link. The curve on the right meets the same saddles (interior edges) and link overcrossings (truncation faces) as the curve on the left.

to be in meridianal form. This will be useful for links, for example, when some components of ∂Z are meridians and some are not. In this mixed case, with some meridian boundary components and some nonmeridians, we will label all intersections with truncation edges by B. However, when a surface has strictly meridianal boundary, we obtain more information using only labels P. We do not allow words in both P and B.

For classical alternating links, the curves of intersection subdivide the surface into disks lying in topological 3-balls above and below the link. The curves of intersection and the position of the disks are uniquely determined by the associated words in S, P, and B and by the position of each letter on the link diagram. The disks are then glued along their boundaries, and the gluing pattern is also uniquely determined by the words in S, P, and B and by the position of each letter on the link diagram. For weakly generalized alternating links, there are also words associated to boundary components of Z_i, and the position of each letter on the knot projection. But the boundary curves subdivide the surface Z into subsurfaces Z_i that might have positive genus, and multiple boundary components. Moreover, the subsurfaces are not in topological 3-balls anymore: rather, they are in chunks which might have complicated topology themselves.

6 Normal form, saddles, and meridians

For a surface Z, associate a pair of natural numbers $(s(Z), t(Z))$ where $s(Z)$ is the number of intersections of Z with interior edges, and $t(Z)$ is the number of intersections with truncation edges. Call the pair $(s(Z), t(Z))$ the *weight* of the embedding with respect to a chunk decomposition, and order lexicographically. In this section, we ensure that a surface can be isotoped into normal form while preserving meridianal form for the surface, by an isotopy that does not increase weight. Note that there are many (isotopic) ways to put a surface in normal form or standard position, and in applications, often the one that minimizes weight is chosen.

For convenience, we collect assumptions here. They will be used throughout the rest of the paper, and will be part of the hypothesis in each lemma, proposition and theorem we prove.

Assumption 6.1 • Let Y and Π satisfy Assumption 2.1. We assume that $\pi(L)$ is a weakly generalized alternating projection of a link L onto Π in Y.

• We consider $(Z, \partial Z)$ to be an essential surface (either orientable or nonorientable) embedded in $(Y - N(L), \partial N(L))$, where the boundary of Z is possibly empty. Note the boundary of Z (if any) is disjoint from ∂Y.

• We also assume that there is a chunk decomposition of $Y - N(L)$ arising from a weakly generalized alternating diagram as in Section 3.1. The Z_i are the connected subsurfaces of Z cut out by chunks, each closed or with boundary, and possibly with multiple boundary components.

Theorem 6.2 *A surface Z in $Y - N(L)$ can be isotoped into normal form such that*:

(a) *The full isotopy consists of a sequence of subisotopies given by discrete steps, and no step of the isotopy increases the weight.*

(b) *If all meridian components of ∂Z begin in meridianal form, then after the isotopy, meridianal components of ∂Z remain in meridianal form.*

Proof The proof that an isotopy exists putting the surface into normal form is a standard innermost disk / outermost arc argument that is similar to arguments in [5; 14], but we provide it to verify (a) and (b).

We need to check the requirements of Definition 4.2. A small isotopy of Z ensures transversality conditions (1). We may ensure this isotopy does not increase the number of intersections with any edge, for (a), and does not affect meridianal form for (b).

Since Z is essential, condition (0) automatically holds for closed surfaces in any chunk, without any isotopy. If D is an essential compressing disk for $Z \cap C$ within a chunk C, then because Z is incompressible, ∂D bounds a disk E in Z, and by irreducibility of Y, $D \cup E$ bounds a ball. Hence E can be isotoped through the ball and past E to remove the compressing disk. For (a), we say that this full isotopy through the ball is one step. Note the weight may increase temporarily during the isotopy through the ball, but when this step of the isotopy is completed, the number of intersections of Z with interior edges has not increased. The isotopy does not affect the boundary of Z at all, completing (a) and giving (b).

For condition (2) of Definition 4.2, suppose some Z_k has boundary ∂Z_k lying in a face of C and bounding a disk in that face. Take an innermost such disk D. It is not an essential compressing disk for Z in Y because Z is incompressible. Because Y is irreducible, there is a ball B with $D \subset \partial B$ and $\partial B - D \subset Z$. Isotope Z through B and slightly further, removing all intersections of Z with faces and interior edges that lie within the ball B, and removing the intersection of ∂Z_k on a face of C, all without changing Z outside a small neighborhood of B. Again the full isotopy through B is considered to be one step, for (a). Note this move cannot increase the number of intersections of Z with interior edges, and avoids the boundary of Z entirely, so it cannot increase the weight for (a), and does not affect meridianal form for (b).

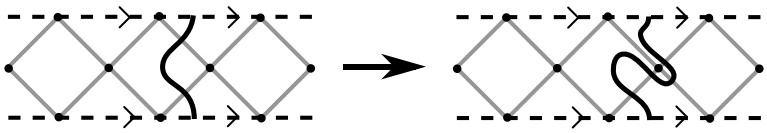


Figure 9: A curve of ∂Z forming a meridian on the harlequin tiling of the boundary is shown on the left. If we isotope across a disk that violates condition (4), the curve is changed as shown on the right. Note further isotopy is required to put it into normal form, meridianal form.

For condition (3) of Definition 4.2, suppose for some Z_k, an arc of ∂Z_k lies in a single face with endpoints on the same edge, and together with a part of the edge cuts off a disk D in that face. Assume D is innermost with this property in that face, meaning its interior does not intersect Z.

There are three cases depending on the type of edge and the type of face.

Suppose first that the arc of ∂Z_k has endpoints on the same interior edge. Then use a regular neighborhood of the disk D to isotope Z past that edge (a step of the isotopy), strictly reducing the number of intersections with the interior edge, and not affecting ∂Z, giving (a) and (b).

Next suppose that the arc of ∂Z_k lies in a truncation face, with both endpoints on the same truncation edge. Then again use a regular neighborhood of D in $Y - N(L)$ to isotope ∂Z along this truncation face and past the edge, removing two intersections with truncation edges. This move is a step of the full isotopy. It does not affect intersections with interior edges, and strictly decreases intersections with truncation edges, giving (a). Such an arc is not in meridianal form, so does not arise for (b).

Finally suppose the arc α of ∂Z_k has both endpoints on the same truncation edge but lies in an interior face. Then the disk D has the form of a boundary compression disk for Z. Since Z is boundary incompressible, there must be an arc $\beta \subset \partial Z$ whose endpoints agree with those of α, and such that $\alpha \cup \beta$ bounds a disk D' in Z. Then $D \cup D'$ forms a disk with boundary on $\partial N(L)$. Because $Y - N(L)$ is boundary irreducible, by [14, Corollary 3.16], the disk $D \cup D'$ must cobound a ball with a disk of $\partial N(L)$. Use this ball to isotope D' in Z past D to remove the two intersections with the truncation face, as well as any other intersections of Z with edges and faces within the ball. For (a), the isotopy through this ball is a step, and when complete, the number of intersections with interior and truncation edges does not increase. As for (b), the arc β of ∂Z cannot be in meridianal form, since both its endpoints are on the same truncation edge, so this situation does not arise for (b).

Condition (4) of Definition 4.2 requires the most care, since it could arise for surfaces in meridianal form. Nevertheless, we show that an isotopy also addresses this case while preserving meridianal form and without increasing weight. The argument requires careful consideration of the combinatorics of the chunk decomposition around a truncation face and an adjacent interior edge. We step through it below, using a number of figures.

Suppose an arc γ of ∂Z_k in an interior face of a chunk C has an endpoint on a truncation edge and an endpoint on an adjacent interior edge, cutting off a disk D with these two edges. We may take γ to be outermost with this property, so that the interior of D is disjoint from Z. Isotope Z through a regular neighborhood of D, by sliding (a neighborhood of an arc of) ∂Z along the adjacent truncation faces and

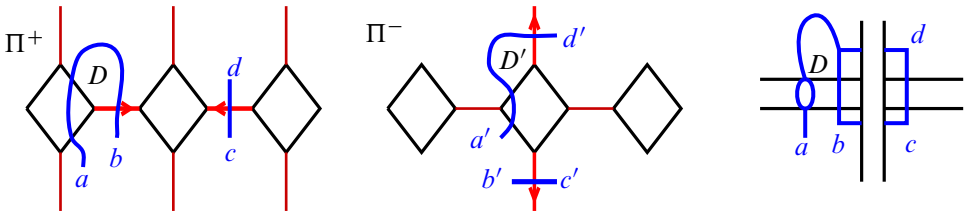


Figure 10: When Z is in meridianal form, and an arc runs from a truncation edge to an adjacent interior edge, Z must intersect Π^+ and Π^- as shown on the left (possibly with roles of Π^+, Π^- switched). The right shows the saddle and the adjacent meridian in $Y - N(L)$.

past the interior edge, then sliding the rest of a neighborhood of γ in Z through a neighborhood of D to follow. The effect of this move on the harlequin tiling of the boundary is shown in Figure 9. This removes an intersection of Z with an interior edge, giving (a). This will introduce an arc in a truncation face with both endpoints on the same truncation edge, as in Figure 9. Such an arc can be eliminated as above, reducing weight. Thus if the surface does not begin in meridianal form, we are done at this point.

To complete the proof of (b), we need to analyse the isotopy of Z through $N(D)$ more carefully, ensuring that if D is adjacent to a truncation face that Z meets in meridianal form, then after the isotopy the result is still in meridianal form.

Suppose Z is in meridianal form. We set up some notation. Assume that $D \subset \Pi^+$. The arc γ of ∂Z_k that cuts off D continues through the truncation face in meridianal form. Further, because the interior face containing D is glued to another interior face on Π^-, there is another such arc cutting off a disk $D' \subset \Pi^-$, and that arc extends to run through another truncation face in meridianal form. Indeed, these two arcs in meridianal form together form a meridian boundary component of ∂Z. Finally, on each of Π^+ and Π^- there is another interior edge identified to the one meeting D or D'. The surface Z must also run through that interior edge. The setup must therefore appear as in Figure 10.

That is, there are arcs ab and cd of $Z_k \cap \Pi^+$ meeting interior edges identified to the same crossing arc of a saddle, and a subarc of ab is an arc of the boundary of the disk D. There are additional arcs $a'd'$ and $b'c'$ in Π^-, also meeting interior edges identified to the same crossing arc, with an arc of $a'd'$ forming an arc of the boundary of disk D'. Moreover, a is glued to a', b to b', c to c' and d to d'.

For the leftmost and middle pictures in Figure 10, note that the endpoints of arcs labeled by a, b, c, d are in the same quadrants as a', b', c', d' respectively. This is because two interior faces that are identified correspond to the same region of the link diagram. Also note that the arcs ab, cd, $a'd'$, $b'c'$ and their

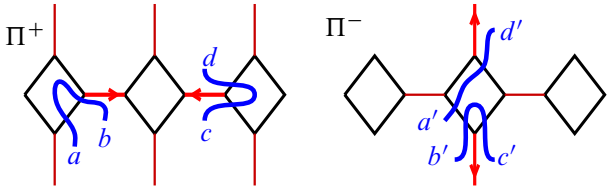


Figure 11: The isotopy through D and D' has the effect shown.

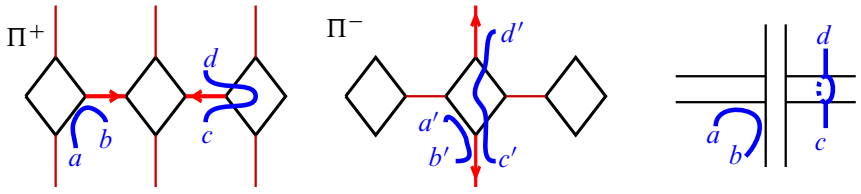


Figure 12: Isotoping further into normal form yields surface in meridianal form.

intersection pattern with interior edges and truncation edges of the chunks correspond to rotating the boundary of every interior face under the gluing one notch, as described in Section 3.2.

From the middle picture in Figure 10, rotate quadrant I one notch counterclockwise and identify it with quadrant I in the leftmost picture. Similarly, rotate quadrant III one notch counterclockwise and identify it with quadrant III in the leftmost picture. Here interior edges are rotating to the next interior edges in the counterclockwise direction, and truncation edges are rotating to the next truncation edge in the counterclockwise direction. Rotate quadrants II and IV one notch clockwise to the leftmost picture.

Now we isotope Z across the disk D, to remove intersections with the interior edge identified to the crossing arc. Note this simultaneously isotopes across D' in the other chunk. This isotopes arcs cd and $b'c'$ to run through the adjacent truncation faces; this is shown in Figure 11. Away from these arcs, the intersection of the surface Z with Π^+ and Π^- is unchanged.

Note that the surface is not in normal form; on the left of Figure 11, an arc of $\partial Z'$ in the truncation face meets the same truncation edge twice. We perform an isotopy to slide this arc off this truncation face and onto the truncation face on Π^-, shown in the center on the right of Figure 11. The final result of this isotopy, in Π^-, Π^+ and in $Y - N(L)$, is shown in Figure 12.

The surface is now again in meridianal form, for (b). In the link complement, the entire procedure swept a meridian past a crossing arc, removing the intersection with the corresponding interior edge without introducing new intersections with interior edges, therefore decreasing weight. We call this a step of the isotopy, and note (a) also holds for this step. □

Assumption 6.3 From now on, we assume the essential surface $(Z, \partial Z)$ is in normal form. Moreover, when we put it into normal form,

(1) we do it so that the surface is in meridianal form;

(2) out of all ways to do it so that the surface is in meridianal form, we choose one with least weight.

7 Generalizing standard position

In this section we generalize properties of standard position for closed surfaces from [20], and for surfaces with nonmeridianal boundary from Propositions 2.1–2.2 of [22].

Proposition 7.1 *The following holds for a surface Z in $Y - N(L)$, and any connected subsurface Z_i of Z in a chunk C:*

(1) *Let R be a component of $\partial N(L)$. If $\partial Z \cap R$ is nonempty and nonmeridianal, then there are no meridianal compressions of Z to R.*

(2) *Z_i cannot be a sphere or projective plane.*

(3) *If Z is in normal position, each curve of ∂Z_i meets truncation edges in pairs. Thus labels B occur in pairs.*

Proof For (1), suppose Z is a surface with nonmeridianal boundary on a component R of $\partial N(L)$, and suppose that Z admits a meridianal compression to R. Such a compression defines an embedded annulus A with one boundary component on Z and one on R, with interior disjoint from Z. But a meridian intersects any essential nonmeridianal closed curve on the link boundary torus. If ∂Z meets this component R of $\partial N(L)$ and is nonmeridianal, ∂Z must intersect the annulus A, and hence Z meets the interior of A. (Here we are using incompressibility and boundary incompressibility of $(Z, \partial Z)$ to ensure intersections of ∂Z with A cannot be removed.) This contradicts the definition of A.

For (2), recall that Π was chosen such that $Y - \Pi$ is irreducible, thus any chunk is irreducible. Therefore a spherical component cannot bound anything but a ball. But a 2-sphere is incompressible if it does not bound a ball. Hence item (0) of Definition 4.2 implies that the intersection of a normal surface with a chunk has no spherical components. Similarly, because $Y - \Pi$ is orientable, the boundary of a regular neighborhood of any embedded incompressible projection sphere would be an essential 2-sphere, which cannot exist in the irreducible $Y - \Pi$.

For (3), every time ∂Z_i meets a truncation edge, it runs into a truncation face and then must meet another truncation edge as it runs out of that truncation face. □

The next result is analogous to Menasco's result [20, Lemma 1] that no subsurface in standard position meets the same crossing bubble in more than one arc. See also Menasco and Thistlethwaite [22, Proposition 2.2(ii)].

Fix a truncation face on a chunk. The face is adjacent to exactly two interior edges, say e_1 and e_2, that are identified together. Denote by γ the arc formed from the union of e_1, e_2 and (a choice of) two truncation edges connecting e_1 with e_2. See Figure 13(a). We call the connected simplex consisting of these four edges a *saddle complex*.

Proposition 7.2 *Suppose Z is meridianally incompressible. Suppose an arc α of $Z \cap \partial C$ intersects a saddle complex γ in exactly two points. Then α cannot cobound a disk with a subarc β of γ.*

Proof Suppose the arcs cobound a disk. There are six ways that an arc α might cobound a disk with a subarc of γ, shown in Figure 13. Take D to be an innermost disk with the property that ∂D consists of an arc α on $Z \cap \partial C$ and a subarc β of γ. That is, $D \cap Z = \alpha$.

Suppose that for D innermost, α has both endpoints on the same interior edge e, as in (a) of Figure 13. If α happens to lie in only one face, then α gives a contradiction to our assumption that Z be in normal form, Assumption 6.3. So α lies in more than one face. Then D intersects multiple faces and edges. In this case, we will isotope Z to remove two intersections of Z with e, illustrated in Figure 14.

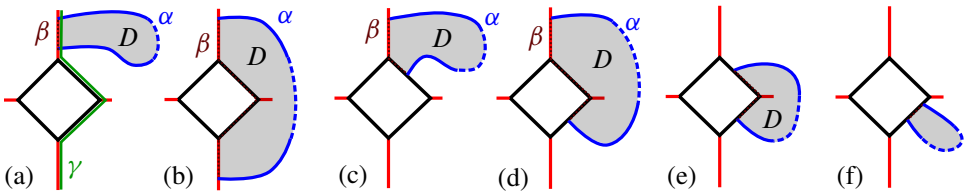


Figure 13: Shown are the ways that an arc α of $Z \cap \partial C$ can cobound a disk D with a subarc β of γ, where γ is a saddle simplex.

Carefully, push the disk D slightly into the interior of the chunk, obtaining a new disk D', shown in darker gray on the left of Figure 14. Then isotope Z in a small regular neighborhood of D', removing two intersections with e, as shown on the right of the figure. Note that within $Y - N(L)$, the edge e is identified to other edges to obtain a crossing arc that contains β, and β runs between two saddles of Z intersecting the crossing arc. This move eliminates the intersections of Z at endpoints of β. Note Figure 14 does not show what happens in $Y - N(L)$, only the chunk decomposition, prior to the edge identifications. We call the move of Figure 14 a *band move*, as in [22, Figure 2.3(iii)].

Alternatively, the isotopy of a band move is equivalent to the following move. Take a small product neighborhood of D'. This is a ball $B = D' \times (-\epsilon, \epsilon)$, for some sufficiently small $\epsilon > 0$. The boundary ∂B consists of a disk E on Z of the form $E = \alpha' \times (-\epsilon, \epsilon)$, and a second disk $F = \partial B - E$. Replace the surface Z by removing the disk E from Z and replacing it with the disk F. Push slightly through the boundary of the chunk. The resulting surface is isotopic to F, but meets the edge e two fewer times.

The resulting surface is still essential and meridianally incompressible, and any of its meridian boundary curves are still in meridianal form, so it can be isotoped further into normal form without increasing intersections with interior edges as in Theorem 6.2. This is a contradiction, since Z was assumed to have minimal weight.

Next suppose that for D innermost, α has endpoints on distinct interior edges of γ, as in (b) of Figure 13. Then glue interior edges to obtain a crossing arc. Because the disk D is innermost, the two endpoints of α on interior edges are then identified. The arc of β on the truncation faces forms a meridian. Thus

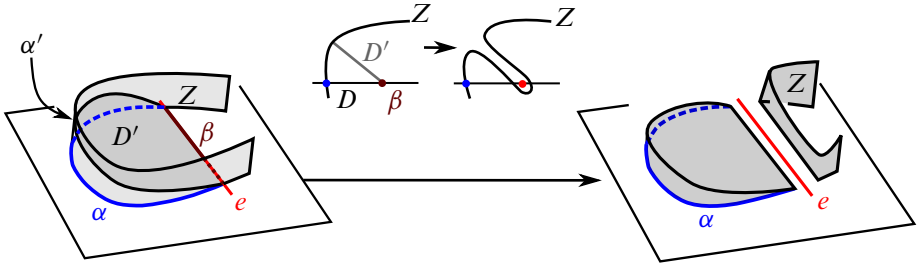


Figure 14: Shows how we isotope Z through the disk D' to remove intersections with e, in 3-dimensions and a 2-dimensional cross section. In the leftmost and rightmost figures, e is the edge in red, with β slightly darker red on e. In the center, a cross-section is depicted, and e becomes a single point.

after gluing, $\alpha \subset Z$ bounds a meridianal compression disk. Because Z is meridianally incompressible, the meridianal compression disk must be inessential, and α is parallel through an annulus $A \subset Z$ to a meridian component of ∂Z, and together $D - N(L)$ and this annulus A cobound a thickened annulus. Isotope Z through this thickened annulus, past D, removing all intersections of Z with faces, edges, and truncation edges within the thickened annulus. The number of intersections of Z with interior edges strictly decreases, and the meridian component of ∂Z slides into meridianal form on this truncation face and an adjacent one. The resulting surface is still essential and meridianally incompressible, and still in meridianal form, so it can be isotoped further into normal form in a manner that does not increase intersections with interior edges, using Theorem 6.2. This contradicts the minimum weight assumption.

Next suppose α has both endpoints on truncation edges, as in (e) or (f) of Figure 13. (Note that in the classical setting, this case is automatically excluded by the definition of standard position [22].) Away from β, push D slightly into the interior of the chunk C, similar to the move on the left of Figure 14. Then the newly isotoped disk has the form of a boundary compression disk D' for Z. Since Z is boundary incompressible, we may push off this disk using a standard argument (see for example [10, Figure 1.9]), as follows. By boundary incompressibility of Z, $\partial D' \cap Z$ must cobound a disk $E \subset Z$ with an arc δ of $Z \cap \partial N(L)$. Because $Y - N(L)$ is boundary irreducible by [14, Corollary 3.16], $D' \cup E$ must be parallel to a disk $F \subset \partial N(L)$. Then $D' \cup E \cup F$ is a sphere in $Y - N(L)$ bounding a ball; use the ball to isotope Z, moving the disk E through the ball and slightly past D'. After this isotopy, Z has at least two fewer intersections with truncation edges, and no aditional intersections with other interior edges, so the weight has strictly decreased. Observe also that the arc $\delta \subset \partial Z$ could not have been in meridianal form, so this case will not affect meridianal boundary components of Z, and hence Z remains in meridianal form after isotopy. Then further isotopy as in Theorem 6.2 puts Z back into normal form but with strictly reduced weight, contradicting our assumption that Z had minimal weight. Hence we may assume that an innermost instance of D does not appear as in (e) or (f) of Figure 13.

Finally suppose an innermost D has the form of either (c) or (d) in Figure 13, with one endpoint of α on an interior edge and one on a truncation edge. We will isotope Z to remove the intersection with the interior edge in two steps. The first step is to slide ∂Z through $\partial N(L)$ in $Y - N(L)$. For (c), we slide through a neighborhood of $\beta \cap \partial N(L)$ as shown on the left of Figure 15. For (d), note that because D is innermost, the arc of $Z \cap \partial C$ that meets the truncation face cannot exit the truncation face at the edge vertically adjacent (so the arc cannot be in meridianal form, meaning the corresponding component of ∂Z is not a meridian). So the arc on the truncation face continues to exit through one of the two other edges on the truncation face. One of these is shown in Figure 15(d). In both cases, we slide ∂Z through the truncation face, past the point where it meets the interior edge, as shown on the far right in the figure. The effect of these moves on the component of $Z \cap \partial C$, on the disk D, and on ∂Z in $\partial N(L)$ is shown. Observe that this first step of the isotopy temporarily increases weight by one.

The next step is identical to the process in case (a) in Figure 13: there is now a disk D'' in ∂C with boundary consisting of an arc on an interior edge and an arc of $Z \cap \partial C$. Away from the interior edge, push this disk into C. Use it to isotope Z via a band move as in Figure 14. This will remove all intersections

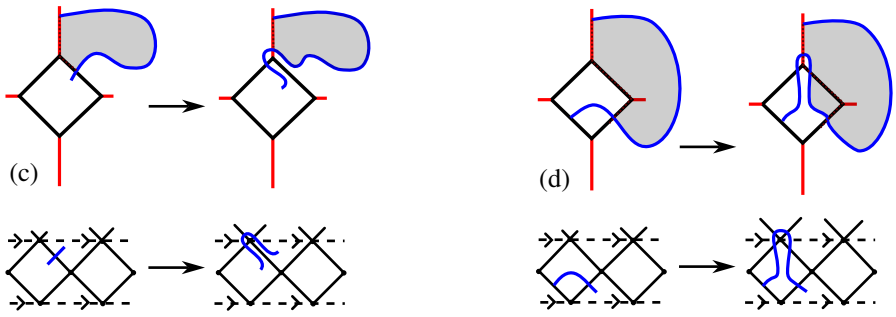


Figure 15: Isotope Z by sliding ∂Z in a neighborhood of $\beta \cap \partial N(L)$. Effect is shown on the component in the chunk on top, and on $\partial N(L)$ on bottom.

of Z with interior edges meeting D''. Thus it removes two intersections, and the weight strictly decreases from its initial value. If the corresponding component of ∂Z was not a meridian in meridianal form, this gives a contradiction to the fact that it was least weight: further isotopy as in Theorem 6.2 will put Z back into normal form without increasing weight.

If the corresponding component of ∂Z is a meridian, and so it began in meridianal form, we may need to isotope further to put it back into meridianal form to apply Theorem 6.2. As noted when we introduced case (d), if this component of ∂Z is a meridian then a disk as in case (d) in Figure 13 cannot be innermost. Thus we are in case (c). Then we follow the proof of Theorem 6.2: the initial isotopy of ∂Z will adjust a curve in meridianal form exactly as in Figure 9, yielding an arc on a truncation face with both endpoints on the same truncation edge. Just as in the proof of Theorem 6.2, further isotopy will put the surface Z back into meridianal form exactly as shown in Figures 11 and 12. This gives a surface that is essential and in meridianal form with smaller weight, and so further isotopy using Theorem 6.2 will give a normal surface with smaller weight; a contradiction. □

8 Angled chunks and combinatorial area

So far, we have isotoped essential surfaces into normal form, and have considered some of the resulting subsurfaces in chunks and their boundary curves. In this section, we recall an important tool to reduce the options for such subsurfaces, namely combinatorial area.

In its most general form, combinatorial area can be defined for surfaces in any chunk with dihedral angles assigned to interior edges that satisfy certain conditions; this is called an *angled chunk decomposition*, and it is described in full generality in [14]. In our setting, label each interior edge with angle $\pi/2$. By [14, Proposition 3.15], the decomposition satisfies the requirements to be an angled chunk decomposition.

We now review a few consequences.

8.1 Combinatorial area

Let Z be in normal form with respect to the chunk decomposition of $Y - N(L)$. Write $Z = \bigcup_{j=1}^{m} Z_j$, where each Z_j is a connected normal surface embedded in a chunk. Consider ∂Z_j. Each component of

∂Z_j meets interior edges a total of $n_S = n_S(Z_j)$ times; this is the number of instances of S on the words decorating ∂Z_j. It meets truncation edges a total of $n_T = n_T(Z_j)$ times, where n_T must be even, since Z_j enters and exits each truncation face by meeting a truncation edge. In case Z is meridianal, $n_T/2$ is the total number of instances of P in the words decorating ∂Z_j. Otherwise, n_T is the total number of instances of B.

The *combinatorial area* of Z_j is defined to be

$$a(Z_j) = \tfrac{1}{2}\pi n_S + \tfrac{1}{2}\pi n_T - 2\pi\chi(Z_j). \tag{8.1}$$

Denote the number of S's in Z_j by $\#S$, and the number of P's by $\#P$. If Z is meridianal, we rewrite (8.1) as

$$a(Z_j) = \tfrac{1}{2}\pi(\#S) + \pi(\#P) - 2\pi\chi(Z_j). \tag{8.2}$$

The *combinatorial area* of Z is defined to be

$$a(Z) = \sum_{j=1}^{m} a(Z_j). \tag{8.3}$$

The combinatorial area satisfies a Gauss–Bonnet formula [14, Proposition 3.12]:

$$a(Z) = -2\pi\chi(Z). \tag{8.4}$$

We also have the following results that follow from [14].

Lemma 8.5 *Let Z_j be a subsurface of a connected normal surface Z. The combinatorial area of Z_j satisfies the following.*

(1) *If $\chi(Z_j) < 0$, then $a(Z_j) \geq 2\pi$.*

(2) *If $\chi(Z_j) \geq 0$, then either $a(Z_j) \geq \pi/2$, or $a(Z_j) = 0$.*

(3) *Additionally, in the case when $a(Z_j) = 0$, Z_j is either*

(a) *an essential torus or Klein bottle embedded in a chunk, hence $Z = Z_j$ is a torus or Klein bottle,*

(b) *an annulus or Möbius band with boundary meeting no edges, or*

(c) *a disk such that ∂Z_j meets exactly four edges of the chunk decomposition.*

Proof Equation (8.1) in the definition of combinatorial area implies the first item.

Note that since Z_j is a connected surface, with or without boundary, and Z_j is never a sphere or projective plane by Proposition 7.1 (2), we have $\chi(Z_j) \leq 1$.

If $\chi(Z_j) = 0$, then Z_j is either an annulus, Möbius band, torus, or Klein bottle. If Z_j is an annulus or Möbius band and ∂Z_j meets an edge, then $a(Z_j) \geq \pi/2$ by (8.1). If Z_j is a torus or Klein bottle, it lies in the interior of the chunk, so meets no edges, and $a(Z_j) = 0$. Similarly if it is an annulus or Möbius band such that ∂Z_j meets no edges, then $a(Z_j) = 0$.

If $\chi(Z_j) = 1$, then Z_j is a disk. Then $a(Z_j) = k\pi/2 - 2\pi$ by formula (8.1), where k is the number of edges (interior or truncation) met by ∂Z_j. Thus if $k > 4$, then $a(Z_j) \geq \pi/2$. If $k = 4$, then $a(Z_j) = 0$ and (3)(c) holds. Finally, the cases $k = 0, 1, 2, 3$ are ruled out by [14, Proposition 3.11]. □

The Gauss–Bonnet formula, (8.4), implies that there are no normal spheres or normal disks embedded in a weakly generalized alternating link complement, for such a surface must have negative combinatorial area, and hence it must have a subsurface of negative combinatorial area, which is impossible by Lemma 8.5.

Howie and Purcell observe that if a weakly generalized alternating link is reducible, it must contain a normal sphere, and if it is boundary reducible, it must contain a normal disk [14, Theorem 3.8].[1] Additionally, that paper studies surfaces with combinatorial area zero, namely annuli and tori, to analyze when weakly generalized alternating knots are hyperbolic [14, Section 4]. By contrast, the work in this paper allows us to study more surfaces: higher-genus surfaces, punctured spheres, etc. For a surface with fixed Euler characteristic, the Gauss–Bonnet formula (8.4) restricts potential normal subsurfaces to those with combinatorial area no more than the original. This paper helps us analyze the structure of potential normal subsurfaces, and how they interact with the diagram. We will see that of key importance are subsurfaces with zero combinatorial area.

Lemma 8.6 *In a weakly generalized alternating link complement, the only normal subsurfaces that are disks with zero combinatorial area have boundaries labeled PP, PSS, SSSS, BBBB, and BBSS.*

Recall that we do not allow both P and B labels in the same word. Thus we will regard *PBB* disks as instances of *BBBB* disks, as in Remark 5.2.

Proof of Lemma 8.6 By Lemma 8.5, a zero area disk meets four edges. If these are all interior edges, the disk is labeled *SSSS*. If it meets truncation edges, it must do so in pairs, thus either two adjacent instances of B or a single P (encoding two truncation edges for a curve in meridianal form). Thus the possibilities are as claimed. □

9 Eliminating *SSSS* disks

In this section, we restrict disks of the form *SSSS*.

Theorem 9.1 *Suppose Z is meridianally incompressible surface in normal form with respect to the chunk decomposition of $Y - N(L)$. Then any SSSS disk is an essential compression disk for Π meeting the diagram $\pi(L)$ exactly four times. Thus if the representativity satisfies $r(\pi(L), \Pi) > 4$, there are no SSSS disks.*

Proof Suppose Z_i is a normal disk meeting exactly four interior edges and suppose Z_i is not a compressing disk for Π. Then Z_i is parallel into the boundary surface Π^+ or Π^- of a chunk, without loss of generality say Π^+, so the curve ∂Z_i bounds a disk on Π^+ (meeting edges, faces, etc. in its interior).

[1] This does not follow from our proof of Theorem 6.2, because we assumed this result to prove Theorem 6.2. However, the proof is similar, using surgery on disks rather than isotopy as in [5]. Hence a corollary is that weakly generalized alternating link complements are irreducible and boundary irreducible [14, Corollary 3.16].

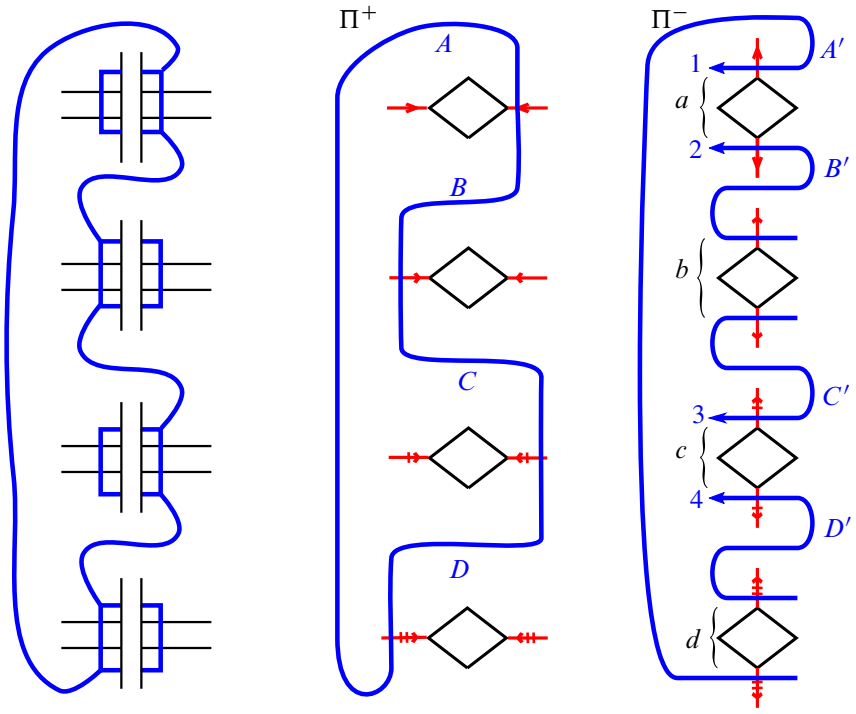


Figure 16: Left: disk with boundary *SSSS* in $Y - N(L)$. Middle: form in chunk decomposition with boundary Π^+. Right: glued to Π^- as shown.

The form of Z_i in the diagram $\pi(L)$ is shown on the left of Figure 16, where the curve $\partial Z_i \cap \Pi$ is shown in blue along with four saddles. In particular, if one follows the component of $\partial Z_i \cap \Pi$, the overpasses of $\pi(L)$ alternate between being on the right and on the left due to the diagram $\pi(L)$ being alternating. The form of ∂Z_i in one chunk is shown in the middle, where ∂Z_i lies on some component Π_j^+ of Π^+. Denote the four arcs of ∂Z_i between the intersection points with edges by A, B, C, D as on the middle figure. The surface Π_j^+ is glued to Π_j^- via a gluing that rotates the boundary of each face as in Section 3.2. Thus the arcs A, B, C, D are glued to four arcs A', B', C', D' as shown on the right. The blue arcs in the right figure are arcs of boundary curves of other normal subsurfaces Z_k in the chunk decomposition; thus they are portions of simultaneously embedded closed curves.

First, observe that while A', B', C', and D' no longer necessarily connect to bound a disk in Π^-, the union of the arcs A', B', C', and D' along with meridianal arcs through the four truncation faces shown on the right of Figure 16, and pieces of interior edges between them, all bound a disk on Π_j^-.

Label the blue arcs that run into this disk in Π^- by 1, 2, 3, 4 as in Figure 16, right. Because the arcs are subsets of a disjoint union of embedded closed curves (the boundaries of normal surfaces in the chunk), the arcs 1, 2, 3, and 4 must exit the disk region shown. They might exit either by connecting to each other, for example 1 might connect to 2, or they might exit by running through the region between 1 and 2, between arcs B' and C', between 3 and 4, or between arcs D' and A'. We call these regions zones a, b, c, and d, and they are shown in Figure 16, right, as well.

The arcs labeled 1 and 2 cannot connect to each other or exit through zone a by Proposition 7.2, since they each meet the interior edge in that region once already. Similarly, the arcs 3 and 4 cannot connect with each other or exit through zone c. Note also that arcs 1 and 4 cannot run to zone d by the same result, since arc A' already meets the edge of this region, and similarly for D'. Similarly arcs 2 and 3 cannot run through zone b.

It follows that the arc labeled 1 runs into zone b or c. If b, then the arc labeled 2 must run into zone a or b in order to remain disjoint from the arc labeled 1. But this is impossible: these are exactly the zones that 2 is not allowed to enter. Hence the arc labeled 1 runs to zone c. But then the arc labeled 4 must run into zone c or d in order to remain disjoint from 1. Again this is impossible: these are exactly the zones that 4 is not allowed to enter. It follows that there can be no *SSSS* disk that is not a compressing disk for Π. □

10 Eliminating disks with words in the letter *P*

In this section, we prove that if Z has meridianal boundary, there are no disks whose boundaries are curves of intersection labeled with two or three letters in P and S. That is, there are no words of type *SS*, *PP*, *PS*, *PPP*, *PSS*, *SSS*, or *PPS*. The proof for *SS* and *SSS* holds in more generality, and does not require meridianal boundary. We also prove that for any subsurface Z_i of Z, there are no boundary components ∂Z_i labeled with an odd number of instances of P and S. Finally, we use this to show all weakly generalized alternating links are prime.

Theorem 10.1 *Suppose Z is in normal form with respect to the chunk decomposition of $Y - N(L)$. Then no normal subsurface of Z is a disk whose boundary is labelled SS, nor a disk in meridianal form whose boundary is labelled PP or PS.*

Proof Recall that the boundary of a chunk is decorated by the diagram graph, where interior edges corresond to diagram edges, and truncation faces correspond to neighborhoods of crossings. Therefore the boundary of a disk labeled by two letters *PP*, *PS*, or *SS* gives a curve on Π meeting the diagram $\pi(L)$ at edges or at crossings.

In the case of a label P, there is a corresponding arc through a truncation face in meridianal form. We isotope this curve very slightly off the crossing in a direction determined by the side of the truncation face that is met by the curve. See Figure 17, where an example is shown of a boundary *PP*. The isotopy moves ∂Z_i slightly to the curve γ, which we think of as lying on Π and meeting two edges of the diagram graph $\pi(L)$.

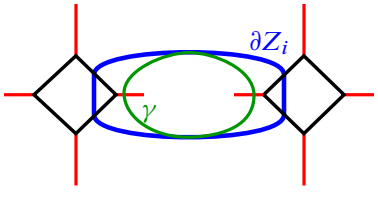


Figure 17: If ∂Z_i is labeled *PP*, it determines a curve γ meeting the diagram exactly twice.

Then in all cases, the disk Z_i determines a curve γ on Π meeting the diagram exactly twice transversely in diagram edges. Because Z_i is a disk, γ must also bound a disk. Because the diagram has representativity at least 4, by definition of weakly generalized alternating, the disk must be parallel to Π. Then the fact that the diagram is weakly prime implies that γ bounds a disk that does not meet any crossings.

In the case that the curve is of the form *SS* or *PS*, this gives an immediate contradiction: the curve ∂Z_i is not in normal form, violating condition (3) or (4) of Definition 4.2, respectively.

In the case the curve is of the form *PP*, then it has the form shown in Figure 17. Both arcs are in meridianal form, thus they cut off a corner that is not identified to another corner of the truncation face. However, consider the interior edge encircled by the curve ∂Z_i. Because this is an interior edge of a chunk decomposition of a weakly generalized alternating link, at one of its endpoints, this edge is identified to another edge across the corresponding truncation face, as described in Section 3.2; see also Figure 3. Thus one of its endpoints meets a corner of a truncation face that is identified to another corner of the same truncation face. This is a contradiction: because ∂Z_i is in meridianal form, neither endpoint can have this property. □

Note that the previous proof relies heavily on the alternating condition, when using the chunk decomposition of Section 3.2.

Theorem 10.2 *Suppose Z is in normal form with respect to the chunk decomposition of $Y - N(L)$. Then no normal subsurface of Z is a disk whose boundary is labelled SSS, nor a disk in meridianal form whose boundary is labelled PPP, PPS, or PSS.*

More generally, no normal subsurface Z_i of Z has a boundary component meeting an odd number of letters S and P.

Proof As in the proof of Theorem 10.1, a curve ∂Z_i determines an embedded curve γ on the diagram graph $\pi(L) \subset \Pi$, with letters S running transversely through diagram edges, and letters P running through a diagram edge immediately to the left or right of a crossing, determined by the position of the arc on the truncation face in meridianal form as in Figure 17.

In each case, consider the checkerboard coloring of the diagram. Passing through a letter P or S changes the color of the face meeting ∂Z_i. If the word labeling ∂Z_i is made up of exactly three letters, then the color must change exactly three times. But this is impossible: if we start in a white face and traverse ∂Z_i, it changes to shaded when it meets the first letter, to white when it meets the second letter, to shaded when it meets the third, and then it must close up, implying that the starting and ending face is both shaded and white. This contradiction is illustrated in the case *PSS* in Figure 18.

More generally, it is impossible for a curve of ∂Z_i to meet an odd number of letters S and P, since again the existence of such a curve would contradict the checkerboard coloring of the diagram. □

The above theorem gives us a quick way to prove the fact that a weakly generalized alternating link with a cellular diagram is prime. When the hat-representativity satisfies $\hat{r}(\pi(L), \Pi) > 4$, this result was originally proved in Howie–Purcell [14, Corollary 4.7]. We can extend now to all weakly generalized alternating links without the additional hat-representativity condition.

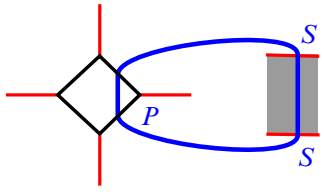


Figure 18: A *PSS* disk should have three arcs in three faces of distinct colors, but this is impossible for a checkerboard colored diagram.

Theorem 10.3 *A weakly generalized alternating link is prime.*

Proof Suppose not. Then there exists an essential meridianal annulus Z. Put it into normal form with respect to the chunk decomposition. We may assume it decomposes into normal subsurfaces that are meridianal by Theorem 6.2, and have zero combinatorial area by the Gauss–Bonnet formula, (8.4). Because it meets two meridians of the link, at least one of the subsurfaces Z_i must have a boundary curve ∂Z_i whose labeling includes at least one instance of P. Because both boundary components are meridianal, there will be no instances of B. By Lemma 8.5, Z_i must be a disk meeting exactly four edges of the chunk decomposition. By Lemma 8.6, the possibilities are disks with boundaries labeled *PP* or *PSS*. Disks labeled *PP* are ruled out by Theorem 10.1. Disks labeled *PSS* are ruled out by Theorem 10.2. This gives a contradiction. □

11 Disks with *BBBB* and *BBSS* words

This section concerns disks whose boundaries are *BBBB* or *BBSS* words. As mentioned in Remark 5.2, an instance of P may be replaced with two instances of B, and the results for *BB* go through. Hence the results for *BBBB* disks immediately apply to *PBB* disks, and we will use this in the sequel.

For a surface Z with nonmeridianal boundary, consider a normal disk Z_i that meets exactly four truncation edges, and no interior edges. Such a subsurface is a disk that corresponds to a *BBBB* word. For an example, see Figure 19, left. Similarly, a normal disk that meets exactly two truncation edges and exactly two interior edges corresponds to a *BBSS* word. An example is shown in Figure 19, right.

As described in [8], topologically the label B means that ∂Z_i meets $\partial N(L)$. An arc *BB* can be a part of ∂Z, where ∂Z is on $\partial N(L)$. In a chunk decomposition, such an arc travels between two truncation edges in a truncation face. An arc *BB* can also connect two *BB* arcs of the previous type, and then it lies in an interior face of a chunk. Each *BBBB* disk is hence a quadrilateral with two opposite sides on truncation faces and the other pair of opposite sides on interior faces. Since a *BB*-arc that lies in a truncation face

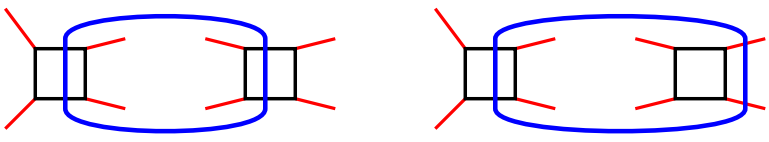


Figure 19: Left: an example of a normal disk of type *BBBB*. Right: an example of type *BBSS*.

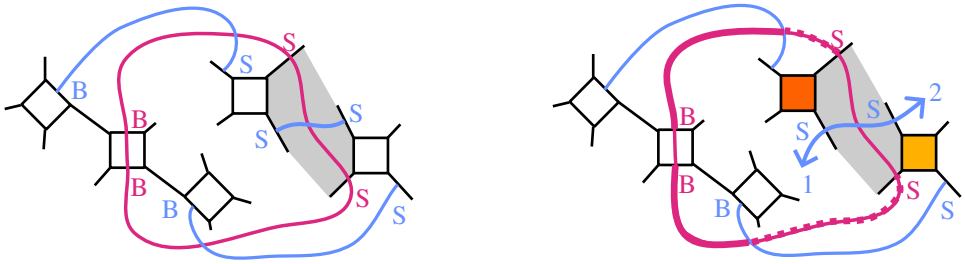


Figure 20: Left: ∂D is shown in red in Π^+. Arcs of subsurfaces meeting Π^- that are glued to ∂D are superimposed on Π^+, and shown in light blue. Right: if the blue *SS* arc connects to form the boundary of a *BBSS* disk, it cannot meet thick or dashed red arcs.

is a part of ∂Z, it is not glued to any other arc of ∂Z_j, for any j. On the other hand, a *BB*-arc that lies inside an interior face, as well as a *BS*-arc, is not a part of ∂Z, and is rather in the interior of Z. Such an arc therefore must be glued to a similar arc of some Z_k.

We say that two *BBBB* disks are *connected* if they share a common arc on an interior face. A collection of such disks is *connected* if the union of disks is connected in Y. A collection of connected disks is *maximal* if it is not a strict subset of a connected collection. We define connected and maximal collections of *BBSS* disks similarly.

Lemma 11.1 *Suppose D is a BBSS disk that is parallel to the surface Π. Then D is not connected to another BBSS disk along its SS arc.*

Proof A *BBSS* disk meets a truncation face along the arc *BB*, two opposite faces adjacent to that truncation face of the same color — say white — along the two arcs *BS*, and a face containing the *SS* arc that is shaded. Without loss of generality, say the *BBSS* disk parallel to Π lies on Π^+. Then the three arcs of the disk that lie in interior faces are glued to arcs in Π^-. As in Section 3.2, the gluing of chunks in the neighborhood of edges is via a rotation of the respective faces: in the clockwise direction for the two arcs *BS*, and in a counterclockwise direction for the arc *SS*. Superimpose all these arcs on Π as in Figure 20, left. The boundary of the disk *BBSS* is shown in red, and the arcs in Π^- that it is glued to are shown in light blue.

By way of contradiction, suppose the *SS* arc glues to an *SS* arc contained in another *BBSS* disk E, with boundary on Π^-. By hypothesis, ∂E is labelled with only two instances of S, hence ∂E cannot contain either of the two blue arcs BS in white faces obtained by rotations of arcs of D. Thus the disk E must be disjoint from these arcs. Also note that (the superimposed copy of) ∂E intersects ∂D. Because D is a disk parallel into Π, ∂E must meet ∂D one additional time.

Note that the arc of ∂E labelled 1 cannot run immediately into the dark orange truncation face; this would contradict the fact that E is normal. Similarly the arc labelled 2 cannot run immediately into the light orange truncation face, or E is not normal. This means ∂E does not meet ∂D on the dashed arcs shown on the right of Figure 20.

Suppose the arc labelled 1 runs to meet one of the thicker red arcs shown. Then there are two cases to consider.

Case 1. Suppose arc 1 intersects the upper left red arc. Then arc 1 cannot intersect the blue arc, so it must run into a truncation face or interior edge after intersecting the thick red arc. Because arc 1 has already met two instances of S, it cannot meet an interior edge and remain the boundary of a *BBSS* disk. Since meeting the truncation face uses up all the letters *BBSS* in its boundary, this means that the arc labelled 1 meets no edges between 1 and the thick red arc. It follows that the two white faces on opposite sides of the dark orange truncation face are the same, since one white face contains arc 1 and the other white face contains the thick red arc. But this would contradict the link being weakly prime.

Case 2. Suppose arc 1 meets the thick red arc in the lower left part of the figure. Similarly to case 1, arc 1 then must enter the region between that arc and the blue arc through a truncation face, using up all the letters *BBSS*. But then arc 1 must connect immediately to the arc labelled 2, without meeting additional interior or truncation edges. This means that the white face containing the arc labelled 2 and the white face containing the thick red arc in the lower left part of the figure are the same. But these are the two white faces adjacent to the light orange truncation face, contradicting the link being weakly prime.

Therefore, arc 1 does not meet any of the two thicker red arcs. The only remaining place that the arc labelled 1 could meet the boundary of the red disk would be on the truncation face between instances of B, and it must do so by entering and exiting opposite sides of that truncation face, not the sides meeting the red arc. But those opposite sides lie in shaded faces, whereas arc 1 is in a white face. This is also impossible. □

Lemma 11.2 *Assume that the hat-representativity $\hat{r}(\pi(L), \Pi) > 4$, and that $\pi(L)$ is not a string of bigons on Π. Let Z be in normal position, with a nonmeridianal boundary component. Then the following form disks*:

(1) *a maximal connected collection of BBBB disks for Z or*

(2) *a maximal connected collection of BBSS disks for Z.*

Proof As explained above, the *BBBB* or *BBSS* disks are connected only along arcs on interior faces, since the arcs on truncation faces are a part of the surface boundary.

For (1), a maximal connected collection is therefore glued end to end along opposite interior faces. The only way it could not form a disk is if it forms an essential annulus or Möbius band, made entirely of *BBBB* disks. In the case of an annulus, it was shown in [14, Theorem 4.6] that $\pi(L)$ is a string of bigons on Π (here we use the hypothesis $\hat{r}(\pi(L), \Pi) > 4$). In fact the proof also implies the same result for the case of the Möbius band, because it shows that a string of *BBBB* disks connect to bound a chain of bigons in the diagram graph.

For (2), arcs in the interior faces have the form *BS* or *SS*. By Lemma 11.1, *SS* arcs cannot glue to other *SS* arcs in the maximal collection. All *BBSS* disks in the collection must be glued along arcs of the form *BS*. If the connected collection of *BBSS* disks is not a disk, it will form a normal annulus made up of *BBSS* disks, at least one of which is parallel into Π because $\hat{r}(\pi(L), \Pi) > 4$. But then by [14, Lemma 4.5], the diagram $\pi(L)$ is a string of bigons on Π, contradicting our hypotheses. □

Proposition 11.3 *For a surface Z in normal position, no BBBB region is connected to a BBSS region.*

Proof Two normal subsurfaces Z_i, Z_j in a chunk can only be connected across arcs in interior faces, not across arcs in truncation faces, which are left unglued and become the boundary of the normal surface. The arcs in the interior faces for a *BBBB* region have both endpoints on a truncation edge. The arcs in the interior faces for a *BBSS* region have one end on a truncation edge and one end on an interior edge, or both ends on interior edges. Because interior edges glue to interior edges, each arc of a *BBSS* region must glue to an arc with an endpoint on an interior edge. Thus it cannot glue to a *BBBB* region. □

Theorem 11.4 *Assume that the hat-representativity $\hat{r}(\pi(L), \Pi) > 4$, and $\pi(L)$ is not a string of bigons on Π. Let Z be in normal form. Then all subsurfaces Z_i that are neither BBBB disks nor BBSS disks, together with the link L, determine the surface Z up to isotopy.*

Proof By Lemma 11.2 (1), maximal connected collections of *BBBB* disks form disjoint disks. Similarly for *BBSS* disks by Lemma 11.2 (2). By Proposition 11.3, no such disks are connected. Therefore, if we remove maximal connected collections of *BBBB* and *BBSS* regions, we are left with a surface with a number of disjoint disks removed.

Suppose Z' is another surface in normal form, whose normal subsurfaces agree with those of Z away from *BBBB* and *BBSS* disks. Then Z and Z' are both obtained by adding disks to the same surface with boundary, along the same boundary components. The link L has irreducible complement by [14, Theorem 3.14]. Therefore, the two surfaces disagree except in disks that can be isotoped to bound balls. Thus the surfaces are isotopic. □

12 Essential Conway spheres

In this section, we apply our results above to generalize the work of Menasco [20] that allows the characterization of essential 4-punctured spheres, or essential Conway spheres, for alternating links as "visible" and "hidden", with terminology due to Thistlethwaite [27]. We call an essential Conway sphere *visible* if its intersection with $\Pi^{\pm}$ consists of only one *PPPP* curve, and *hidden* if it consists of exactly two *PSPS* curves. See Figure 21 and compare to [27, Figure 3(i)–(ii)].

Note that if curves as in Figure 21 appear in the classical alternating case, the region exterior to any dashed line will also be a disk, hence will contain an alternating tangle. In the weakly generalized alternating setting, this region may have higher genus.

In this section, we will restrict to diagrams that are *cellular*, meaning all regions of the diagram are disks. This is the only place in the paper where we use cellular.

Theorem 12.1 *Let L be a weakly generalized alternating link on a connected projection surface Π with a cellular diagram and hat-representativity $\hat{r}(\pi(L), \Pi) > 4$. Then any essential Conway sphere has one of two forms: visible or hidden.*

We will prove the theorem by considering how essential Conway spheres decompose in a chunk decomposition, ruling out certain subsurfaces and labels on their boundaries. The proof will follow from a few lemmas.

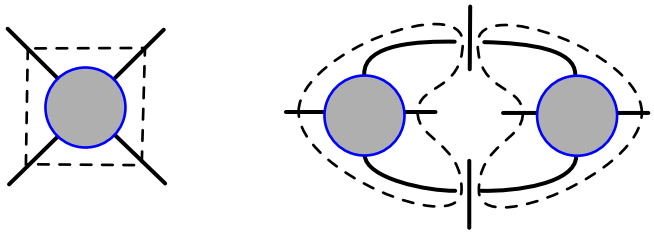


Figure 21: Left: a visible essential Conway sphere is made up of two disks labeled *PPPP*, one on either side of the projection surface. Right: a hidden Conway sphere is made up of four disks labeled *PSPS*, two on either side of the projection surface. Here we show how they meet one side of the diagram. The gray disk denotes a portion of the diagram contained in a disk in the projection surface.

Remark 12.2 We use the hypothesis that the diagram is cellular in Theorem 12.1 to simplify the enumeration of potential essential Conway spheres. It is required in the proof we give below. However, we have no counterexamples to the theorem when this hypothesis is dropped. It is an interesting open question as to whether a cellular diagram is necessary for Theorem 12.1 to hold.

Lemma 12.3 *Under the hypotheses of Theorem 12.1, suppose $Z = \bigcup Z_i$ is an essential Conway sphere in normal form with respect to the chunk decomposition. Then all the subsurfaces Z_i must be disks.*

Proof An essential Conway sphere is an essential 4-punctured sphere, hence it will have combinatorial area $a(Z) = -2\pi\chi(S) = 4\pi$. Because it meets the knot in meridianal punctures, by Assumption 6.3 we assume it is in normal form with all components of ∂Z in meridianal form. It decomposes into normal subsurfaces within chunks that are genus-zero surfaces. Thus the only subsurfaces that may arise are disks, annuli, and 3-holed spheres. Also, on each of $\Pi^{\pm}$ it must meet four instances of the letter P, so eight instances of P total.

In a cellular link diagram, no boundary component of a genus-zero subsurface in normal position can lie entirely in a face of the diagram. This means that each boundary component of an annulus or 3-holed sphere must meet instances of S or P, and by Theorem 10.2, it must meet at least two such instances. We will prove any annulus or 3-holed sphere subsurface contributes too much area, to rule these out.

By (8.2), the smallest possible area contribution from an annulus has area 2π, meeting two instances of S on each boundary component. But we still need eight instances of the letter P, and the total area of all subsurfaces must add up to 4π. By Lemma 8.5, all additional contributions to area are nonnegative. Since each P contributes area π by (8.2), annuli have too much area. Similarly for 3-holed spheres. Thus all subsurfaces must be disks. □

Lemma 12.4 *Under the hypotheses of Theorem 12.1, suppose $Z = \bigcup Z_i$ is an essential Conway sphere in normal form with respect to the chunk decomposition. Then each of the disks Z_i must be labeled SSSS, PSSS, PPSS, PSPS, PPPS, or PPPP.*

Proof When Z is an essential Conway sphere, by Lemma 12.3, all Z_i are disks. Also, Z must meet eight instances of P, with four instances each on $\Pi^{\pm}$. The total area is $a(Z) = 4\pi$. Thus there is some Z_i with positive area at most 2π.

Consider the possible boundary words for disks with area at most 2π. Lemma 8.6, along with the results of Section 10, rules out all zero area disks *PP* and *PSS*. Any odd number of letters is ruled out by Theorem 10.2. If the representativity of the diagram is 4, then 0-area *SSSS* disks are possible, not ruled out by Theorem 9.1, although they will be ruled out for higher representativity. The only disk contributing area $\pi/2$ has boundary *PSSS*. The possibilities contributing area π are *PPSS*, *PSPS*, and *SSSSSS*. Those contributing area $3\pi/2$ are *PPPS* and *PSSSSS*. Those contributing area 2π include *PPPP*, as well as other words with at most three instances of P and some instances of S.

Each instance of P in the disks enumerated above contributes at least $\pi/2$ to the sum. Returning to the case of an essential Conway sphere, using the fact that there must be eight instances of P in total, in fact each of the disks with positive area must contain a letter P, and the area of the disk will be $\pi/2$ times the number of instances of P. Because there must be four instances of P on each side of Π, the only possibilities for disks overall are 0-area *SSSS* disks, and positive area *PSSS*, *PPSS*, *PSPS*, *PPPS*, and *PPPP* disks. □

By the assumption that $\hat{r}(\pi(L), \Pi) > 4$ and the assumption that Π is connected, the boundaries of essential compression disks on one side of Π meet the diagram in more than four points. So on this side of Π, any *PSSS*, *PPSS*, or *SSSS* curve bounds a disk parallel to the projection surface. Without loss of generality, say this side is Π^+. We note that unlike in the classical case of alternating links in S^3, here the proofs for Π^+ and Π^- from this moment are not analogous.

We say that a *PPSS* or *PSSS* disk is *innermost* on Π^+ if the boundary curve bounds a disk on the surface Π^+ that contains no other intersections with Z. We say it is *outermost* if it bounds a disk D on Π^+ such that $\Pi^+ - D$ is a surface with boundary containing no intersections with Z.

Lemma 12.5 *Under the hypotheses of Theorem 12.1, if there exists a PPSS disk or a PSSS disk with boundary on Π^+, then there exists one that is innermost or one that is outermost with boundary on Π^+.*

Proof Suppose first that there is a *PPSS* disk with boundary on Π^+. Because any disk with boundary meeting Π^+ at most four times is parallel to Π^+ (by the assumption on $\hat{r}$), we may assume the *PPSS* curve bounds a disk in Π^+. Since there are four instances of P on each side of Π, there will either be another *PPSS* disk, or a *PSPS* disk, or two *PSSS* disks with boundary on Π^+ by Lemma 12.4. In the first two cases, the second disk lies on one side of the *PPSS* disk and so there are no curves ∂Z_i on the other side. In the last case, there may be a *PSSS* disk on both sides, but then such a *PSSS* disk is innermost. This proves the lemma when there is a *PPSS* disk.

So now suppose there is a *PSSS* disk with boundary on Π^+. Again, its boundary curve encloses a disk on Π^+. Then within the same chunk there could be a single *PPPS* disk, which means the *PSSS* disk is innermost or outermost. There could be another *PSSS* disk and a disk with two instances of P (namely *PSPS* or *PPSS*), in which case one of the two *PSSS* disks will be innermost or outermost. Or there could be three additional *PSSS* disks, in which case one is innermost or outermost. □

The next lemma can be viewed as an extension of a lemma of Menasco [20, Lemma 2] in our setting.

Lemma 12.6 *Under the hypotheses of Theorem 12.1, suppose $Z = \bigcup Z_i$ is an essential Conway sphere in normal form with respect to the chunk decomposition. Then there are no PSSS, PPSS, PPPS, or SSSS disks.*

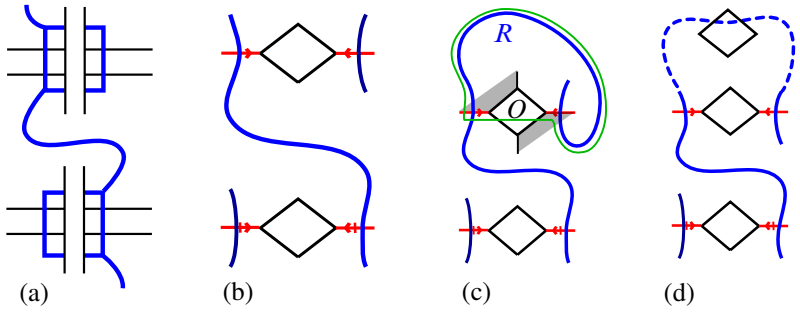


Figure 22: The *SS* portion of a *PSSS* or *PPSS* disk, shown in (a) with saddles in the link complement, and in (b) on the chunk decomposition. In (c) and (d), in fact, it must be a *PSSS* disk, either with P away from the arc between identified edges as in (c), or with P meeting this arc as in (d).

Proof We will first rule out *PPSS* and *PSSS* disks with their boundaries in Π^+. Suppose there is such a disk. Then Lemma 12.5 implies that there is either an innermost or outermost such disk.

The portion of the boundary in the *PPSS* or *PSSS* disk that runs between two instances of S must lie in a single region of the link diagram; here we are using the fact that the edges and ideal vertices on Π^+ come from $\pi(L)$ as in Section 3.1. Because the diagram is alternating, it must meet the first instance of S in a saddle with the link on one side, and in the second instance the link lies on the other side; see Figure 22(a) and (b).

Each instance of S lies on an interior edge of a chunk that is glued to another interior edge. Hence there must be at least one other curve of ∂Z_i running through that glued edge, as shown in Figure 22(b). But this gives two curves lying on opposite sides of the *PSSS* or *PPSS* curve that we consider. Because our disk is innermost or outermost, this is possible only if one of these two curves is actually still part of our original *PPSS* or *PSSS* boundary. If the arc of the *PSSS* disk between the two identified interior edges cuts off a disk on the projection surface Π with those edges, then we have a contradiction to Proposition 7.2. This was Menasco's argument in [20]. In our setting, it may not be the case that this arc cuts of a disk with the edge, and so Proposition 7.2 may not apply.

So assume that the arc does not cut off a disk on Π. So far we have found three instances of S. It follows that the original disk is a *PSSS* disk rather than a *PPSS* disk.

Suppose first that the instance of P does not lie between the two instances of S on identified edges. Hence there is an arc, say R, on the boundary of the *PSSS* disk meeting a single (disk) face of the chunk decomposition, with endpoints on interior edges that are identified to a crossing arc under the gluing. Denote the crossing of $\pi(L)$ that correpsonds to this crossing arc by O. By assumption, the disk on Π bounded by the *PSSS* curve does not contain the crossing O, or else the arc R would cut off a disk on Π that contradicts Proposition 7.2. Because of checkerboard coloring, the endpoints of the arc R must lie on opposite sides of the crossing O, as shown in Figure 22(c). Then the union of this arc R and another arc running between the endpoints of R through the crossing O forms a closed curve on the surface Π that meets the arc R only once (after slight isotopy). This curve is shown by the thin green line in Figure 22(c). This is impossible: either this curve meets the boundary of the *PSSS* disk only once on Π, which is

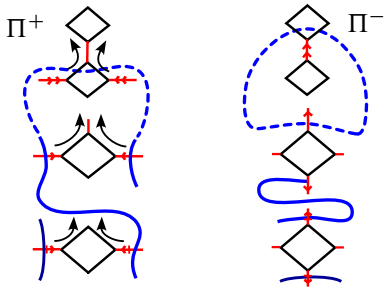


Figure 23: In case (d) of Figure 22, the arcs of the *PSSS* disk in Π^+ shown on the left must glue to the arcs in Π^- shown on the right.

impossible for a disk boundary, or it meets it again as the *PSSS* disk boundary exits through one of the identified edges at the crossing O. But in the latter case, the disk bounded by the *PSSS* curve in Π would then again give a contradiction to Proposition 7.2.

So the instance of P lies between these two instances of S; see Figure 22(d). This figure shows Π^+. We now consider how these arcs are glued to Π^-. Arcs in Π^+ and in Π^- in this case are shown in Figure 23. Because the surface Z is in meridianal form, the instance of P on Π^+ is glued across truncation edges to an instance of P in Π^-; see Figure 7. In particular, the arc in Π^+ shown in dashed lines in Figure 22(d), running from S to P and back to an identified S, is mapped to a single arc in Π^- running from some S to P and back to some S. However, on Π^-, the two endpoints of that arc must be identified together; this is due to the fact that the gluing of chunks rotates the boundary of each face in the clockwise direction in white faces, and the counterclockwise direction in shaded faces. Thus this arc in Π^+ is identified to a closed curve in Π^-. It must be the boundary of a disk Z_i by Lemma 12.3. Then this is a *PS* disk meeting Π^-. But by Theorem 10.1, there are no *PS* disks, and we have a contradiction. It follows that there are no *PPSS* or *PSSS* disks with boundaries in Π^+.

Now consider *SSSS* disks. There is no such disk on Π^+, by Theorem 9.1 and the fact that any compressing disk meets Π^+ more than four times. Observe that we have now ruled out *SSSS*, *PPSS*, and *PSSS* disks on Π^+. By Lemma 12.4, there are no further options for disks with two adjacent instances of S.

Since the portion of ∂Z_i running between two instances of S must be glued to another boundary curve running between two instances of S, there cannot be a disk with two adjacent instances of S on the other side of Π as well. This rules out *SSSS*, *PSSS*, and *PPSS* disks on Π^-.

Finally, since there are no *PSSS* disks, there can be no *PPPS* disks, because there must be exactly four instances of P on each side of the projection surface, and there are no options for disks meeting only one instance of P to pair with *PPPS*. □

Proof of Theorem 12.1 Suppose Z is an essential Conway sphere. By Lemma 12.3, it meets the diagram in normal disks. By Lemmas 12.4 and 12.6, the only possible disks are labeled *PPPP* or *PSPS*.

In the case of a *PPPP* disk, the fact that there must be four instances of P on each side of Π implies that there are exactly two *PPPP* disks, lying on opposite sides of the projection surface. They cut off a visible essential Conway sphere.

In the case of a *PSPS* disk, there must be four such disks, with one *PSPS* disk meeting another across each saddle S on the same side of the projection surface. It must additionally meet another two disks at the saddle S on the opposite side of the projection surface. The only possibility is that the four disks fit together with two on one side as in Figure 21, right. This is a hidden essential Conway sphere. □

Acknowledgements

We thank the anonymous referees whose suggestions have greatly improved the paper. Purcell was partially supported by the Australian Research Council, grants DP160103085 and DP210103136. Tsvietkova was partially supported by the National Science Foundation (NSF) of the United States, grants DMS-2142487 (CAREER), DMS-1664425 (previously 1406588) and DMS-2005496, the Institute of Advanced Study under NSF grant DMS-1926686, and the Okinawa Institute of Science and Technology.

References

[1] **C C Adams**, *Toroidally alternating knots and links*, Topology 33:2 (1994) 353–369 MR

[2] **C Adams**, **C Albors-Riera**, **B Haddock**, **Z Li**, **D Nishida**, **B Reinoso**, **L Wang**, *Hyperbolicity of links in thickened surfaces*, Topology Appl. 256 (2019) 262–278 MR

[3] **A Champanerkar**, **I Kofman**, **J S Purcell**, *Geometry of biperiodic alternating links*, J. Lond. Math. Soc. (2) 99:3 (2019) 807–830 MR

[4] **A Champanerkar**, **I Kofman**, **J S Purcell**, *Right-angled polyhedra and alternating links*, Algebr. Geom. Topol. 22:2 (2022) 739–784 MR

[5] **D Futer**, **F Guéritaud**, *Angled decompositions of arborescent link complements*, Proc. Lond. Math. Soc. (3) 98:2 (2009) 325–364 MR

[6] **W Haken**, *Theorie der Normalflächen*, Acta Math. 105 (1961) 245–375 MR

[7] **J Hass**, **A Thompson**, **A Tsvietkova**, *The number of surfaces of fixed genus in an alternating link complement*, Int. Math. Res. Not. 2017:6 (2017) 1611–1622 MR

[8] **J Hass**, **A Thompson**, **A Tsvietkova**, *Alternating links have at most polynomially many Seifert surfaces of fixed genus*, Indiana Univ. Math. J. 70:2 (2021) 525–534 MR

[9] **J Hass**, **A Thompson**, **A Tsvietkova**, *Tangle decompositions of alternating link complements*, Illinois J. Math. 65:3 (2021) 533–545 MR

[10] **A Hatcher**, *Notes on basic* 3*-manifold topology*, notes (2023) https://pi.math.cornell.edu/∼hatcher/3M/3Mdownloads.html

[11] **C Hayashi**, *Links with alternating diagrams on closed surfaces of positive genus*, Math. Proc. Cambridge Philos. Soc. 117:1 (1995) 113–128 MR

[12] **J A Howie**, *Surface-alternating knots and links*, PhD thesis, University of Melbourne (2015) http://hdl.handle.net/11343/58991

[13] **J A Howie**, *A characterisation of alternating knot exteriors*, Geom. Topol. 21:4 (2017) 2353–2371 MR

[14] **J A Howie**, **J S Purcell**, *Geometry of alternating links on surfaces*, Trans. Amer. Math. Soc. 373:4 (2020) 2349–2397 MR

[15] **T Kindred**, *Primeness of alternating virtual links* (2022) arXiv 2210.03225

[16] **M Lackenby**, *Word hyperbolic Dehn surgery*, Invent. Math. 140:2 (2000) 243–282 MR

[17] **W Lin**, *A meridian lemma for fully alternating links in thickened surfaces* (2022) arXiv 2203.06858

[18] **M T Lozano**, **J H Przytycki**, *Incompressible surfaces in the exterior of a closed* 3*-braid, I: Surfaces with horizontal boundary components*, Math. Proc. Cambridge Philos. Soc. 98:2 (1985) 275–299 MR

[19] **W W Menasco**, *Polyhedra representation of link complements*, from "Low-dimensional topology" (San Francisco, CA, 1981) (J Lomonaco, Samuel J, editor), Contemp. Math. 20, Amer. Math. Soc., Providence, RI (1983) 305–325 MR

[20] **W Menasco**, *Closed incompressible surfaces in alternating knot and link complements*, Topology 23:1 (1984) 37–44 MR

[21] **W Menasco**, *Determining incompressibility of surfaces in alternating knot and link complements*, Pacific J. Math. 117:2 (1985) 353–370 MR

[22] **W W Menasco**, **M B Thistlethwaite**, *Surfaces with boundary in alternating knot exteriors*, J. Reine Angew. Math. 426 (1992) 47–65 MR

[23] **M Ozawa**, *Non-triviality of generalized alternating knots*, J. Knot Theory Ramifications 15:3 (2006) 351–360 MR

[24] **M Ozawa**, *Rational structure on algebraic tangles and closed incompressible surfaces in the complements of algebraically alternating knots and links*, Topology Appl. 157:12 (2010) 1937–1948 MR

[25] **J S Purcell**, *Hyperbolic knot theory*, Graduate Studies in Mathematics 209, Amer. Math. Soc., Providence, RI (2020) MR

[26] **J S Purcell**, **A Tsvietkova**, *Polynomial bounds for surfaces in cusped* 3*-manifolds* (2023) arXiv 2311.08567

[27] **M B Thistlethwaite**, *On the algebraic part of an alternating link*, Pacific J. Math. 151:2 (1991) 317–333 MR

[28] **W P Thurston**, *The geometry and topology of three-manifolds*, lecture notes, Princeton University (1979) https://url.msp.org/gt3m

[29] **F Waldhausen**, *On irreducible* 3*-manifolds which are sufficiently large*, Ann. of Math. (2) 87 (1968) 56–88 MR

JESSICA S. PURCELL jessica.purcell@monash.edu
School of Mathematics, Monash University, Clayton, VIC, Australia

ANASTASIIA TSVIETKOVA n.tsvet@gmail.com
Department of Mathematics and Computer Science, Rutgers University, Newark, Newark, NJ, United States

Received: May 26, 2022 Revised: November 8, 2024

msp

Algebraic & Geometric Topology **26**:3 (2026) 863–954
DOI: 10.2140/agt.2026.26.863
Published: April 1, 2026

Geometric rigidity of quasi-isometries in horospherical products

TOM FERRAGUT

We prove that quasi-isometries of horospherical products of hyperbolic spaces are geometrically rigid in the sense that they are uniformly close to product maps. This is a generalisation of a result obtained by Eskin, Fisher and Whyte (2012). Our work covers the case of solvable Lie groups of the form $\mathbb{R} \ltimes (N_1 \times N_2)$, where N_1 and N_2 are nilpotent Lie groups, and where the action on $\mathbb{R}$ contracts the metric on N_1 while extending it on N_2. We obtain new quasi-isometric invariants and classifications for these spaces.

MSC2020: 22E25, 51F30, 53C24.

Introduction

Let (X, d_X) and (Y, d_Y) be two Gromov hyperbolic spaces. Their *horospherical product*, denoted by $X \bowtie Y$, is constructed by combining X and Y, and lies in the direct product $X \times Y$. It no longer has negative curvature, however its geometry is still very rigid (see Section 1.2 for the definition). This way of combining two hyperbolic spaces appears to unify the construction of metric spaces such as Diestel–Leader graphs, treebolic spaces and Sol geometries, which are the horospherical products constructed out of a regular infinite tree or the hyperbolic plane $\mathbb{H}_2$.

Quasi-isometric classification and existing rigidity results

In [14], a mainstay of geometric group theory, Gromov points out the importance of quasi-isometric invariants in groups. The quasi-isometric classification of groups, or metric spaces, has since been a wide and prolific research domain (see [17] for a nice survey on this topic). For the family of solvable groups, there are still a lot of open cases.

The first result was obtained in [10] where Farb and Mosher provided a quasi-isometric classification of solvable Baumslag–Solitar groups $\mathrm{BS}(1, n)$. Then Eskin, Fisher and Whyte obtained the quasi-isometric classification of lamplighter groups and Sol geometries in [8; 9]. In [8; 10], the horospherical product construction of their respective groups is crucial in their proofs.

Eskin, Fisher and Whyte [8] also answered a question asked by Woess in [23] about the existence of vertex-transitive graphs not quasi-isometric to any Cayley graph: when m and n are coprime integers, the Diestel–Leader graphs $T_m \bowtie T_n$ are such graphs.

Peng [21; 22] and Dymarz [7], using similar methods as in [8; 9], generalised the description of the quasi-isometries for Lie groups of the form $\mathbb{R} \ltimes \mathbb{R}^p$. Peng [21; 22] proved that a subgroup of finite index of the quasi-isometry group of Lie groups of the form $\mathbb{R}^m \ltimes \mathbb{R}^n$ is a product of groups of bilipschitz maps.

Statement of results

The main goal of our work is to generalise the methods and techniques developed by Eskin, Fisher and Whyte to a wider set of horospherical products $X \bowtie Y$. In order to do that, the spaces X and Y are endowed with appropriate measures (see Definition 3.1). Once endowed with suitable measures, X and Y are called horopointed admissible spaces.

To be more precise let X (respectively X', Y, Y') be a horopointed admissible space with exponential growth parameter m (respectively m', n, n'). When X is a regular tree, the parameter m is related to the degree of X. When X is a negatively curved Lie group $\mathbb{R} \ltimes_A N$, the parameter m is $\mathrm{tr}(A)$, the trace of A.

Let $\Phi : X \bowtie Y \to X' \bowtie Y'$ be a quasi-isometry. The map Φ is called a *product map* if and only if there exist two maps $\Phi^X : X \to X$ (or $\Phi^X : X \to Y$) and $\Phi^Y : Y \to Y$ (or $\Phi^Y : Y \to X$) such that for all $(x, y) \in X \bowtie Y$ we have either

$$\Phi(x, y) = (\Phi^X(x), \Phi^Y(y)) \quad \text{or} \quad \Phi(x, y) = (\Phi^Y(y), \Phi^X(x)).$$

Our main theorem states that, when $m > n$ and $m' > n'$, any quasi-isometry $\Phi : X \bowtie Y \to X' \bowtie Y'$ is close to a product map.

Theorem A (geometric rigidity) *Let X, X', Y and Y' be horopointed admissible measured metric spaces with $m > n$ and $m' > n'$ and let $\Phi : X \bowtie Y \to X' \bowtie Y'$ be a quasi-isometry. Then there exist two quasi-isometries $\Phi^X : X \to X'$ and $\Phi^Y : Y \to Y'$ such that*

$$d_{\bowtie}(\Phi, (\Phi^X, \Phi^Y)) < +\infty.$$

This is a generalisation of Theorems 2.1 and 2.3 of [8]. While completing the proof of this result, we obtained a first quasi-isometry invariant in horospherical products.

Theorem B *When $m > n$, the parameter $\frac{m}{n}$ is a quasi-isometry invariant.*

Let $\mathbb{R} \ltimes_{A_1} N_1$ and $\mathbb{R} \ltimes_{A_2} N_2$ be two simply connected, negatively curved, solvable Lie groups (also called *Heintze groups*). In Section 5 we show that this couple of Heintze groups is admissible, and that the condition $m > n$ is equivalent to $\operatorname{tr}(A_1) > \operatorname{tr}(A_2)$. We obtain a necessary condition for the existence of a quasi-isometry on solvable Lie groups. The horospherical product of these two Heintze groups is isomorphic to

$$G := \mathbb{R} \ltimes_{\operatorname{Diag}(A_1, -A_2)} (N_1 \times N_2),$$

defined by the diagonal action of $\mathbb{R}$, $t \mapsto (\exp(tA_1), \exp(-tA_2))$ on $N_1 \times N_2$.

We say that G is *Carnot-Sol type* if N_1 and N_2 are Carnot groups and if A_1 and A_2 are multiples of Carnot derivations of N_1 and N_2 respectively. In the literature (see [19] for example), Carnot type stands for Lie groups with $N_2 = \{1\}$. Here we extend the denominations to nonhyperbolic Lie groups.

Using the previous quasi-isometry invariants we obtain the following quasi-isometry classification.

Theorem C *Let $G = \mathbb{R} \ltimes_{\operatorname{Diag}(A_1, -A_2)} (N_1 \times N_2)$ and $G' = \mathbb{R} \ltimes_{\operatorname{Diag}(A_1', -A_2')} (N_1' \times N_2')$ be Carnot-Sol type, nonunimodular Lie groups. Then*

$$(1) \qquad G \text{ and } G' \text{ are quasi-isometric} \iff G \text{ and } G' \text{ are isomorphic.}$$

The case where $N_2 = \{1\}$ is treated in Corollary 12.4 of [19].

Recall that a group G is called *metabelian* if $[G, G]$ is abelian (when both N_1 and N_2 are euclidean spaces). In this case, a similar quasi-isometry classification is deduced from [21; 22]. Both the quasi-isometry classification for the metabelian groups and for Carnot-Sol type groups are special cases of Conjecture 19.113 of [5] that we recall.

Conjecture 0.1 *Let S and S' be completely solvable Lie groups. Then S and S' are quasi-isometric if and only if they are isomorphic.*

Classifying completely solvable Lie groups up to quasi-isometry would yield the quasi-isometry classification of all connected Lie groups; see [4].

For $i \in \{1, 2\}$, let N_i and N_i' be two simply connected, nilpotent groups and let $A_i \in \operatorname{Lie}(N_i)$ and $A_i' \in \operatorname{Lie}(N_i')$ be derivations. Let $G := \mathbb{R} \ltimes_{\operatorname{Diag}(A_1, -A_2)} (N_1 \times N_2)$ and $G' := \mathbb{R} \ltimes_{\operatorname{Diag}(A_1', -A_2')} (N_1' \times N_2')$.

In this general setting of horospherical products of Heintze groups we have the following necessary conditions for being quasi-isometric.

Proposition D *Let us assume that* $\operatorname{tr}(A_1) > \operatorname{tr}(A_2)$ *and* $\operatorname{tr}(A_1') > \operatorname{tr}(A_2')$. *If* G *and* G' *are quasi-isometric, then we have that, for* $i \in \{1, 2\}$,

(1) N_i *and* N_i' *are bilipschitz*;

(2) A_i *and* $\frac{\operatorname{tr}(A_1)}{\operatorname{tr}(A_1')} A_i'$ *share the same characteristic polynomial.*

With the same setting, using the geometric rigidity on self quasi-isometries of this family of solvable Lie groups, we provide a characterisation of their quasi-isometry group.

Recall that for F a metric space, $\mathrm{QI}(F)/\sim$ is the group of self quasi-isometries of F, up to finite distance. (This equivalence relation is required since a quasi-isometry only has a coarse inverse.) Recall also that $\mathrm{Bilip}(F)$ stands for the group of self bi-Lipschitz maps of F. Then we have:

Theorem E *If* $\operatorname{tr}(A_1) \neq \operatorname{tr}(A_2)$,

$$\mathrm{QI}(\mathbb{R} \ltimes_{\mathrm{Diag}(A_1, -A_2)} (N_1 \times N_2))/\sim = \mathrm{Bilip}(N_1) \times \mathrm{Bilip}(N_2). \tag{2}$$

Here we choose the horospherical product metric on $\mathbb{R} \ltimes_{\mathrm{Diag}(A_1,-A_2)} (N_1 \times N_2)$.

In the course of this proof we also obtain that any self quasi-isometry of $\mathbb{R} \ltimes_{\mathrm{Diag}(A_1,-A_2)} (N_1 \times N_2)$ is a rough isometry. Le Donne, Pallier and Xie [18] proved that when you change the left-invariant Riemannian metric of one of these solvable Lie groups, the identity map is a rough similarity. Hence self quasi-isometries are rough isometries with respect to any left-invariant distances.

Outline of the proof

Let X and Y be two Gromov hyperbolic spaces, and let $\beta_X : X \to \mathbb{R}$ and $\beta_Y : Y \to \mathbb{R}$ be two Busemann functions. We call height functions h_X and h_Y the opposite of the Busemann functions. The horospherical product of X and Y, denoted by $X \bowtie Y$, is defined as the set of points in $X \times Y$ such that the two Busemann functions (or the height functions) add up to zero:

$$X \bowtie Y := \{(x, y) \in X \times Y \mid \beta_X(x) + \beta_Y(y) = 0\}.$$

A Busemann function is associated with a unique point on the boundary. We call any geodesic ray in the equivalence class of this point a *vertical* geodesic ray.

In order to generalise the proof of Eskin, Fisher and Whyte developed in [8; 9], the horospherical products have to be equipped with appropriate measures presented in Definition 3.1.

Briefly speaking, for the measured space (X, μ^X), the measure μ^X must verify three assumptions. Assumption (E1) allows us to disintegrate μ^X on its horospheres, assumption (E2) provides us with a bounded geometry on horospheres and (E3) ensures an exponential contraction (of exponent m) of the horospheres' measures in the upward vertical direction.

Let X (respectively X', Y, Y') be a horopointed admissible space with exponential growth parameter m (respectively m', n, n').

Most of this paper focuses on proving Theorem A. To do so we will use three major tools:

- We use *coarse vertical quadrilaterals*, which are realised by four points (the vertices) whose neighbourhoods are linked by vertical geodesics (the edges). In Proposition 2.11, we show that coarse vertical quadrilaterals are rigid: two of the four points almost share the same X-coordinate and the two other almost share the same Y-coordinate.
- We use *box tilings* of different scales for $X \bowtie Y$, suitable for the vertical flow. The boxes correspond to euclidean rectangular cuboids in the Sol geometry.
- We use *coarse differentiation*: given a quasi-isometry $\Phi : X \bowtie Y \to X' \bowtie Y'$, there exists a suitable scale R for the box tiling of $X \bowtie Y$. Suitable here means that the image by Φ of most vertical geodesic segments of length R are close to a vertical geodesic segment.

With these tools, the proof can be summarised as follows. Let $\Phi : X \bowtie Y \to X' \bowtie Y'$ be a quasi-isometry.

Step 1 By the *coarse differentiation*, there exists a scale R such that in the box tiling at scale R of $X \bowtie Y$, the quasi-isometry Φ mostly preserves the vertical direction on most of the boxes at scale R. This means that on most of the boxes, most vertical geodesic segments are sent close to a vertical geodesic segment by Φ.

Step 2 Then in most of the boxes at scale R, most of the vertical quadrilaterals are sent close to vertical quadrilaterals by Φ. Therefore, by the rigidity property of these configurations, on most of the boxes $\mathcal{B}$ the quasi-isometry Φ is close to a product map $\widehat{\Phi}_{|\mathcal{B}} = (\widehat{\Phi}^X, \widehat{\Phi}^Y)$ or $(\widehat{\Phi}^Y, \widehat{\Phi}^X)$.

Step 3 If $m > n$ and $m' > n'$ then all product maps have the form $\widehat{\Phi}_{\mathcal{B}} = (\Phi^X, \Phi^Y)$. Therefore by *gluing* them together, we show that there exists $L \gg R$ such that on *all* boxes at scale L, the map Φ is close to a product map $\widehat{\Phi} = (\Phi^X, \Phi^Y)$.

Step 4 We show that Φ quasi-respects the height, and then we use this last result on Φ^{-1} to show that Φ sends *all* vertical geodesics close to vertical geodesics. Therefore all vertical quadrilateral configurations are preserved by Φ, and hence Φ itself is close to a product map on all $X \bowtie Y$.

A major technical issue in this proof is to manage the notion of "almost all" vertical geodesic segments having a certain property. The disintegrable measure μ of assumption (E1) is not suited for this role since it concentrates the measure of a box on its bottom part. Therefore we introduce another disintegrable measure λ, constructed from μ, which (almost) equally weights the level-sets of the height function h in boxes.

Such a measure λ^X on X, together with a similar measure λ^Y on Y, allows us to define a suitable measure (later denoted by η) on the family of vertical geodesics contained in a box $\mathcal{B} \subset X \bowtie Y$.

The geometric rigidity has useful consequences when we understand the boundaries of X and Y. In this case, Theorem A leads to a description of the quasi-isometry-group of $X \bowtie Y$. In the last section of this paper, we detail such a description for the horospherical product of two Heintze groups.

Organisation of the paper

This work, about the geometric rigidity of quasi-isometries between two horospherical products, is organised as follows.

- In Section 2 we display the coarse differentiation in our context, and we discuss particular quadrilateral configurations of $X \bowtie Y$.
- Section 3 focuses on developing all the measure-theoretical tools required to achieve the rigidity results.
- Then, in Section 4, we follow the structure of the proof proposed by Eskin, Fisher and Whyte [8], invoking technical tools of previous chapters when required.
- In the last section we present an application of our theorem by providing new quasi-isometric classifications for some families of solvable Lie groups. We also provide a description of the quasi-isometry group of a wider family of solvable Lie groups.

1 Context

1.1 Gromov hyperbolic, Busemann spaces

Let $\delta > 0$, and let (X, d_X) and (Y, d_Y) be two δ-hyperbolic spaces (see [1, Part III, H, p. 399; 12] for more details on Gromov hyperbolic spaces). We present here the context in which we will construct our horospherical product. We require that X and Y are both *proper*, *geodesically complete*, *Busemann spaces*.

- A metric space is called proper if all closed metric balls are compact.
- A *geodesic line*, respectively *ray*, *segment*, of X is the isometric image of a Euclidean line, respectively half Euclidean line, interval, in X. We denote by $[x_1, x_2]$ a geodesic segment linking $x_1 \in X$ to $x_2 \in X$.
- A metric space X is called geodesically complete if all geodesics are infinitely extendable.
- A metric space is called Busemann if the distance between any couple of geodesics parametrised by arclength is a convex function. (See [20, Chapters 8 and 12] for more details on Busemann spaces.)

An important property of Gromov hyperbolic spaces is that they admit a nice compactification thanks to their *Gromov boundary*. We call two geodesic rays of X equivalent if their images are at finite Hausdorff distance. Let $w \in X$ be a base point. We define $\partial_w X$, the Gromov boundary of X, as the set of equivalence classes of geodesic rays starting from w. While $\partial_w X$ as a set depends on the choice of the base point w, it is topologically independent of w under the cone topology. We denote the Gromov boundary simply by ∂X when the choice of w does not matter topologically. The cone topology on $X \cup \partial X$ restricts to the natural topology on X, and with this topology, $X \cup \partial X$ is compact (see [1] for further details on the cone topology). In this context, the Gromov boundary coincides with the visual boundary.

Let us fix a point $a \in \partial X$ on the boundary. We call *vertical geodesic ray*, respectively *vertical geodesic line*, any geodesic ray in the equivalence class a, respectively any geodesic line with one of its half-lines in a. The study of these specific geodesic rays is central in this work.

The Busemann assumption removes some technical difficulties in a significant number of proofs in this work. If X is a Busemann space in addition to being Gromov hyperbolic, for all $x \in X$ there exists a unique vertical geodesic ray, denoted by V_x, starting at x. In fact the distance between two vertical geodesics starting at x is a convex and bounded function, hence decreasing and therefore constant equal to 0.

The construction of the *horospherical product* of two Gromov hyperbolic space X and Y requires the so called *Busemann functions*. Their definition is simplified by the Busemann assumption. Let us consider ∂X, the Gromov boundary of X (which, in this setting, is the same as the visual boundary). Both the boundary ∂X and $X \cup \partial X$, endowed with the natural Hausdorff topology, are compact. Then, given $a \in \partial X$ a point on the boundary, and $w \in X$ a base point, we define a Busemann function $\beta_{(a,w)}$ with respect to a and w by

$$\beta_{(a,w)}(x) := \limsup_{t \to +\infty} \big(d(x, V_w(t)) - t\big) \quad \text{for all } x \in X,$$

where V_w is the unique vertical geodesic ray starting from w. In all our results, X and Y will be proper, geodesically complete, Gromov hyperbolic, Busemann spaces, with some additional assumptions from time to time.

1.2 Horospherical products

Let $a^X \in \partial X$, $a^Y \in \partial Y$ be points on the boundaries and let $w^X \in X$, $w^Y \in Y$ be base points. Let us denote by $h^X := -\beta_{(a^X, w^X)}$ and $h^Y := -\beta_{(a^Y, w^Y)}$ the two corresponding height functions. The *horospherical product* of X and Y, relative to (a^X, w^X) and (a^Y, w^Y), denoted by $X \bowtie Y$ is defined by

$$X \bowtie Y := \{(x, y) \in X \times Y \mid h^X(x) + h^Y(y) = 0\}.$$

The set $X \bowtie Y$ can be seen as a diagonal in $X \times Y$. It is constructed by gluing X with an upside down copy of Y along their respective horospheres. This construction, illustrated in Figure 1, can also be seen as the union of the direct products between opposite horospheres in X and Y,

$$X \bowtie Y = \bigsqcup_{z \in \mathbb{R}} X_z \times Y_{-z}.$$

From now on, with a slight abuse, we omit the reference to the base points and points on the boundaries in the construction of the horospherical product.

To study the geometry of a horospherical product $X \bowtie Y$, we make additional assumptions on X and Y. We require them to be Gromov hyperbolic, Busemann, *geodesically complete* and *proper* metric spaces.

(1) X is *geodesically complete* if and only if all geodesic segments of X can be extended into a geodesic bi-infinite line.

(2) X is *proper* if and only if all closed metric balls of X are compact.

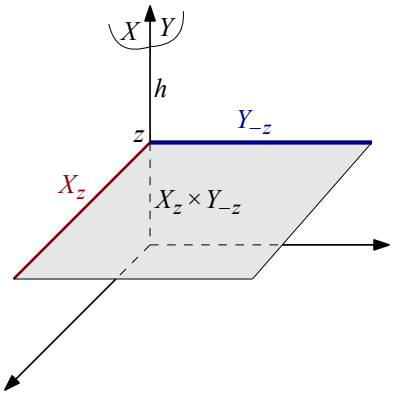


Figure 1: Horospherical product $X \bowtie Y$.

If X and Y satisfy these two additional conditions, the horospherical product $X \bowtie Y$ is connected (see [11, Property 3.11]).

Example 1.1 Let X be a Gromov hyperbolic, Busemann, geodesically complete and proper metric space. Then $X \bowtie \mathbb{R}$ is isometric to X. In particular, if V^Y is a vertical geodesic line of Y, the product $X \bowtie V^Y$ is an isometric embedding of X in $X \bowtie Y$.

The three (nontrivial) first examples of horospherical products appeared independently in the literature. They correspond to the case where X and Y are either a regular infinite tree T_m of degree m or the hyperbolic plane $\mathbb{H}^2$. We list them here:

(1) $T_m \bowtie T_n$ is the Diestel–Leader graph $\mathrm{DL}(m,n)$. When $m=n$, this horospherical product is a Cayley graph of the lamplighter group $\mathbb{Z} \wr \mathbb{Z}_m$. See Figure 2 for a subset of $T_3 \bowtie T_3$.

(2) $\mathbb{H}^{2,m} \bowtie \mathbb{H}^{2,n}$ is the Lie group $\mathbb{R} \ltimes_{(m,n)} \mathbb{R}^2 = \mathrm{Sol}(m,n)$, one of the eight Thurston geometries when $m=n$. By $\mathbb{H}^{2,m}$ we mean the manifold $\mathbb{R}^2$ endowed with the infinitesimal Riemannian metric $\mathrm{d}s^2 = e^{-2mz}\mathrm{d}x^2 + \mathrm{d}z^2$. The action associated to the aforementioned semidirect product is described by $(z,(x,y)) \mapsto (e^{mz}x, e^{-nz}y)$.

(3) $T_m \bowtie \mathbb{H}_2$ is a Cayley 2-complex of the Baumslag–Solitar group $\mathrm{BS}(1,m)$.

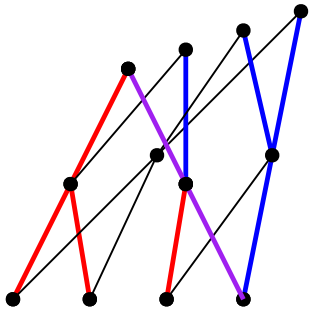


Figure 2: Small neighbourhood in $T_3 \bowtie T_3$.

The awareness of them being identically constructed from Gromov hyperbolic spaces came later, and a survey on these three examples is provided by Wolfgang Woess [24].

Another approach is to consider the hyperbolic plane $\mathbb{H}^{2,m}$ as the affine Lie group $\mathbb{R} \ltimes_m \mathbb{R}$ with action by multiplication $(z, x) \mapsto e^{mz}x$, and the Sol geometry $\mathrm{Sol}(m, n)$ as the Lie group $\mathbb{R} \ltimes_{(m,n)} \mathbb{R}^2$. In this context we have that $(\mathbb{R} \ltimes_m \mathbb{R}) \bowtie (\mathbb{R} \ltimes_n \mathbb{R}) = \mathbb{R} \ltimes_{(m,n)} \mathbb{R}^2$. The natural next step is to consider which Lie group can be taken as a component in a horospherical product.

A *Heintze group* is a Lie group of the form $\mathbb{R} \ltimes_A N$ with N a nilpotent Lie group, with A the derivation of the Lie algebra and where all eigenvalues of A have positive real part. Heintze [16] proved that any simply connected, negatively curved solvable Lie group is isomorphic to a Heintze group.

Moreover, a Busemann metric space is simply connected, hence any Gromov hyperbolic, Busemann Lie group is isomorphic to a Heintze group. Consequently, Heintze groups are natural candidates for the two components from which a horospherical product is constructed. Let $\mathbb{R} \ltimes_{A_1} N_1$ and $\mathbb{R} \ltimes_{A_2} N_2$ be two Heintze groups. We have

$$(\mathbb{R} \ltimes_{A_1} N_1) \bowtie (\mathbb{R} \ltimes_{A_1} N_1) = \mathbb{R} \ltimes_{\mathrm{Diag}(A_1,-A_2)} (N_1 \times N_2),$$

where $\mathrm{Diag}(A_1, -A_2)$ is the block diagonal matrix containing A_1 and $-A_2$ on its diagonal.

Xie [25] classified the subfamily of all negatively curved Lie groups $\mathbb{R} \ltimes \mathbb{R}^n$ up to quasi-isometry. In Section 5, we provide a description of the quasi-isometry group of the horospherical product of two Heintze groups, namely the solvable Lie groups $\mathbb{R} \ltimes_{\mathrm{Diag}(A_1,-A_2)} (N_1 \times N_2)$.

1.3 Settings

In this chapter we recall some material about horospherical products.

In order to lighten the notation, we will not fully describe the multiplicative and additive constants involved in inequalities. We will use the following notation instead.

Notation 1.2 Let $A, B \in \mathbb{R}$ and e a parameter (set, real numbers, ...). Let us write:

(1) $A \preceq_e B$ if and only if there exists a constant $M(e)$ depending only on e such that $A \leq M(e)B$.

(2) $A \asymp_e B$ if and only if $B \preceq_e A \preceq_e B$.

If the constant M is a specific integer such as 2, we will simply write $A \preceq B$, and similarly $A \succeq B$, $A \asymp B$. The notation $\preceq_e$ might also appear for parameters in several results of this paper. In this context it means that there exists a constant depending only on e such that the implied result holds.

A metric space is called geodesically complete if all its geodesic segments can be extended into geodesic lines, therefore when the space is also Gromov hyperbolic and Busemann space, with respect to $a \in \partial X$, any point is included in a vertical geodesic line (not necessarily unique).

We define the relative distance between two points x_1 and x_2 of X as

$$d_r(x_1, x_2) = d(x_1, x_2) - \Delta h(x_1, x_2).$$

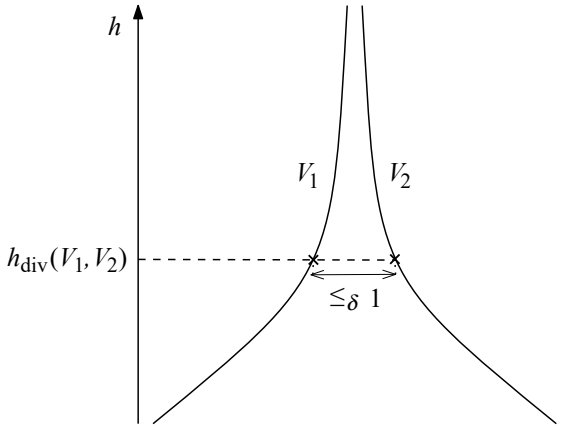


Figure 3: Figure of Corollary 1.4.

It can be understood as the distance along a level-set of the Busemann function. Let us recall Lemma 4.7 of [11].

Lemma 1.3 *Let X be a proper, δ-hyperbolic, Busemann space. Let V_1 and V_2 be two vertical geodesics of H. Let $t_1, t_2 \in \mathbb{R}$ and let us define $D := \frac{1}{2} d_r(V_1(t_1), V_2(t_2))$. Then, for all $t \in [0, D]$,*

$$|d_r(V_1(t_1 + D - t), V_2(t_1 + D - t)) - 2t| \leq 288\delta. \tag{3}$$

Corollary 1.4 *Let V_1, V_2 be two vertical geodesics of X. Then there exists a height $h_{\mathrm{div}}(V_1, V_2) \in \mathbb{R}$ from which V_1 and V_2 diverge from each other*:

(1) $\forall t \geq h_{\mathrm{div}}(V_1, V_2)$, *we have* $d(V_1(t), V_2(t)) \preceq_\delta 1$;

(2) $\forall t \leq h_{\mathrm{div}}(V_1, V_2)$, *we have* $|d(V_1(t), V_2(t)) - 2(h_{\mathrm{div}}(V_1, V_2) - t)| \preceq_\delta 1$.

This corollary is illustrated in Figure 3. We also have a more quantitative version.

Lemma 1.5 [11, Lemma 4.3] *Let H be a δ-hyperbolic and Busemann metric space, let x and y be two elements of H such that $h(x) \leq h(y)$, and let α be a geodesic linking x to y. Let us define $z = \alpha\big(\Delta h(x, y) + \frac{1}{2} d_r(x, y)\big)$, $x_1 := V_x\big(h(y) + \frac{1}{2} d_r(x, y)\big)$ the point of V_x at height $h(y) + \frac{1}{2} d_r(x, y)$ and $y_1 := V_y\big(h(y) + \frac{1}{2} d_r(x, y)\big)$ the point of V_y at the same height $h(y) + \frac{1}{2} d_r(x, y)$. Then*

(1) $\sup_{p \in \alpha}(h(p)) \geq h(y) + \frac{1}{2} d_r(x, y) - 96\delta$;

(2) $d(z, x_1) \leq 144\delta$;

(3) $d(z, y_1) \leq 144\delta$;

(4) $d(x_1, y_1) \leq 288\delta$.

We list here some notation we will use in later sections.

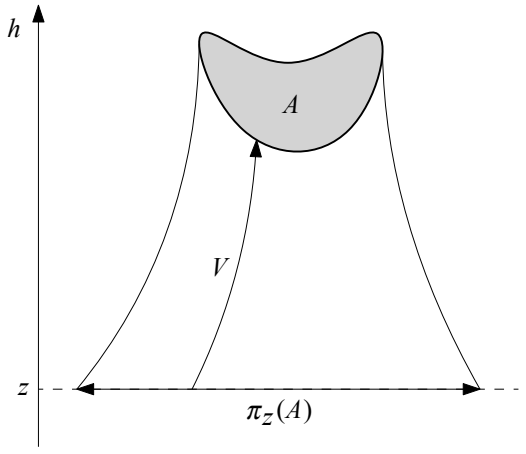


Figure 4: Projection of A on X_z.

Notation 1.6 Let X be a proper, geodesically complete, δ-hyperbolic, Busemann space.

(1) Let us denote the r-neighbourhood of U for all $U \subset X$ and for all $r \geq 0$ by

$$\mathcal{N}_r(U) := \{x \in X \mid d(x, U) \leq r\}. \tag{4}$$

(2) For all $x \in X$ let us denote by V_x the unique vertical geodesic ray such that $V_x(0) = x$.

(3) For a subset $A \subset X$, let us define

$$h^-(A) := \inf_{x \in A}(h(x)); \quad h^+(A) := \sup_{x \in A}(h(x)). \tag{5}$$

(4) For a subset $A \subset X$ and a height $z \in \mathbb{R}$, we denote the slice of A at the height z by $A_z := A \cap h^{-1}(z)$. Therefore the horospheres of X are denoted by X_z for $z \in \mathbb{R}$.

(5) Given a point $p \in X$ and a radius $r \in \mathbb{R}^+$, let us denote the ball of radius r included in the horosphere $X_{h(p)}$ by $D_r(p) := \{x \in X \mid h(x) = h(p) \text{ and } d(x, p) \leq r\} = B(p, r) \cap X_{h(p)}$.

(6) $\forall z \in \mathbb{R}$, $\forall U \subset X_z$, $\forall r > 0$, the r-interior of U in X_z is defined by

$$\mathrm{Int}_r(U) := \{p \in U \mid d(p, q) \geq r,\ \forall q \in X_z \setminus U\}.$$

Vertical geodesics of X can be understood as being normal to horospheres of X.

Definition 1.7 (projection on horospheres) Let X be a Gromov hyperbolic, Busemann, proper, geodesically complete metric space. Then, for all $A \subset X$ and all $z \leq h^-(A)$,

$$\pi_z(A) := \{x \in X_z \mid V_x \cap A \neq \varnothing\}. \tag{6}$$

The definition of this projection along the vertical flow is illustrated in Figure 4. The following lemma shows that the projection of a disk on a horosphere is almost a disk, It will be used in further subsequent sections.

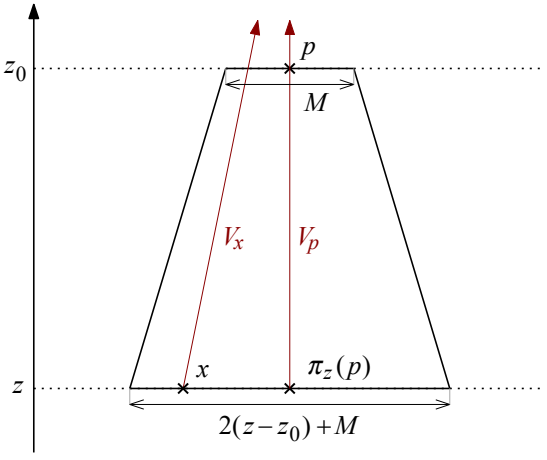


Figure 5: Proof of Lemma 1.8.

Lemma 1.8 *Let X be a Gromov hyperbolic, Busemann, proper, geodesically complete metric space. Let $z_0 \in \mathbb{R}$ and $p \in X_{z_0}$. Then for $M \geq 288\delta$ we have that, for all $z \leq z_0$ and for all $p_z \in \pi_z(\{p\})$,*

$$D_{2(z_0-z)-M}(p_z) \subset \pi_z(D_M(p)) \subset D_{2(z_0-z)+M}(p_z).$$

Proof This lemma is a corollary of Lemma 1.3 and is illustrated in Figure 5. Let $M = 288\delta$ be the constant involved in Lemma 1.3.

Let us prove the first inclusion. Let $x \in D_{2(z_0-z)-M}(p_z)$. Then $d(x, p_z) \leq 2(z_0 - z) - M$. Let us denote by V_x a vertical geodesic containing x and V_p a vertical geodesic containing p and p_z. We apply Lemma 1.3 with $t_1 = t_2 = z$, $V_1 = V_x$ and $V_2 = V_p$, then $D = \frac{d(x,p_z)}{2}$. Moreover

$$z + D = z + \frac{d(x, p_z)}{2} \leq z + (z_0 - z) - \frac{M}{2} \leq z_0.$$

Therefore, by the Busemann convexity of X, the distance between vertical geodesic ray is convex and bounded, hence decreasing. Therefore

$$\begin{aligned} d(V_x(z_0), p) = d(V_x(z_0), V_p(z_0)) &\leq d(V_x(z + D), V_p(z + D)), \\ &\leq M \quad \text{(by Lemma 1.3 used with } t = 0), \end{aligned}$$

which means that $x \in \pi_z(D_M(p))$.

Let us now prove the second inclusion, which is

$$\pi_z(D_M(p)) \subset D_{2(z_0-z)+M}(p_z). \tag{7}$$

Let $x \in \pi_z(D_M(p))$. Then $d(V_x(z_0), V_p(z_0)) \leq M$. Therefore by the triangle inequality

$$\begin{aligned} d(x, p_z) = d(V_x(z), V_p(z)) &\leq d(V_x(z), V_x(z_0)) + d(V_x(z_0), V_p(z_0)) + d(V_p(z_0), V_p(z)) \\ &\leq (z_0 - z) + M + (z_0 - z) = 2(z_0 - z) + M. \end{aligned}$$

Hence $x \in D_{2(z_0-z)+M}(p_z)$. □

Notation 1.6 can be extended to horospherical products.

Notation 1.9 Let X and Y be two proper, hyperbolic, geodesically complete, Busemann spaces.

(1) We denote the r-neighbourhood of U, for all $U \subset X \bowtie Y$ and for all $r \geq 0$, by

$$\mathcal{N}_r(U) := \{p \in X \bowtie Y \mid d_{\bowtie}(p, U) \leq r\}. \tag{8}$$

(2) The difference of height between two points $a, b \in X \bowtie Y$ is still denoted by $\Delta h(a, b) := |h(a) - h(b)|$.

(3) We still denote, for all $z \in \mathbb{R}$ and $A \subset X \bowtie Y$, by $A_z := A \cap h^{-1}(z)$ the "slice" of A at the height z.

(4) We still denote, for all $r \geq 0$ and $p \in X \bowtie Y$, by

$$D_r(p) := \{x \in X \mid h(p) = h(x) \text{ and } d_{\bowtie}(p, x) \leq r\} = B(p, r) \cap (X \bowtie Y)_{h(p)}$$

the ball of radius r in the height level set containing p.

We recall other useful results of [11] that we will use later. First the fact that the height function is Lipschitz.

Lemma 1.10 [11, Lemma 3.6] *Let N be an admissible norm, and let $d_{\bowtie}$ the distance on $X \bowtie Y$ induced by N. Then the height function is 1-Lipschitz with respect to the distance $d_{\bowtie}$, i.e.,*

$$\forall\, p, q \in X \bowtie Y, \quad d_{\bowtie}(p, q) \geq \Delta h(p, q). \tag{9}$$

Here is a description of the distance in Horospherical products.

Theorem 1.11 [11, Corollary 4.13] *For all $p, q \in X \bowtie Y$,*

$$|d_{\bowtie}(p, q) - (d_Y(p^Y, q^Y) + d_X(p^X, q^X) - \Delta h(p, q))| \preceq_{\bowtie} 1.$$

Here is one central result of [1]. Let us denote by $l(c)$ the length of a path c.

Proposition 1.12 [1, Proposition 1.6, p. 400] *Let X be a δ-hyperbolic geodesic space. Let c be a continuous path in X. If $[p, q]$ is a geodesic segment connecting the endpoints of c, then, for every $x \in [p, q]$,*

$$d(x, \mathrm{im}(c)) \leq \delta |\log_2 l(c)| + 1.$$

We also provide two more definitions that will be used in future sections. First a projection on level-sets of the height function.

Definition 1.13 Let $z_0, z \in \mathbb{R}$ and let $U \subset (X \bowtie Y)_{z_0}$. Then we define the projection of U on $(X \bowtie Y)_z$ by

$$\pi_z^{\bowtie}(U) := \{p \in (X \bowtie Y)_z \mid \exists V \text{ a vertical geodesic such that } p \in V \text{ and } V \cap U \neq \varnothing\}.$$

Then we define X-horospheres and Y-horospheres as horospheres of hyperbolic spaces embedded in $X \bowtie Y$, illustrated in Figure 6.

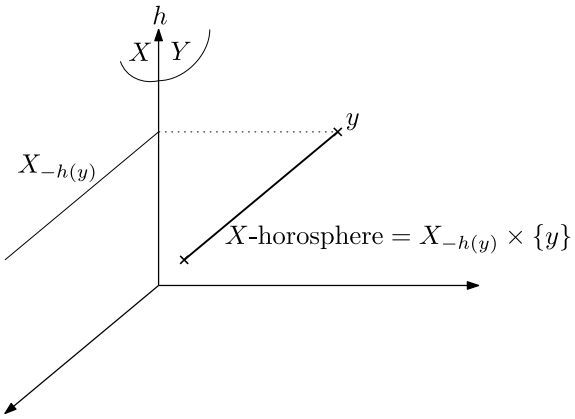


Figure 6: X-horosphere in $X \bowtie Y$.

Definition 1.14 The set $H \subset X \bowtie Y$ is called

(1) an X-horosphere if there exists $y \in Y$ such that $H = X \bowtie \{y\} = X_{-h(y)} \times \{y\}$,

(2) a Y-horosphere if there exists $x \in X$ such that $H = \{x\} \bowtie Y = \{x\} \times Y_{-h(x)}$.

From now on, we will work in a horospherical product $X \bowtie Y$ of two proper, geodesically complete, δ-hyperbolic and Busemann spaces.

2 Metric aspects and metric tools in horospherical products

Throughout this section we fix two constants $k \geq 1$ and $c \geq 0$. We recall the notions of quasi-isometry and quasi-geodesic.

Definition 2.1 ((k,c)-quasi-isometry) Let (E, d_E) and (F, d_F) be two metric spaces. A map $\Phi : E \to F$ is called a (k,c)-quasi-isometry if and only if

(1) for all $x, x' \in E$, we have $k^{-1} d_E(x,x') - c \leq d_F(\Phi(x), \Phi(x')) \leq k d_E(x,x') + c$;

(2) for all $y \in F$, there exists $x \in E$ such that $d(\Phi(x), y) \leq c$.

A map satisfying Definition 2.1(1) is called a quasi-isometric embedding of E.

Definition 2.2 ((k,c)-quasigeodesic) Let E be a metric space. A (k,c)-quasigeodesic segment, respectively ray, line, of E is a (k,c)-quasi-isometric embedding of a segment, respectively $[0; +\infty)$, $\mathbb{R}$, into E.

Gouëzel and Shchur [13, Lemma 2.1] proved that any (k,c)-quasigeodesic segment is included in the $2c$-neighbourhood of a continuous $(k, 4c)$-quasigeodesic segment sharing the same endpoints. Therefore, without loss of generality, we may consider that all quasi-geodesic segments are continuous.

This section gathers several geometric results on horospherical products, including the generalisation in our context of Lemmas 4.6, 3.1 and the coarse differentiation previously obtained by Eskin, Fisher

and Whyte in [8]. Propositions 2.6, 2.11 and Corollary 2.7 of this section will be especially useful in the following proofs.

At first, a reader who is more interested in the rigidity result on horospherical product can take these propositions for granted and jump to the next sections.

When $A \asymp_e B$, and $e = (X \bowtie Y, d)$ is a horospherical product, we shall write $A \asymp_{\bowtie} B$ as a short-cut, and similarly $\preceq_{\bowtie}$, $\succeq_{\bowtie}$ and $M(\bowtie)$ for a constant depending only on the metric horospherical product $(X \bowtie Y, d_{\bowtie})$.

2.1 ε-monotonicity

We introduce ε-monotone quasigeodesics, which happen to remain close to vertical geodesics. This fact plays a key role in our argument and will be proved later.

Definition 2.3 (ε-monotone quasigeodesic) Let $\varepsilon \geq 0$ and let $\alpha : [0, R] \to X \bowtie Y$ be a quasigeodesic segment. Then α is called ε-monotone if and only if

$$\forall t_1, t_2 \in [0, R], \quad \big(h(\alpha(t_1)) = h(\alpha(t_2))\big) \implies (|t_1 - t_2| \leq \varepsilon R). \tag{10}$$

Since α is assumed to be continuous, a 0-monotone quasigeodesic has monotone height, $h \circ \alpha$ is either decreasing or increasing. We first show that in $X \bowtie Y$, the projections on X and Y of an ε-monotone quasigeodesic are also quasigeodesics.

Theorem 2.4 *Let $\varepsilon > 0$, $R > \frac{1}{\varepsilon}$, and $\alpha = (\alpha^X, \alpha^Y) : [0, R] \to X \bowtie Y$ be an ε-monotone (k, c)-quasigeodesic segment. Then there exists a constant $M(\bowtie, k, c)$ (depending only on $\bowtie$, k and c) such that α^X and α^Y are $(4k, M\varepsilon R)$-quasigeodesics.*

A portion of the proof of Theorem 2.4 is illustrated in Figure 7.

Proof We know that $\forall p_1 = (p_1^X, p_1^Y),\ p_2 = (p_2^X, p_2^Y) \in X \bowtie Y$ we have (this is the admissible assumption we made on the norm underneath the distance $d_{\bowtie}$)

$$d_{\bowtie}(p_1, p_2) \geq \frac{d_X(p_1^X, p_2^X) + d_Y(p_1^Y, p_2^Y)}{2}. \tag{11}$$

Therefore we have that α^X satisfies the upper-bound assumption of quasigeodesics

$$\forall s_1, s_2 \in [0, R],\ d_X(\alpha^X(s_1), \alpha^X(s_2)) \leq 2d_{\bowtie}(\alpha(s_1), \alpha(s_2)) \leq 2k|s_1 - s_2| + 2c.$$

We want to find an appropriate $c' \geq c$ such that α^X satisfies the lower-bound condition of a $(4k, c')$-quasigeodesic. Let $c' \geq c$ and let us assume that α^X does not satisfy the lower-bound condition of a $(4k, c')$-quasigeodesic, we will show that this provides us with an upper-bound on c'. Indeed, consider $s_1, s_2 \in [0, R]$ such that

$$0 \leq d_X(\alpha^X(s_1), \alpha^X(s_2)) \leq \frac{1}{4k}|s_1 - s_2| - c'. \tag{12}$$

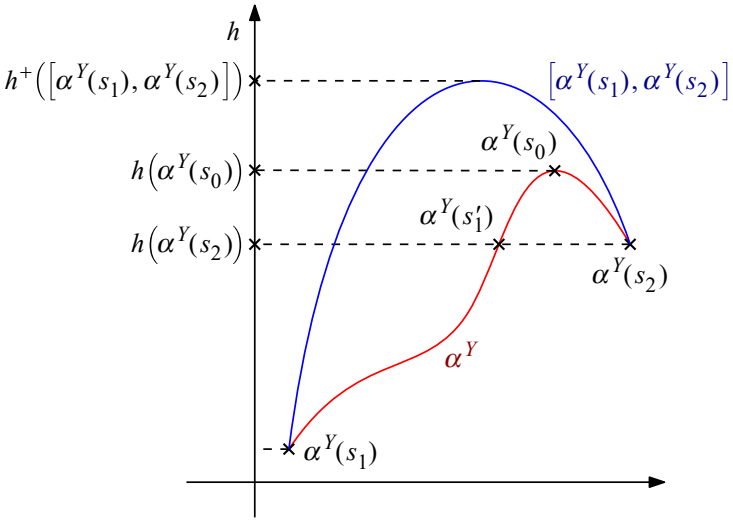


Figure 7: Proof of Theorem 2.4.

Therefore, by the Lipschitz property of h,

$$\begin{aligned}\Delta h(\alpha^X(s_1),\alpha^X(s_2)) &\le d_X(\alpha^X(s_1),\alpha^X(s_2)) \le \frac{1}{4k}|s_1-s_2|-c'\\ &\le \tfrac{1}{4}d_{\bowtie}(\alpha(s_1),\alpha(s_2))+\frac{c}{4}-c' \quad \text{(since } \alpha \text{ is a } (k,c)\text{-quasigeodesic).}\end{aligned} \tag{13}$$

Theorem 1.11 gives us the existence of a constant $M(\bowtie)$ depending only on X, Y and the underlying norm of $d_{\bowtie}$ such that

$$\begin{aligned}&d_Y(\alpha^Y(s_1),\alpha^Y(s_2))\\ &\quad\ge d_{\bowtie}(\alpha(s_1),\alpha(s_2))-d_X(\alpha^X(s_1),\alpha^X(s_2))+\Delta h(\alpha(s_1),\alpha(s_2))-M\\ &\quad\ge d_{\bowtie}(\alpha(s_1),\alpha(s_2))-d_X(\alpha^X(s_1),\alpha^X(s_2))-M\\ &\quad\ge d_{\bowtie}(\alpha(s_1),\alpha(s_2))-\frac{1}{4k}|s_1-s_2|+c'-M \quad \text{(by assumption (12))}\\ &\quad\ge d_{\bowtie}(\alpha(s_1),\alpha(s_2))-\tfrac{1}{4}d_{\bowtie}(\alpha(s_1),\alpha(s_2))-\frac{c}{4k}+c'-M \quad \text{(since } \alpha \text{ is a } (k,c)\text{-quasigeodesic)}\\ &\quad\ge \tfrac{1}{2}d_{\bowtie}(\alpha(s_1),\alpha(s_2))-\frac{c}{4}+c'-M \quad \text{(since } k\ge 1\text{).}\end{aligned} \tag{14, 15}$$

Without loss of generality, we may assume that

$$\max\big(h(\alpha^Y(s_1)),h(\alpha^Y(s_2))\big)=h(\alpha^Y(s_2)).$$

Applying Lemma 1.5 on the geodesic $[\alpha^Y(s_1),\alpha^Y(s_2)]$ of Y gives us

$$h^+([\alpha^Y(s_1),\alpha^Y(s_2)])\ge h(\alpha^Y(s_2))+\tfrac{1}{2}\big(d_Y(\alpha^Y(s_1),\alpha^Y(s_2))-\Delta h(\alpha^Y(s_1),\alpha^Y(s_2))\big)-M(\bowtie).$$

However α^Y is a continuous path between $\alpha^Y(s_1)$ and $\alpha^Y(s_2)$, then by Proposition 1.12, there exists $s_0 \in [s_1, s_2]$ such that

$$h(\alpha^Y(s_0)) \geq h(\alpha^Y(s_2)) + \tfrac{1}{2}\big(d_Y(\alpha^Y(s_1), \alpha^Y(s_2)) - \Delta h(\alpha^Y(s_1), \alpha^Y(s_2))\big) \\ - \delta \log_2\big(d_Y(\alpha^Y(s_1), \alpha^Y(s_2))\big) - M(\bowtie).$$

Therefore, by inequalities (13) and (15),

$$\begin{aligned} h(\alpha^Y(s_0)) &\geq h(\alpha^Y(s_2)) + \tfrac{1}{4} d_{\bowtie}(\alpha(s_1), \alpha(s_2)) - \tfrac{1}{8} d_{\bowtie}(\alpha(s_1), \alpha(s_2)) - \frac{c}{4} + c' - \frac{c}{8} + \tfrac{1}{2}c' \\ &\quad - \delta \log_2\big(d_Y(\alpha^Y(s_1), \alpha^Y(s_2))\big) - \frac{M(\bowtie)}{2} \\ &\geq h(\alpha^Y(s_2)) + \tfrac{1}{8} d_{\bowtie}(\alpha(s_1), \alpha(s_2)) - \delta \log_2\big(d_Y(\alpha^Y(s_1), \alpha^Y(s_2))\big) + \tfrac{3}{2}c' - M(\bowtie, c). \end{aligned}$$

However

$$2d_{\bowtie} \geq d_X + d_Y \geq d_Y,$$

hence

$$h(\alpha^Y(s_0)) \geq h(\alpha^Y(s_2)) + \tfrac{1}{8} d_{\bowtie}(\alpha(s_1), \alpha(s_2)) - \delta \log_2\big(d_{\bowtie}(\alpha(s_1), \alpha(s_2))\big) + \tfrac{3}{2}c' - M(\bowtie, c). \tag{16}$$

Furthermore, there exists $r_0 \in \mathbb{R}$ depending only on δ such that $\forall r \geq r_0$, $\frac{1}{8}r - \delta \log_2(r) > \frac{1}{10}r$ holds. Therefore, one of the two following statements holds:

(a) $d_{\bowtie}(\alpha(s_1), \alpha(s_2)) < r_0$.

(b) $\frac{1}{8} d_{\bowtie}(\alpha(s_1), \alpha(s_2)) - \delta \log_2\big(d_{\bowtie}(\alpha(s_1), \alpha(s_2))\big) \geq \frac{1}{10} d_{\bowtie}(\alpha(s_1), \alpha(s_2))$.

We will deal with the first case (a) at the end of the proof. Let us assume that

$$d_{\bowtie}(\alpha(s_1), \alpha(s_2)) \geq r_0$$

hence (b), then, by inequality (16),

$$h(\alpha^Y(s_0)) \geq h(\alpha^Y(s_2)) + \tfrac{1}{10} d_{\bowtie}(\alpha(s_1), \alpha(s_2)) + \tfrac{3}{2}c' - M(\bowtie, c). \tag{17}$$

Then we either have $d_{\bowtie}(\alpha(s_1), \alpha(s_2)) \leq M(\bowtie, c)$ (up to multiplying by 10 the constant M), or $h(\alpha^Y(s_0)) \geq h(\alpha^Y(s_2))$. In the case $d_{\bowtie}(\alpha(s_1), \alpha(s_2)) \leq M(\bowtie, c)$, we have

$$|s_1 - s_2| \preceq_{k,c,\bowtie} 1$$

since α is a quasigeodesic, and therefore $c' \preceq_{k,c,\bowtie} 1$ following assumption (12). In the other case we have

$$h(\alpha^Y(s_0)) \geq h(\alpha^Y(s_2)),$$

therefore there exists $s_1' \in [s_1, s_0]$ such that $h(\alpha^Y(s_1')) = h(\alpha^Y(s_2))$, since α is continuous. Hence

$$\begin{aligned}
&d_{\bowtie}(\alpha(s_1'), \alpha(s_2)) \\
&\quad \geq \frac{1}{k}|s_1' - s_2| - c \geq \frac{1}{k}(|s_1' - s_0| + |s_0 - s_2|) - M(c) \quad \text{(since } \alpha \text{ is a quasigeodesic)} \\
&\quad \geq \frac{1}{k^2}\big(d_{\bowtie}(\alpha(s_1'), \alpha(s_0)) + d_{\bowtie}(\alpha(s_0), \alpha(s_2))\big) - M(k, c) \quad \text{(since } \alpha \text{ is a quasigeodesic)} \\
&\quad \geq \frac{1}{k^2}\big(\Delta h(\alpha(s_1'), \alpha(s_0)) + \Delta h(\alpha(s_0), \alpha(s_2))\big) - M(k, c) \quad \text{(by Lemma 1.10)} \\
&\quad \geq \frac{2}{k^2}\Delta h(\alpha(s_0), \alpha(s_2)) - M(k, c) \quad \big(\text{since } h(\alpha(s_1')) = h(\alpha(s_2))\big) \\
&\quad \geq \frac{1}{5k^2} d_{\bowtie}(\alpha(s_1), \alpha(s_2)) + \frac{3}{k^2}c' - M(k, c, \bowtie) \quad \text{(by (17))}. \qquad (18)
\end{aligned}$$

Moreover assumption (12) implies $|s_1 - s_2| \geq 4kc'$. Then

$$d_{\bowtie}(\alpha(s_1), \alpha(s_2)) \geq \frac{1}{k}|s_1 - s_2| - c \geq 4c' - c.$$

Combined with inequality (18) it gives us

$$d_{\bowtie}(\alpha(s_1'), \alpha(s_2)) \geq \frac{19}{5k^2}c' - M(k, c, \bowtie).$$

Since α is ε-monotone and because $h(\alpha^Y(s_1')) = h(\alpha^Y(s_2))$, we have

$$\varepsilon R \geq |s_1' - s_2| \geq d_{\bowtie}(\alpha(s_1'), \alpha(s_2)) \geq \frac{19}{5k^2}c' - M(k, c, \bowtie).$$

Hence

$$c' \leq M(k)\varepsilon R + M(k, c, \bowtie).$$

We proved that if α^X does not satisfy the lower bound inequality for being a $(4k, c')$-quasigeodesic, then

$$c' \leq M(k)\varepsilon R + M(k, c, \bowtie).$$

Thus, when $\varepsilon R \geq 1$, there exists a constant $M(k, c, \bowtie)$ such that α^X is a $(4k, M\varepsilon R)$-quasigeodesic in both subcases of case (b) under consideration. Similarly we show that α^Y is a $(4k, M\varepsilon R)$-quasigeodesic segment of Y.

For case (a), let us assume that each couple of times $(s_1, s_2) \in [0, R]^2$ that contradicts the lower-bound hypothesis of a $(4k, M\varepsilon R)$-quasigeodesic verifies that $d_{\bowtie}(\alpha(s_1), \alpha(s_2)) < r_0$. Then α is a $(4k, r_0)$-quasigeodesic, with r_0 depending only on δ. Therefore α is in both cases a $(4k, M\varepsilon R)$-quasigeodesic, with M depending only on k, c and $X \bowtie Y$. □

In the sequel we denote by d_{Hff} the Hausdorff distance induced by $d_{\bowtie}$. In the proof of Proposition 2.6 we use a quantitative version of the quasigeodesic rigidity in a Gromov hyperbolic space, provided by the main theorem of [13].

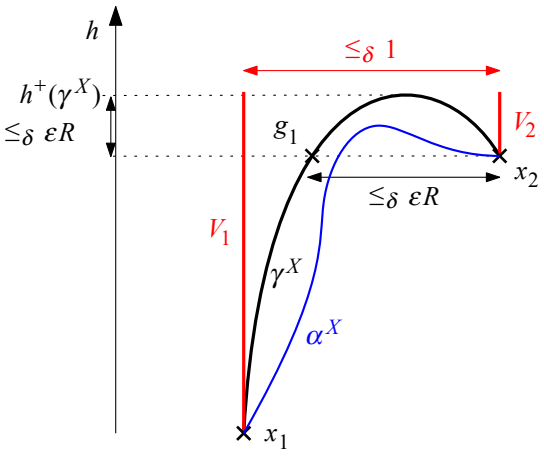


Figure 8: Proof of Proposition 2.6.

Theorem 2.5 [13] *Consider a (k, C)-quasigeodesic segment α in a δ-hyperbolic space X, and γ a geodesic segment between its endpoints. Then the Hausdorff distance $d_{\mathrm{Hff}}(\alpha, \gamma)$ between α and γ satisfies*

$$d_{\mathrm{Hff}}(\alpha, \gamma) \leq 92k^2(C + \delta).$$

This quantitative version allows us to have a linear control with respect to C on the Hausdorff distance, which is mandatory in our cases since $C \asymp \varepsilon R$. Combining this rigidity with the fact that projections α^X and α^Y are also ε-monotone provides us with the existence of vertical geodesic segments close to α.

Proposition 2.6 *Let $\varepsilon > 0$, $R > \frac{1}{\varepsilon}$, and $\alpha : [0, R] \to X \bowtie Y$ be an ε-monotone (k, c)-quasigeodesic segment. Then there exists a vertical geodesic segment $V : [0, R] \to X \bowtie Y$ such that*

$$d_{\mathrm{Hff}}(\mathrm{im}(\alpha), \mathrm{im}(V)) \preceq_{k,c,\delta} \varepsilon R. \tag{19}$$

This proposition corresponds to [8, Lemma 4.6].

Figure 8 is an illustration of the proof.

Proof By Theorem 2.4, α^X is a $(4k, M\varepsilon R)$-quasi-geodesic in X which is δ-hyperbolic, hence by Theorem 2.5 there exists a geodesic γ^X with the same endpoints as α^X such that

$$d_{\mathrm{Hff}}(\mathrm{im}(\alpha^X), \mathrm{im}(\gamma^X)) \preceq_{k,c,\delta} \varepsilon R.$$

Let us define $x_1 := \alpha^X(0)$ and $x_2 := \alpha^X(R)$. The quasigeodesic α^X is also ε-monotone. Furthermore [3, Proposition 2.2, page 19] gives us that γ^X, which links x_1 to x_2, is included in the 24δ-neighbourhood of two vertical geodesic rays V_1 and V_2 such that $V_1(0) = x_1$ and $V_2(0) = x_2$. Let us define $\tau := h^+(\gamma^X)$, and let us recall that $\forall t_1, t_2 \in \mathbb{R}^+$ and for $i \in \{1, 2\}$ we have $\Delta h(V_i(t_1), V_i(t_2)) = |t_1 - t_2|$. Let us also define by slight abuse $\gamma^X := \mathrm{im}(\gamma^X)$, $\alpha^X := \mathrm{im}(\alpha^X)$, $V_1 := \mathrm{im}(V_{1|[0,\tau-h(x_1)]})$ and $V_2 := \mathrm{im}(V_{2|[0,\tau-h(x_2)]})$. Since $\tau = h^+(\gamma^X) = h^+(V_1) = h^+(V_2)$ we have

$$d_{\mathrm{Hff}}(\gamma^X, V_1 \cup V_2) \preceq_\delta 1.$$

Hence by the triangle inequality

$$d_{\mathrm{Hff}}(\alpha^X, V_1 \cup V_2) \preceq_{k,c,\delta} \varepsilon R. \tag{20}$$

Without loss of generality we can assume that $h(x_1) \leq h(x_2)$. Furthermore γ^X is continuous, therefore there exists a point of γ^X close to both vertical geodesics (less than 24δ apart). Furthermore X is Busemann convex, hence the distance between the two vertical geodesics is decreasing. Therefore $d_X\big(V_1(\tau - h(x_1)), V_2(\tau - h(x_2))\big) \preceq_\delta 1$. We will use the ε-monotonicity of α^X to prove that $\tau \approx h(x_2)$. Let us denote by x_1' a point of α^X such that $h(x_1') = h(x_2)$ and such that $d_X(x_1', \mathrm{V}_1) \preceq_{k,c,\delta} \varepsilon R$. Since α^X is ε-monotone and a $(4k, M\varepsilon R)$-quasigeodesic we have that $d_X(x_1', x_2) \preceq_{k,c} \varepsilon R$, hence using the triangle inequality we have

$$\begin{aligned} d_X\big(V_1(h(x_2) - h(x_1)), x_2\big) &\leq d_X\big(V_1(h(x_2) - h(x_1)), x_1'\big) + d_X(x_1', x_2) \\ &\preceq_{k,c,\delta} \varepsilon R. \end{aligned} \tag{21}$$

Let $g_1 \in \mathrm{im}(\gamma^X)$ be the closest point to x_1 at height $h(x_2)$. Then

(1) $d_X\big(g_1, V_1(h(x_2) - h(x_1))\big) \preceq_\delta 1$;

(2) $d_X(g_1, x_2) \geq 2(h^+(\gamma^X) - h(x_2))$.

We recall that $\tau = h^+(\gamma^X)$. Then $d_X(g_1, x_2) \geq 2\tau - 2h(x_2) \geq 0$. Hence

$$\begin{aligned} |\tau - h(x_2)| \leq \tfrac{1}{2} d_X(g_1, x_2) &\leq \tfrac{1}{2} d_X\big(g_1, V_1(h(x_2) - h(x_1))\big) + \tfrac{1}{2} d_X\big(V_1(h(x_2) - h(x_1)), x_2\big) \\ &\preceq_{k,c,\delta} \varepsilon R \quad \text{(by definition of } g_1 \text{ and inequality (21)).} \end{aligned}$$

Hence $V_{2|[0,\tau - h(x_2)]}$ is a vertical geodesic segment of length $\preceq_{k,c,\delta} \varepsilon R$. Furthermore,

$$d_X\big(V_1(\tau - h(x_1)), V_2(\tau - h(x_2))\big) \preceq_\delta .$$

Therefore by the triangle inequality, any point of $V_{2|[0,\tau - h(x_2)]}$ is (up to a multiplicative constant) εR-close to $V_1(\tau - h(x_1))$. Therefore $d_{\mathrm{Hff}}(V_1 \cup V_2, V_1) \preceq_{k,c,\delta} \varepsilon R$. Therefore, by the triangle inequality we can improve inequality (20) as follows:

$$\begin{aligned} d_{\mathrm{Hff}}(\alpha^X, V_1) &\leq d_{\mathrm{Hff}}(\alpha^X, V_1 \cup V_2) + d_{\mathrm{Hff}}(V_1 \cup V_2, V_1) \\ &\preceq_{k,c,\delta} \varepsilon R \quad \text{(by inequality (20)).} \end{aligned}$$

We deduce similarly that α^Y is included in the $M\varepsilon R$-neighbourhood of a vertical geodesic segment V_2'. Therefore, α is included in the $M\varepsilon R$-neighbourhood of the vertical geodesic segment (V_1, V_2'). □

As a corollary, we show that the height function along an ε-monotone quasigeodesic is a quasi-isometry embedding of a segment into $\mathbb{R}$.

Corollary 2.7 *Let $\alpha : [0, R] \mapsto X \bowtie Y$ be an ε-monotone (k, c)-quasigeodesic segment. Then there exists a constant $M(k, c, \delta)$ such that the height function satisfies, for all $t_1, t_2 \in [0, R]$,*

$$\frac{1}{k}|t_1 - t_2| - M\varepsilon R \leq \Delta h(\alpha(t_1), \alpha(t_2)) \leq k|t_1 - t_2| + M\varepsilon R. \tag{22}$$

Proof Let $t_1, t_2 \in [0, R]$. The quasigeodesic upper-bound inequality

$$\Delta h(\alpha(t_1), \alpha(t_2)) \leq d_{\bowtie}(\alpha(t_1), \alpha(t_2)) \leq k|t_1 - t_2| + c$$

is straightforward since h is 1-Lipschitz and α is a (k, c)-quasigeodesic. To achieve the lower-bound inequality we use Proposition 2.6, hence there exists a vertical geodesic segment $V : [0, R] \to X \bowtie Y$ and a constant $M(k, c, \delta)$ such that

$$d_{\mathrm{Hff}}(\mathrm{im}(\alpha), \mathrm{im}(V)) \leq M\varepsilon R. \tag{23}$$

For $i \in \{1, 2\}$, let $s_i \in [0, R]$ be such that $d_{\bowtie}(\alpha(t_i), V(s_i)) \leq M\varepsilon R$. Then by the triangle inequality

$$\begin{aligned}\Delta h(\alpha(t_1), \alpha(t_2)) &\geq \Delta h(V(s_1), V(s_2)) - 2M\varepsilon R \\ &= |s_1 - s_2| - 2M\varepsilon R \quad \text{(since } V \text{ is vertical)}.\end{aligned}$$

However we can achieve, on $|s_1 - s_2|$, the lower-bound inequality

$$\begin{aligned}|s_1 - s_2| = d_{\bowtie}(V(s_1), V(s_2)) &\geq d_{\bowtie}(\alpha(t_1), \alpha(t_2)) - 2M\varepsilon R \quad \text{(by the triangle inequality)} \\ &\geq \frac{1}{k}|t_1 - t_2| - c - 2M\varepsilon R \quad \text{(since } \alpha \text{ is a quasigeodesic)},\end{aligned}$$

which provides us with

$$\Delta h(\alpha(t_1), \alpha(t_2)) \geq |s_1 - s_2| - 2M\varepsilon R \geq \frac{1}{k}|t_1 - t_2| - 5M\varepsilon R. \qquad \square$$

2.2 Coarse differentiation of a quasigeodesic segment

The coarse differentiation of a quasigeodesic α consists in finding a scale $r > 0$ such that a subdivision by pieces of length r of α contains almost only ε-monotone components (which are therefore close to vertical geodesic segments). See Figure 9.

Proposition 2.9 provides us with the existence of such an appropriate scale.

Lemma 2.8 *Let $k \geq 1$, $c \geq 0$ and $\varepsilon > 0$. There exists $M(k, c, \bowtie, \varepsilon)$ such that for all $r \geq M$, $N \geq M$ and for all non ε-monotone, (k, c)-quasigeodesic segment $\alpha : [0, r] \to X \bowtie Y$ we have*

$$\sum_{j=0}^{N-1} \Delta h\left(\alpha\left(\frac{jr}{N}\right), \alpha\left(\frac{(j+1)r}{N}\right)\right) - \Delta h(\alpha(0), \alpha(r)) \succeq_{k,c\bowtie} \varepsilon r. \tag{24}$$

This proposition corresponds to [8, Lemma 4.7].

Proof Since α is non ε-monotone, there exist $t_1, t_3 \in [0, r]$ such that

$$h(\alpha(t_1)) = h(\alpha(t_3)) \quad \text{and} \quad |t_1 - t_3| > \varepsilon r. \tag{25}$$

We can assume without loss of generality that $h(\alpha(0)) \leq h(\alpha(t_1)) \leq h(\alpha(r))$ with $t_1 < t_3$. Since α is a (k, c)-quasigeodesic we have $d_{\bowtie}(\alpha(t_1), \alpha(t_3)) \geq \frac{\varepsilon r}{k} - c$. By Theorem 1.11, there exists $M(\bowtie)$ such that

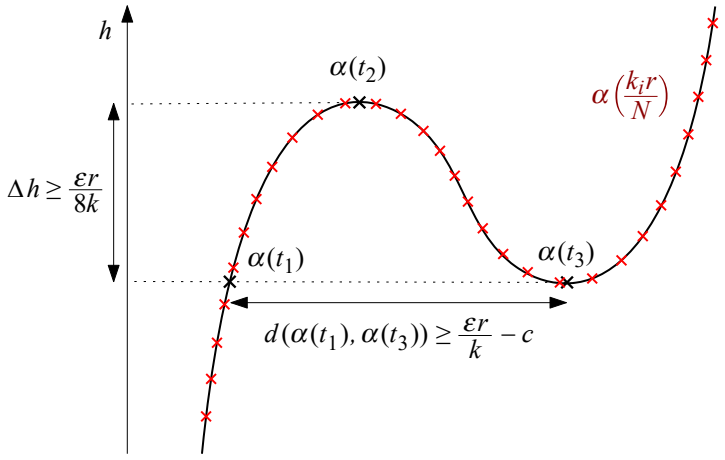


Figure 9: Subdivision of a quasi-geodesic.

$d_{\bowtie} \leq d_X + d_Y + M$. Then at least one of the two following inequalities holds:

(1) $d_X(\alpha^X(t_1), \alpha^X(t_3)) \succeq_{\bowtie} \frac{\varepsilon}{2k} r - M(\bowtie, c)$;

(2) $d_Y(\alpha^Y(t_1), \alpha^Y(t_3)) \succeq_{\bowtie} \frac{\varepsilon}{2k} r - M(\bowtie, c)$.

Let us assume the first inequality is true. By Lemma 1.5 applied to the geodesic segment $[\alpha^X(t_1), \alpha^X(t_3)]$,

$$\begin{aligned} h^+([\alpha^X(t_1), \alpha^X(t_3)]) &\geq \max\big(h(\alpha^X(t_1)), h(\alpha^X(t_3))\big) + \tfrac{1}{2}\big(d_X(\alpha^X(t_1), \alpha^X(t_3)) - \Delta h(\alpha^X(t_1), \alpha^X(t_3))\big) - 96\delta \\ &= h(\alpha^X(t_1)) + \tfrac{1}{2} d_X(\alpha^X(t_1), \alpha^X(t_3)) - 96\delta. \end{aligned}$$

Hence, there exists $t_2 \in [t_1, t_3]$ such that the assumed inequality provides us with

$$\begin{aligned} \Delta h(\alpha(t_1), \alpha(t_2)) &\succeq_{\bowtie} d_X(\alpha^X(t_2), [\alpha^X(t_1), \alpha^X(t_3)]) + \frac{\varepsilon r}{k} - M(\bowtie, c) \\ &\succeq_{\bowtie} \frac{\varepsilon r}{k} - \delta \log_2\big(d_{\bowtie}(\alpha^X(t_1), \alpha^X(t_3))\big) - M(\bowtie, c) \quad \text{(by Proposition 1.12)} \\ &\succeq_{\bowtie} \frac{\varepsilon r}{k} - \delta \log_2(r) - M(\bowtie, c). \end{aligned}$$

Similarly, assuming the second inequality provides us with the same lower-bound on $\Delta h(\alpha(t_1), (t_2))$. Furthermore there exists $M(\varepsilon, \bowtie, c)$ such that for $r \geq M$ we have $\frac{1}{2}\varepsilon r \geq \delta \log_2(r) + M(\varepsilon, \bowtie, c)$, hence

$$\Delta h(\alpha(t_1), \alpha(t_2)) \succeq_{\bowtie} \frac{\varepsilon r}{2k}. \tag{26}$$

Furthermore $\forall i \in \{1, 2, 3\}$ there exists $n_i \in \{0, \dots, N-1\}$ such that

$$\frac{n_i r}{N} \leq t_i \leq \frac{(n_i + 1) r}{N}.$$

Computing the sum of the successive differences of heights provides us with

$$\begin{aligned}
&\sum_{j=0}^{N-1} \Delta h\left(\alpha\left(\frac{jr}{N}\right), \alpha\left(\frac{(j+1)r}{N}\right)\right) \\
&\qquad \geq \Delta h\left(\alpha(0), \alpha\left(\frac{n_1 r}{N}\right)\right) + \Delta h\left(\alpha\left(\frac{n_1 r}{N}\right), \alpha\left(\frac{n_2 r}{N}\right)\right) + \Delta h\left(\alpha\left(\frac{n_2 r}{N}\right), \alpha\left(\frac{n_3 r}{N}\right)\right) \\
&\qquad\qquad + \Delta h\left(\alpha\left(\frac{n_3 r}{N}\right), \alpha(r)\right) \\
&\qquad \geq \Delta h(\alpha(0), \alpha(t_1)) + \Delta h(\alpha(t_1), \alpha(t_2)) + \Delta h(\alpha(t_2), \alpha(t_3)) + \Delta h(\alpha(t_3), \alpha(r)) - 6\left(\frac{kr}{N} + c\right) \\
&\qquad\qquad \text{(because } h \text{ is Lipschitz, } \alpha \text{ is a quasigeodesic and by the triangle inequality)} \\
&\qquad \geq \Delta h(\alpha(0), \alpha(r)) + 2\Delta h(\alpha(t_1), \alpha(t_2)) - 6\left(\frac{kr}{N} + c\right) \quad \big(\text{since } h(\alpha(t_1)) = h(\alpha(t_3))\big).
\end{aligned}$$

Using inequality (26) we have

$$\begin{aligned}
\sum_{j=0}^{N-1} \Delta h\left(\alpha\left(\frac{jr}{N}\right), \alpha\left(\frac{(j+1)r}{N}\right)\right) - \Delta h(\alpha(0), \alpha(r)) &\succeq_{\bowtie} \frac{\varepsilon r}{2k} - \frac{6kr}{N} - 6c \\
&\succeq_{k,c,\bowtie} \varepsilon r,
\end{aligned}$$

where the last line is since we assumed $N \geq M(k, c, \bowtie, \varepsilon)$. □

The next lemma asserts that, at some scale, most segments of a quasigeodesic are ε-monotone.

Proposition 2.9 *Let $k \geq 1$, $c \geq 0$, $\varepsilon > 0$ and let S be an integer. There exists $M(k, c, \bowtie, \varepsilon)$ such that for $r_0 \geq M$ and $N \geq M$ the following occurs. Let us define $L = N^S r_0$. Let $\alpha : [0, L] \to X \bowtie Y$ be a (k, c)-quasigeodesic segment. For all $s \in \{0, \dots, S\}$ we cut $[0, L]$ into segments of length $N^s r_0$, and we denote by A_s the set of these segments, that is,*

$$A_s := \left\{\alpha\big([kN^s r_0, (k+1)N^s r_0]\big) \mid k \in \{0, \dots, N^{S-s} - 1\}\right\},$$

and let $\delta_s(\alpha)$ be the proportion of segments in A_s which are not ε-monotone

$$\delta_s(\alpha) := \frac{\#\{\beta \in A_s \mid \beta \text{ is not } \varepsilon\text{-monotone}\}}{\#A_s}. \tag{27}$$

Then

$$\sum_{s=1}^{S} \delta_s(\alpha) \preceq_{k,c,\bowtie} \frac{1}{\varepsilon}. \tag{28}$$

Proof The idea is to cut $[0, L]$ into N segments of equal length, then to apply Lemma 2.8 to the elements of this decomposition which are not ε-monotone. Afterwards we decompose every piece of this decomposition into N segments of equal length to which we apply Lemma 2.8 if they are not ε-monotone.

The result follows by doing this subdecomposition S times in a row. To begin with, we need to deal with α being ε-monotone or not. Hence $\delta_S(\alpha) = 0$ or 1 and in either case thanks to Lemma 2.8 we have

$$\sum_{j=0}^{N-1} \Delta h\big(\alpha(jN^{S-1}r_0), \alpha((j+1)N^{S-1}r_0)\big) \succeq_{k,c,\bowtie} \Delta h(\alpha(0), \alpha(L)) + \delta_S(\alpha)\varepsilon L. \tag{29}$$

Then for all $j \in \{0, \dots, N-1\}$ such that $\alpha([jN^{S-1}r_0, (j+1)N^{S-1}r_0])$ is not ε-monotone

$$\sum_{k=0}^{N-1} \Delta h\big(\alpha(kN^{S-2}r_0 + jN^{S-1}r_0), \alpha((k+1)N^{S-2}r_0 + jN^{S-1}r_0)\big) \succeq_{k,c,\bowtie} \Delta h\big(\alpha(jN^{S-1}r_0), \alpha((j+1)N^{S-1}r_0)\big) + \frac{\varepsilon L}{N},$$

which happens $N\delta_{S-1}(\alpha)$ times. Therefore we have that

$$\begin{aligned}\sum_{i=0}^{N^2-1} \Delta h\big(\alpha(iN^{S-2}r_0), \alpha((i+1)N^{S-2}r_0)\big) &\succeq_{k,c,\bowtie} \Delta h(\alpha(0), \alpha(r)) + \delta_S(\alpha)\varepsilon L + N\delta_{S-1}(\alpha)\frac{\varepsilon L}{N} \\ &\succeq_{k,c,\bowtie} \Delta h(\alpha(0), \alpha(r)) + (\delta_S(\alpha) + \delta_{S-1}(\alpha))\varepsilon L.\end{aligned}$$

By doing this another $S-2$ times we obtain

$$\sum_{i=0}^{N^S-1} \Delta h\big(\alpha(ir_0), \alpha((i+1)r_0)\big) \succeq_{k,c,\bowtie} \Delta h(\alpha(0), \alpha(r)) + \varepsilon L \sum_{s=1}^{S} \delta_s(\alpha).$$

Furthermore we have, using the Lipschitz property of h,

$$\sum_{i=0}^{N^S-1} \Delta h\big(\alpha(ir_0), \alpha((i+1)r_0)\big) \le \sum_{i=0}^{N^S-1} d_{\bowtie}\big(\alpha(ir_0), \alpha((i+1)r_0)\big) \le N^S(kr_0 + c) \le 2kL \quad \left(\text{with } r_0 \ge \frac{c}{k}\right).$$

Hence

$$\sum_{s=1}^{S} \delta_s(\alpha) \preceq_{k,c,\bowtie} \frac{1}{\varepsilon L} 2kL \preceq_{k,c,\bowtie} \frac{1}{\varepsilon}. \tag{30}$$

□

2.3 Height respecting tetrahedric quadrilaterals

In this subsection we show that a coarse tetrahedric quadrilateral whose sides are vertical geodesics has two vertices on the same X-horosphere, and the other two on the same Y-horosphere (see Definition 1.14 for the definition of such horospheres). We call such a configuration a *vertical quadrilateral*.

Definition 2.10 (orientation) We define the orientation function on the paths of $X \bowtie Y$ as follows. For all $T > 0$ and $\gamma : [0, T] \to X \bowtie Y$ we have

$$\text{orientation}(\gamma) = \begin{cases} \uparrow & \text{if } h(\gamma(0)) < h(\gamma(T)), \text{ upward,} \\ \downarrow & \text{if } h(\gamma(0)) > h(\gamma(T)), \text{ downward.} \end{cases} \tag{31}$$

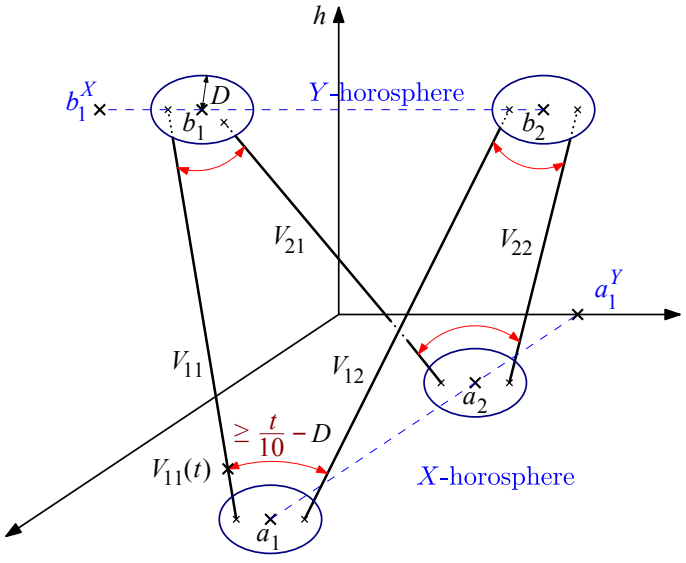


Figure 10: A coarse vertical quadrilateral of Proposition 2.11.

This lemma is strongly inspired by [8, Lemma 3.1], which establishes a similar result in the context of Diestel–Leader graphs and Sol geometries.

Proposition 2.11 (vertical quadrilateral lemma) *Let a_1, a_2, b_1, $b_2 \in X \bowtie Y$. Let $D > 1$ and for $i, j \in \{1,2\}$, let $V_{ij} : [0, l_{ij}] \to X \bowtie Y$ be vertical geodesic segments linking the D-neighbourhood of a_i to the D-neighbourhood of b_j, and diverging quickly from each other. More specifically, we assume, for all $i, j \in \{1,2\}$,*

(a) $d(V_{ij}(0), a_i) \leq D$;

(b) $d(V_{ij}(l_{ij}), b_j) \leq D$;

(c) $d(V_{i1}(t), \mathrm{im}(V_{i2})) \geq \frac{t}{10} - D, \ \forall t \in [0, l_{i1}]$;

(d) $d(V_{1j}(l_{1j} - t), \mathrm{im}(V_{2j})) \geq \frac{t}{10} - D, \ \forall t \in [0, l_{1j}]$.

If for all $i, j \in \{1,2\}$, $l_{ij} > 2D$ and the vertical geodesic segments V_{ij} share the same orientation, then there exists a constant $M(\bowtie)$ such that one of the two following statements holds:

(1) *The four vertical geodesics V_{ij} are upward oriented and a_2 is in the (MD)-neighbourhood of the X-horosphere containing a_1, and b_2 is in the (MD)-neighbourhood of the Y-horosphere containing b_1. Otherwise stated, we have $d_Y(a_1^Y, a_2^Y) \leq MD$ and $d_X(b_1^X, b_2^X) \leq MD$.*

(2) *The four vertical geodesics V_{ij} are downward oriented and a_2 is in the (MD)-neighbourhood of the Y-horosphere containing a_1, and b_2 is in the (MD)-neighbourhood of the X-horosphere containing b_1. Otherwise stated, we have $d_X(a_1^X, a_2^X) \leq MD$ and $d_Y(b_1^Y, b_2^Y) \leq MD$.*

Proposition 2.11 is illustrated in Figure 10.

Proof For all $i, j \in \{1,2\}$ let us define

$$a_i = (a_i^X, a_i^Y); \quad b_j = (b_j^X, b_j^Y); \quad V_{ij} = (V_{ij}^X, V_{ij}^Y). \tag{32}$$

The hypothesis (a) gives us

$$d(V_{i1}(0), V_{i2}(0)) \leq d(V_{i1}(0), a_i) + d(a_i, V_{i2}(0)) \leq 2D. \tag{33}$$

By hypothesis (b)

$$d(V_{1j}(l_{1j}), V_{2j}(l_{2j})) \leq 2D.$$

Without loss of generality we can assume that for all $i, j \in \{1,2\}$, we have orientation$(V_{ij}) = \uparrow$, which means that $h(a_i) \leq h(b_j)$. Then $\forall i, j \in \{1,2\}$ and $t \in [0, l_{ij}]$ we have $h(V_{ij}(t)) = t + h(V_{ij}(0))$, hence

$$h(V_{ij}^X(t)) = t + h(V_{ij}(0));$$
$$h(V_{ij}^Y(t)) = -t - h(V_{ij}(0)).$$

Since X and Y are Busemann convex spaces, $\forall i, j \in \{1,2\}$,

$$t \mapsto d_Y(V_{i1}^Y(t), V_{i2}^Y(t)) \text{ is convex on } [0, \min(l_{i1}, l_{i2})];$$
$$t \mapsto d_X(V_{1j}^X(l_{1j} - t), V_{2j}^X(l_{2j} - t)) \text{ is convex on } [0, \min(l_{1j}, l_{2j})].$$

Moreover, by the Busemann assumption, these maps remain convex up to a linear reparametrisation. The chosen pair of vertical geodesics, whether in X or Y, have endpoints separated by at most $2D$. Consequently, these vertical geodesics remain at a distance of at most $2D$ throughout the entire interval. Therefore

$$\begin{aligned} &\forall t \in [0, \min(l_{i1}, l_{i2})], \quad d_Y(V_{i1}^Y(t), V_{i2}^Y(t)) \leq 2D; \\ &\forall t \in [0, \min(l_{1j}, l_{2j})], \quad d_X(V_{1j}^X(l_{1j} - t), V_{2j}^X(l_{2j} - t)) \leq 2D. \end{aligned} \tag{34}$$

We can assume without loss of generality that $l_{11} \leq l_{21}$ and that $l_{12} \leq l_{22}$. Then

$$d_X(V_{11}^X(0), V_{21}^X(l_{21} - l_{11})) \leq 2D; \tag{35}$$

$$d_X(V_{12}^X(0), V_{22}^X(l_{22} - l_{12})) \leq 2D. \tag{36}$$

Let us define $\Delta l_1 = l_{21} - l_{11}$ and $\Delta l_2 = l_{22} - l_{12}$. Our goal is to show that these two real numbers are sufficiently close. We have, $\forall i, j \in \{1,2\}$,

$$\Delta h(a_i, b_j) - 2D \leq l_{ij} \leq \Delta h(a_i, b_j) + 2D.$$

By subtracting these inequalities we get

$$-h(a_2) + h(a_1) - 4D \leq l_{21} - l_{11} \leq -h(a_2) + h(a_1) + 4D;$$
$$-h(a_2) + h(a_1) - 4D \leq l_{22} - l_{12} \leq -h(a_2) + h(a_1) + 4D.$$

Then $|\Delta l_1 - \Delta l_2| \leq 8D$. However

$$d_X(V_{21}^X(\Delta l_1), V_{22}^X(\Delta l_1)) \leq d_X(V_{21}^X(\Delta l_1), V_{11}^X(0)) + d_X(V_{11}^X(0), V_{12}^X(0)) \\ + d_X(V_{12}^X(0), V_{22}^X(\Delta l_2)) + d_X(V_{22}^X(\Delta l_2), V_{22}^X(\Delta l_1)).$$

By the inequalities (35) and (36) we obtain

$$\begin{aligned} d_X(V_{21}^X(\Delta l_1), V_{22}^X(\Delta l_1)) \leq & 2D + d_X(V_{11}^X(0), V_{12}^X(0)) + 2D + |\Delta l_1 - \Delta l_2| \\ \leq & 4D + 2D + 8D \leq 14D. \end{aligned} \tag{37}$$

By using assumption (c) and the characterisation of the distance on horospherical products we have

$$\begin{aligned} -D + \frac{\Delta l_1}{10} \leq & d_{\bowtie}(V_{21}(\Delta l_1), V_{22}(\Delta l_1)) \\ \leq & d_X(V_{21}^X(\Delta l_1), V_{22}^X(\Delta l_1)) + d_Y(V_{21}^Y(\Delta l_1), V_{22}^Y(\Delta l_1)) \\ & - \Delta h(V_{21}(\Delta l_1), V_{22}(\Delta l_1)) + M(\bowtie) \quad \text{(by Theorem 1.11)} \\ \leq & d_X(V_{21}^X(\Delta l_1), V_{22}^X(\Delta l_1)) + 2D + M \quad \text{(by inequality (34))} \\ \leq & 16D + M \quad \text{(by inequality (37)),} \end{aligned}$$

which provides us with $\Delta l_1 \leq 10(16D + M + D) = 170D + 10M$. We have

$$\begin{aligned} d_X(a_1^X, a_2^X) &\leq d_X(a_1^X, V_{11}^X(0)) + d_X(V_{11}^X(0), V_{21}^X(0)) + d_X(V_{21}^X(0), a_2^X) \\ &\leq d_X(V_{11}^X(0), V_{21}^X(\Delta l_1)) + d_X(V_{21}^X(\Delta l_1), V_{21}^X(0)) + 2D \\ &\leq 2D + 170D + 10M + 2D \leq 174D + 10M \quad \text{(by inequality (35)).} \end{aligned}$$

From this inequality we deduce that $|h(a_1) - h(a_2)| \leq 174D + 10M \preceq_{\bowtie} D$. Similarly we deduce

$$d_Y(b_1^Y, b_2^Y) \preceq_{\bowtie} D;$$
$$|h(b_1) - h(b_2)| \preceq_{\bowtie} D. \qquad \square$$

Four points which satisfies the assumption of Proposition 2.11 are called a coarse vertical quadrilateral with nodes of scale D.

2.4 Orientation and tetrahedric quadrilaterals

From now on we fix a (k, c)-quasi-isometry $\Phi : X \bowtie Y \to X \bowtie Y$. Let us consider a tetrahedric configuration consisting of two points on an X-horosphere, each connected by vertical geodesic segments to two points on a Y-horosphere, forming a total of four points and segments.

The following Proposition 2.13 states that if two points on an X-horosphere are sufficiently far from each other, if two points on an Y-horosphere are sufficiently far from each other and if the vertical geodesic segments have ε-monotone images under a (k, c)-quasi-isometry Φ, then all the images of the vertical geodesic segments by Φ share the same orientation.

We first show that their exists a constant $M(k,c,\bowtie)$ such that the concatenation of two consecutive ε-monotone quasigeodesic segments sharing the same orientation is an $M\varepsilon$-monotone quasigeodesic segment. This result will only be used in the proof of Proposition 2.13.

Lemma 2.12 *Let $k \geq 1$, $c \geq 0$, $D > 0$, $\varepsilon > 0$, $T \geq \frac{D+2c}{3\varepsilon}$ and let $\gamma : [0,T] \mapsto X \bowtie Y$ and $\gamma' : [0,T] \mapsto X \bowtie Y$ be two ε-monotone, (k,c)-quasigeodesic segments such that*

(1) orientation(γ) = orientation(γ');

(2) $d_{\bowtie}(\gamma(T), \gamma'(0)) \leq D$.

Let $\widetilde{\gamma} : [0,2T] \to X \bowtie Y$ be the concatenation of γ and γ', that is,

$$\widetilde{\gamma}(t) = \begin{cases} \gamma(t) & \text{if } t \in [0,T], \\ \gamma'(t-T) & \text{if } t \in]T,2T]. \end{cases} \tag{38}$$

Then there exists $M(k,c,D,\bowtie)$ such that $\widetilde{\gamma}$ is an $M\varepsilon$-monotone, $(k, M\varepsilon T)$-quasigeodesic segment.

Proof We can assume without loss of generality that γ and γ' are upward oriented, we first show that there exists $M(k,c,\bowtie)$ such that $\widetilde{\gamma}$ is $M\varepsilon$-monotone. Let $t_1, t_2 \in [0,2T]$ such that $h(\widetilde{\gamma}(t_1)) = h(\widetilde{\gamma}(t_2))$. If both t_1 and t_2 are in $[0,T]$ or both are in $]T,2T]$, there is nothing to do since γ and γ' are ε-monotone. Then we can assume without loss of generality that $t_1 \in [0,T]$ and $t_2 \in]T,2T]$. Since γ is upward oriented we have $h(\gamma(0)) < h(\gamma(T))$, therefore, because γ is ε-monotone and continuous, we have

$$h(\gamma(t_1)) \leq h(\gamma(T)) + k\varepsilon T + c \leq h(\gamma(T)) + 2k\varepsilon T, \tag{39}$$

otherwise, by continuity there exists t_1' in $[0,t_1]$ such that $h(\gamma(t_1')) = h(\gamma(T))$ contradicting the ε-monotonicity. Two cases arise:

(a) $\Delta h(\gamma'(t_2 - T), \gamma'(0)) \leq 2k\varepsilon T$;

(b) $\Delta h(\gamma'(t_2 - T), \gamma'(0)) > 2k\varepsilon T$.

Let us consider the first case (a). We know that $h(\gamma(t_1)) = h(\widetilde{\gamma}(t_1)) = h(\widetilde{\gamma}(t_2)) = h(\gamma'(t_2 - T))$ and that $\Delta h(\gamma(T), \gamma'(0)) \leq D$. Then by the triangle inequality we have

$$\Delta h(\gamma(t_1), \gamma(T)) = \Delta h(\gamma'(t_2 - T), \gamma(T)) \leq \Delta h(\gamma'(t_2 - T), \gamma'(0)) + \Delta h(\gamma'(0), \gamma(T)) \leq 2k\varepsilon T + D.$$

According to Corollary 2.7, h is a $(k, M\varepsilon T)$-quasi-isometry along ε-monotone quasigeodesics. Hence

$$|t_1 - T| \leq k\Delta h(\gamma(t_1), \gamma(T)) + M\varepsilon T \leq (2k^2 + M)\varepsilon T + kD \leq (4k^2 + M)\varepsilon T \quad \text{(by assumption on } T);$$
$$|t_2 - T| \leq k\Delta h(\gamma'(t_2 - T), \gamma'(0)) + M\varepsilon T \leq (2k^2 + M)\varepsilon T.$$

Therefore by the triangle inequality we obtain $|t_1 - t_2| \leq (3k^2 + M)\varepsilon(2T)$.

We consider now the second case (b). By Corollary 2.7, h is a $(k, M\varepsilon T)$-quasi-isometry, therefore

$$\Delta h(\gamma'(t_2 - T), \gamma'(0)) \geq \frac{1}{k}|t_2 - T| - M\varepsilon T.$$

Furthermore, γ' is upward oriented, hence we have that $h(\gamma'(0)) < h(\gamma'(t_2 - T))$, otherwise, as for γ, by continuity one can construct $t_2' \in [t_2, T + T']$ contradicting the ε-monotonicity of γ'. Hence we have

$$h(\gamma'(t_2 - T)) \geq h(\gamma'(0)) + \frac{1}{k}|t_2 - T| - M\varepsilon T.$$

In combination with inequality (39) it provides us with

$$\begin{aligned} h(\gamma(t_1)) &\leq h(\gamma(T)) + \varepsilon T \leq h(\gamma'(0)) + D + \varepsilon T \\ &\leq h(\gamma'(t_2 - T)) - \frac{1}{k}|t_2 - T| + D + (1 + M)\varepsilon T. \end{aligned}$$

However $h(\gamma(t_1)) = h(\gamma'(t_2 - T))$ by definition of t_1 and t_2, therefore $0 \leq -\frac{1}{k}|t_2 - T| + D + (1 + M)\varepsilon T$, which gives

$$|t_2 - T| \leq (1 + M)k\varepsilon T + kD \leq 3Mk\varepsilon T. \tag{40}$$

Hence

$$\Delta h(\gamma'(t_2 - T), \gamma'(0)) \leq d_{\bowtie}(\gamma'(t_2 - T), \gamma'(0)) \leq k|t_2 - T| + c \leq (3Mk^2 + 1)\varepsilon T.$$

Since $h(\gamma'(t_2 - T)) = h(\gamma(t_1))$, thanks to the triangle inequality we obtain

$$\begin{aligned} \Delta h(\gamma(t_1), \gamma(T)) &\leq \Delta h(\gamma(t_1), \gamma'(0)) + \Delta h(\gamma'(0), \gamma(T)) \\ &\leq (3Mk^2 + 1)\varepsilon T + D \leq (3Mk^2 + 2)\varepsilon T. \end{aligned} \tag{41}$$

Both inequalities (40) and (41) in combination with the fact that h is a $(k, M\varepsilon T)$-quasigeodesic segment provide us with

$$\begin{aligned} |t_1 - t_2| &= |t_1 - T| + |T - t_2| \leq k(3Mk^2 + 2)\varepsilon T + M\varepsilon T + 3Mk\varepsilon T \\ &\leq 9k^3 M\varepsilon T \leq \frac{9k^3 M}{2}\varepsilon(2T) \quad (\text{since } k \geq 1,\ M \geq 1). \end{aligned}$$

In the view of cases (a) and (b) we conclude that $\widetilde{\gamma}$ is $\frac{9k^3 M}{2}\varepsilon$-monotone.

To prove that $\widetilde{\gamma}$ is a $(k, 3M\varepsilon T)$-quasigeodesic segment, we must check the upper-bound and lower bound required. Let $t_1, t_2 \in [0, 2T]$, as for the ε-monotonicity property, since γ and γ' are (k, c)-quasigeodesics, we can assume that $t_1 \in [0, T]$ and $t_2 \in]T, 2T]$. By the triangle inequality, the upper-bound is straightforward:

$$\begin{aligned} d_{\bowtie}(\widetilde{\gamma}(t_1), \widetilde{\gamma}(t_2)) &= d_{\bowtie}(\gamma(t_1), \gamma'(t_2 - T)) \\ &\leq d_{\bowtie}(\gamma(t_1), \gamma(T)) + d_{\bowtie}(\gamma(T), \gamma'(0)) + d_{\bowtie}(\gamma'(0), \gamma'(t_2 - T)) \\ &\leq k(T - t_1) + c + D + k(t_2 - T) + c \\ &= k|t_2 - t_1| + 2c + D \\ &\leq k|t_2 - t_1| + 3\varepsilon T \quad (\text{by the assumed lower bound on } T). \end{aligned}$$

The last inequality holds because γ and γ' are (k,c)-quasigeodesics. To prove the lower-bound we will proceed similarly as for the ε-monotonicity. We have

$$\begin{aligned} d_{\bowtie}(\widetilde{\gamma}(t_1),\widetilde{\gamma}(t_2)) &= d_{\bowtie}(\gamma(t_1),\gamma'(t_2-T)) \\ &\geq \Delta h(\gamma(t_1),\gamma'(t_2-T)) \quad \text{(since } h \text{ is Lipschitz).} \end{aligned}$$

Similarly to inequality (39) we have

$$h(\gamma'(t_2-T)) \geq h(\gamma'(0)) - 2k\varepsilon T. \tag{42}$$

Therefore

$$\begin{aligned} \Delta h(\gamma(t_1),\gamma'(t_2-T)) &\geq h(\gamma'(t_2-T)) - h(\gamma(t_1)) \\ &= \big(h(\gamma'(t_2-T)) + \varepsilon T\big) - h(\gamma'(0)) + h(\gamma'(0)) - h(\gamma(T)) + h(\gamma(T)) - \big(h(\gamma(t_1)) - \varepsilon T\big) - 4k\varepsilon T \\ &= \big|\big(h(\gamma'(t_2-T)) + \varepsilon T\big) - h(\gamma'(0))\big| + \big|h(\gamma(T)) - \big(h(\gamma(t_1)) - \varepsilon T\big)\big| + h(\gamma'(0)) - h(\gamma(T)) - 4k\varepsilon T \\ &\qquad \text{(by inequalities (39) and (42))} \\ &\geq \big|h(\gamma'(t_2-T)) - h(\gamma'(0))\big| + \big|h(\gamma(T)) - h(\gamma(t_1))\big| - D - 8k\varepsilon T \quad \text{(by the triangle inequality)} \\ &\geq \frac{1}{k}|t_2 - T| - M\varepsilon T + \frac{1}{k}|T - t_1| - M\varepsilon T - D - 8k\varepsilon T \quad \text{(because } h \text{ is a } (k, M\varepsilon T)\text{-quasigeodesic).} \end{aligned}$$

Hence

$$\begin{aligned} d_{\bowtie}(\widetilde{\gamma}(t_1),\widetilde{\gamma}(t_2)) &\geq \Delta h(\gamma(t_1),\gamma'(t_2-T)) \\ &\geq \frac{1}{k}(t_2-t_1) - D - (2M+8k)\varepsilon T \geq \frac{1}{k}(t_2-t_1) - M'\varepsilon T, \end{aligned}$$

for M' a constant depending on k, c, D and $\bowtie$. This is the lower-bound we expected and proves that $\widetilde{\gamma}$ is a $(k, M'\varepsilon T)$-quasigeodesic. □

Proposition 2.13 *Let $h \in \mathbb{R}$ and let $k \geq 1$, $c \geq 0$ and $\varepsilon > 0$. Let $\Phi : X \bowtie Y \to X' \bowtie Y'$ be a (k,c)-quasi-isometry. Let $D > 1$ and $R > \frac{k2D+c}{\varepsilon}$. For $i, j \in \{1,2\}$ let a_i, b_j be four points of $X \bowtie Y$ satisfying $d(a_1,a_2) > 10kM\varepsilon R + 2kc$ and $d(b_1,b_2) \geq 10kM\varepsilon R + 2kc$, where M is the constant involved in Lemma 2.12, and let $V_{i,j} : [0,R] \to X \bowtie Y$ be four vertical geodesic segments linking the D-neighbourhood of a_j to the D-neighbourhood of b_i, such that*

- $h(V_{11}(0)) = h(V_{22}(0)) = h(a_1) = h(a_2) = h$;
- $h(V_{11}(R)) = h(V_{22}(R)) = h(b_1) = h(b_2) = h + R$;
- $h(V_{12}(0)) = h(V_{21}(0)) = h$;
- $h(V_{12}(R)) = h(V_{21}(R)) = h + R$;
- *$\Phi \circ V_{i,j}$ is ε-monotone.*

Then

$$\text{orientation}(\Phi \circ V_{11}) = \text{orientation}(\Phi \circ V_{22}).$$

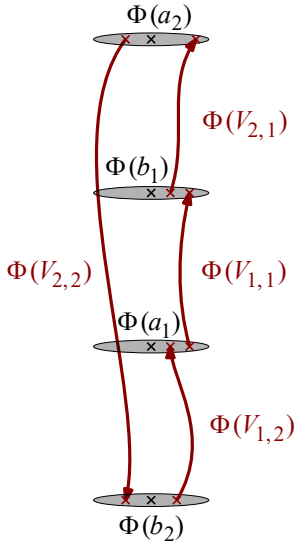


Figure 11: Case (a) in the proof of Proposition 2.13.

Proof Up to the additive constant D, one can consider $V_{1,1} \cup V_{2,1} \cup V_{2,2} \cup V_{1,2}$ as a coarse quadrilateral composed with a_i and b_j as its vertices, and with $V_{i,j}$ as its edges. To make the proof easier to follow, we shall use a vector of arrows to describe the orientations of the edges of the quadrilateral in play as

$$\text{orientation}(V_{1,1}, V_{2,1}, V_{2,2}, V_{1,2}) = (\uparrow, \downarrow, \uparrow, \downarrow).$$

Similarly, we consider orientations of the image of $V_{1,1} \cup V_{2,1} \cup V_{2,2} \cup V_{1,2}$ by Φ as the successive orientations of the paths $\Phi \circ V_{i,j}$. We will proceed by contradiction to prove the lemma. Let us assume that $\text{orientation}(\Phi \circ V_{1,1}) \neq \text{orientation}(\Phi \circ V_{2,2})$. We can assume without loss of generality that $\text{orientation}(\Phi(V_{1,1})) = \uparrow$, therefore $\text{orientation}(\Phi(V_{2,2})) = \downarrow$. Hence there are four possible orientations for $\Phi(V_{1,1} \cup V_{2,1} \cup V_{2,2} \cup V_{1,2})$:

$$\text{(a)}\ \ (\uparrow, \uparrow, \downarrow, \uparrow), \qquad \text{(b)}\ \ (\uparrow, \uparrow, \downarrow, \downarrow), \qquad \text{(c)}\ \ (\uparrow, \downarrow, \downarrow, \uparrow), \qquad \text{(d)}\ \ (\uparrow, \downarrow, \downarrow, \downarrow).$$

Let us consider the case (a) (illustrated in Figure 11). We have

$$\text{orientation}(\Phi(V_{2,1})) = \uparrow \quad \text{and} \quad \text{orientation}(\Phi(V_{1,2})) = \uparrow.$$

Hence we have

$$\text{orientation}(\Phi(V_{1,2})) = \text{orientation}(\Phi(V_{1,1})) = \text{orientation}(\Phi(V_{2,1})).$$

Furthermore Φ is a (k, c)-quasi-isometry and both $V_{1,2}(R)$ and $V_{1,1}(0)$ are close to a_1, hence

$$d_{\bowtie'}\big(\Phi(V_{1,2}(R)), \Phi(V_{1,1}(0))\big) \leq k2D + c.$$

Similarly we have

$$d_{\bowtie'}\big(\Phi(V_{1,1}(R)), \Phi(V_{2,1}(0))\big) \leq k2D + c.$$

Then by Lemma 2.12, there exists $M(k,c,\bowtie)$ such that the concatenation of $\Phi(V_{1,2})$, $\Phi(V_{1,1})$ and $\Phi(V_{2,1})$ is an $M\varepsilon$-monotone $(k, M\varepsilon T)$-quasigeodesic. Therefore by Proposition 2.6, there exists a constant $M(k,c,\bowtie)$ and a vertical geodesic segment $\widetilde{V}$ such that

$$d_{\mathrm{Hff}}(\widetilde{V}, \Phi(V_{1,2}) \cup \Phi(V_{1,1}) \cup \Phi(V_{2,1})) \leq M\varepsilon R. \tag{43}$$

Furthermore, applying Proposition 2.6 on $\Phi(V_{2,2})$ provides us with the existence of a vertical geodesic segment $\widetilde{V}'$ such that

$$d_{\mathrm{Hff}}(\widetilde{V}', \Phi(V_{2,2})) \leq M\varepsilon R. \tag{44}$$

Moreover $d_{\bowtie}(V_{2,2}(0), V_{2,1}(R)) \leq 2D$ (the two points are close to a_2) and $d_{\bowtie}(V_{2,2}(R), V_{1,2}(0)) \leq 2D$ (the two points are close to b_2), and therefore $\widetilde{V}$ and $\widetilde{V}'$ are two vertical geodesics with endpoints $(k2D + c) + 2M\varepsilon R$ close to $\Phi(a_2)$ and $\Phi(b_2)$. Thereby, these two vertical geodesic segments stay close to each other, we have

$$d_{\mathrm{Hff}}(\widetilde{V}, \widetilde{V}') \leq (k2D + c) + 2M\varepsilon R \leq 3M\varepsilon \quad \text{(by assumption on } R\text{)}.$$

Then, we show by the triangle inequality that $\Phi(a_1)$ is close to $\Phi(V_{2,2})$:

$$d_{\bowtie'}(\Phi(a_1), \Phi(V_{2,2})) \leq d_{\bowtie'}(\Phi(a_1), \widetilde{V}) + d_{\mathrm{Hff}}(\widetilde{V}, \widetilde{V}') + d_{\mathrm{Hff}}(\widetilde{V}', \Phi(V_{2,2})) \leq 5M\varepsilon R. \tag{45}$$

However, the assumption $d(a_1,a_2) > 10kM\varepsilon R + 2kc$ gives us that a_1 is sufficiently far from $V_{2,2}$:

$$\begin{aligned} &d_{\bowtie}(a_1, V_{2,2}(t)) \geq \Delta h(a_1, V_{2,2}(t)) = t, \\ &d_{\bowtie}(a_1, V_{2,2}(t)) \geq d_{\bowtie}(a_1,a_2) - d_{\bowtie}(a_2, V_{2,2}(t)) > 10kM\varepsilon R + 2kc - t \end{aligned}$$

for all $t \in [0, R]$. Therefore

$$\begin{aligned} d_{\bowtie'}\big(\Phi(a_1), \Phi(V_{2,2}(t))\big) &\geq k^{-1} d_{\bowtie}(a_1, V_{2,2}(t)) - c \\ &> \frac{t + 10kM\varepsilon R + 2kc - t}{2k} - c = 5M\varepsilon R \end{aligned}$$

for all $t \in [0, R]$, which contradicts inequality (45). Thereby, in case (a), $\Phi \circ V_{1,1}$ and $\Phi \circ V_{2,2}$ share the same orientation.

The other three cases (b), (c) and (d) are treated similarly. We first show $\Phi(V_{1,1} \cup V_{2,1} \cup V_{2,2} \cup V_{1,2})$ is in the $M\varepsilon R$-neighbourhood of two vertical geodesic segments which, depending on the case, have endpoints

(b) close to $\Phi(a_1)$ and $\Phi(a_2)$;

(c) close to $\Phi(b_1)$ and $\Phi(b_2)$;

(d) close to $\Phi(a_1)$ and $\Phi(b_1)$.

Which, depending on the case, contradicts the fact that

(b) $d_{\bowtie}(b_1, V_{2,2}(t)) > 5M\varepsilon R$;

(c) $d_{\bowtie}(a_1, V_{2,2}(t)) > 5M\varepsilon R$;

(d) $d_{\bowtie}(b_2, V_{1,1}(t)) > 5M\varepsilon R$. □

3 Measure and box-tiling

3.1 Appropriate measure and horopointed admissible space

In the setting of horospherical products, an important characteristic is that they are a union of products of horospheres.

As such, if one wants to endow them with a measure, it makes sense that the measure should disintegrate along these horospherical products, and should be related somehow to the measures and the geometries of the initial spaces and its horospheres.

The properties we present are satisfied when our initial spaces are Riemannian manifolds for instance, or graphs of bounded geometry. We will also see in Section 5 that Heintze groups are another set of spaces which satisfies them, making our requirements sound.

Definition 3.1 (admissible horopointed measured metric spaces) Let (X, d) be a δ-hyperbolic, Busemann, proper, geodesically complete, metric space, and let $a \in \partial X$ be a point on the Gromov boundary of X. A Borel measure μ^X on X will be said (X, a) horo-admissible if and only if (E1), (E2) and (E3) are satisfied:

(E1) (There exists a direction $a \in \partial X$ such that) μ^X is disintegrable along the height function h_a, that is, for all $z \in \mathbb{R}$, there exists a Borel measure μ_z^X on $X_z = h^{-1}(z)$ such that, for any measurable set $A \subset X$,

$$\mu^X(A) = \int_{z\in\mathbb{R}} \mu_z^X(A_z)\,\mathrm{d}z.$$

(E2) Controllable geometry for the measures μ_z^X on horospheres, there exists $M_0 \geq 288\delta$ such that

$$\forall x_1, x_2 \in X, \quad \mu^X_{h(x_1)}(D_{M_0}(x_1)) \asymp_X \mu^X_{h(x_2)}(D_{M_0}(x_2)).$$

(E3) There exists $m > 0$ such that for all $z_0 \in \mathbb{R}$, and for all measurable sets $U \subset X_{z_0}$,

$$\forall z \leq z_0, \quad e^{m(z_0-z)}\mu^X_{z_0}(U) \asymp_X \mu^X_z(\pi_z(U)).$$

The space (X, a, d, μ^X) will be called a horopointed admissible metric measured space, or just admissible.

The assumption (E2), in combination with Lemma 1.8, provides us with a uniform control on the measure of disks of any radius.

Lemma 3.2 *Let $r \geq M_0$. Then for all $x \in X$ we have*

$$\mu_{h(x)}(D_r(x)) \asymp_X e^{m\frac{r}{2}}.$$

Proof The proof is illustrated in Figure 12. Let V_x be a vertical geodesic line containing x and let $M_0 \geq 288\delta$ be the constant involved in assumption (E2). Let x_1 be the point of V_x at the height $h(x) + \frac{r+M_0}{2}$ and let x_2 be the point of V_x at the height $h(x) + \frac{r-M_0}{2}$. Applying Lemma 1.8 with $p = x_1$, $z_0 = h(x) + \frac{r+M_0}{2}$ and $z = h(x)$ provides us with

$$D_r(x) = D_{2(z_0-z)-M_0}(x) \subset \pi_{h(x)}(D_{M_0}(x_1)).$$

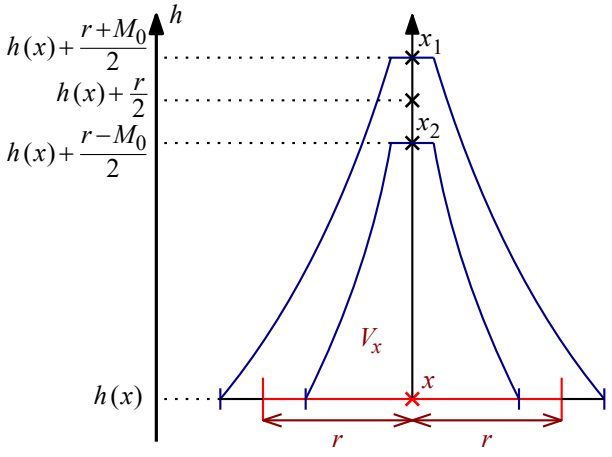


Figure 12: Proof of Lemma 3.2.

Similarly, applying Lemma 1.8 with $p = x_2$, $z_0 = h(x) + \frac{r-M_0}{2}$ and $z = h(x)$ provides us with $\pi_{h(x)}(D_{M_0}(x_2)) \subset D_r(x)$. Furthermore by assumption (E3) then assumption (E2) we have

$$\mu^X_{h(x)}\big(\pi_{h(x)}(D_{M_0}(x_1))\big) \asymp_X e^{m(\frac{r+M_0}{2})}\mu^X_{h(x_1)}(D_{M_0}(x_1)) \asymp_X e^{m\frac{r}{2}},$$

since M_0 depends only on X. Similarly we have $\mu^X_{h(x)}\big(\pi_{h(x)}(D_{M_0}(x_2))\big) \asymp_X e^{m\frac{r}{2}}$, therefore by the two previously obtained inclusions we have $\mu_{h(x)}(D_r(x)) \asymp_X e^{m\frac{r}{2}}$. □

Heuristically, the next lemma asserts that the measure of the boundary of a disk is small in comparison to the measure of the disk.

Lemma 3.3 *Let M_0 be the constant involved in assumption* (E2) *and let M be the constant involved in Corollary 1.4. Let $z_0 \in \mathbb{R}$, $x_0 \in X_{z_0}$ and $\mathcal{C} \subset X_{z_0}$ be a set containing $D_{M_0}(x_0)$ and contained in $D_{2M_0}(x_0)$. Then for all $z_1 \leq z_0$, and for all $r \leq 2|z_1 - z_0| - 2M_0 - M$, we have*

$$\mu^X_{z_1}\big(\mathrm{Int}_r(\pi^X_{z_1}(\mathcal{C}))\big) \asymp_{\bowtie} \mu^X_{z_1}(\pi^X_{z_1}(\mathcal{C})).$$

This lemma might seem to contradict Lemma 3.2, however the r-interior of a disk of radius R is very different from a disk of radius $R-r$ on horospheres, for R sufficiently greater than r.

Proof Let us define $J := \mathrm{Int}_r(\pi^X_{z_1}(\mathcal{C}))$. By definition we have

$$\pi^X_{z_1}(\mathcal{C}) \setminus J := \{x \in \pi^X_{z_1}(\mathcal{C}) \mid d_X(x, \pi^X_{z_1}(\mathcal{C})^c) < r\}. \tag{46}$$

At the height $z_1 + \frac{r}{2}$, let $x_1 \in \pi^X_{z_1+\frac{r}{2}}(\mathcal{C}) \setminus \pi^X_{z_1+\frac{r}{2}}(J)$, then, at the height z_1, there exists $x'_1 \in \pi^X_{z_1}(\mathcal{C}) \setminus J$ such that $x_1 \in V_{x'_1}$. Furthermore by the characterisation (46), there exists $x'_2 \in \pi^X_{z_1}(\mathcal{C})^c$ such that $d(x'_1, x'_2) \leq r$. Then by Corollary 1.4, there exists $M(\delta)$ such that

$$d_X\left(V_{x'_2}\left(z_1 + \frac{r}{2}\right), V_{x'_1}\left(z_1 + \frac{r}{2}\right)\right) = d_X\left(V_{x'_2}\left(z_1 + \frac{r}{2}\right), x_1\right) \leq M, \tag{47}$$

with $V_{x_2'}\left(z_1+\frac{r}{2}\right) \in \pi^X_{z_1+\frac{r}{2}}(\mathcal{C})^c$. Therefore by the triangle inequality and Lemma 1.8

$$\begin{aligned} d(x_1, \pi^X_{z_1+\frac{r}{2}}(x_0)) &\geq -d\left(x_1, V_{x_2'}\left(z_1+\frac{r}{2}\right)\right) + d\left(V_{x_2'}\left(z_1+\frac{r}{2}\right), \pi^X_{z_1+\frac{r}{2}}(x_0)\right) \\ &\geq 2|z_0-z_1| - r - M_0 - M. \end{aligned}$$

Since last inequality holds for all $x_1 \in \pi^X_{z_1+\frac{r}{2}}(\mathcal{C}) \setminus \pi^X_{z_1+\frac{r}{2}}(J)$, we have

$$D_{2|z_0-z_1|-r-M_0-M}(\pi^X_{z_1+\frac{r}{2}}(x_0)) \subset \pi^X_{z_1+\frac{r}{2}}(J).$$

Therefore by Lemma 1.8

$$D_{2|z_0-z_1|-M_0-M}(\pi^X_{z_1}(x_0)) \subset J.$$

Moreover, $J \subset \pi^X_{z_1}(\mathcal{C}) \subset D_{2|z_0-z_1|+M_0}(\pi^X_{z_1}(x_0))$, hence by Lemma 3.2

$$\mu^X_{z_1}(J) \asymp_X e^{|z_0-z_1|m} \asymp_X \mu^X_{z_1}(\pi^X_{z_1}(\mathcal{C})). \qquad \square$$

In order to achieve a rigidity result on horospherical products, we will need another measure λ^X in the same measure class as μ^X.

Definition 3.4 (measure λ^X of X) Let X be an admissible horopointed space. The measure λ^X on X is defined from a set of weighted measures λ^X_z on the level set X_z as

(1) for all $z \in \mathbb{R}$, $\lambda^X_z := e^{mz}\mu^X_z$,

(2) for all measurable sets $A \subset X$, $\lambda^X(A) := \int_{z\in\mathbb{R}} \lambda^X_z(A_z)\,\mathrm{d}z$,

where m is the constant involved in (E3).

For the log model of the hyperbolic plane, this measure λ^X turns out to be the Lebesgue measure on $\mathbb{R}^2$, and the measure μ^X is the Riemannian area. Up to a multiplicative constant, the measure λ^X is constant along the projections. By assumption (E3), the following property is immediate:

Property 3.5 For all measurable set $U \subset X$ we have

$$\forall z_1, z_2 \leq h^-(U), \quad \lambda^X_{z_1}(\pi_{z_1}(U)) \asymp_X \lambda^X_{z_2}(\pi_{z_2}(U)). \tag{48}$$

Otherwise stated we have the following relation between two push-forwards of the measure $(\pi_{z_2})_*\lambda^X_{z_2} \asymp_X (\pi_{z_1})_*\lambda^X_{z_1}$. They are push-forwards from a subset U of X to horospheres below U.

Following the fact that height level sets of $X \bowtie Y$ are direct products of horospheres, we define disintegrable measures on the horospherical products from the disintegrable measures on X and Y. We recall that

$$\forall z \in \mathbb{R}, \quad (X \bowtie Y)_z = X_z \times Y_{-z}.$$

Definition 3.6 (measure μ on $X \bowtie Y$) Let (X, μ^X) and (Y, μ^Y) be two admissible spaces. Then for all measurable sets $U \subset X \bowtie Y$, we define the measure μ on $X \bowtie Y$ by

$$\mu_{X\bowtie Y}(U) := \int_{\mathbb{R}} \mu^X_z \otimes \mu^Y_{-z}(U_z)\,\mathrm{d}z.$$

For all measurable sets $U \subset X \bowtie Y$ we have

$$\mu_{X\bowtie Y}(U) = \int_{\mathbb{R}} \left(\int_{y\in Y_{-z}} \mu_z^X(U_z^y) d\mu_{-z}^Y(y) \right) \mathrm{d}z,$$

where $U_z^y := \{x \in X \mid (x, y) \in U_z\}$. (This measure might be not well defined.)

Remark 3.7 A couple (X, Y) of horopointed admissible spaces is called admissible if the measure $\mu_{X\bowtie Y}$ of Definition 3.6 is well defined.

From now on we fix four horopointed metric spaces X, X', Y and Y', with $m > 0$ (respectively m', n, n') the constant of assumption (E3) for X (respectively X', Y, Y'). We will assume in Section 4.3 and afterwards that (X, Y) and (X', Y') are two admissible couples with $m > n$ and $m' > n'$.

We define similarly a measure $\lambda_{X\bowtie Y}$ on $X \bowtie Y$.

Definition 3.8 (measure λ on $X \bowtie Y$) Let (X, μ^X) and (Y, μ^Y) be two admissible spaces. Then, for all measurable subsets $U \subset X \bowtie Y$,

$$\lambda_{X\bowtie Y}(U) := \int_{\mathbb{R}} \lambda_z^X \otimes \lambda_{-z}^Y(U_z)\, \mathrm{d}z = \int_{\mathbb{R}} e^{(m-n)z} \mu_z^X \otimes \mu_{-z}^Y(U_z)\, \mathrm{d}z.$$

For all measurable subsets $U \subset X \bowtie Y$ we have

$$\lambda_{X\bowtie Y}(U) = \int_{\mathbb{R}} \left(\int_{y\in Y_{-z}} \lambda_z^X(U_z^y) d\lambda_{-z}^Y \right) \mathrm{d}z.$$

From now on, we will simply denote by μ the measure $\mu_{X\bowtie Y}$ and by λ the measure $\lambda_{X\bowtie Y}$.

3.2 Box-tiling of X

In this subsection we tile a proper, geodesically complete, Gromov hyperbolic and Busemann space X with pieces called boxes. This is inspired from [8, Lemma 3.4], which constructs these tilings for trees and the hyperbolic space.

Definition 3.9 (box at scale R) Let X be admissible horopointed space. Let M_0 be the constant of (E2), let $R > 0$, let x be a point of X and let $\mathcal{C}(x)$ be a subset of $X_{h(x)}$ containing $D_{M_0}(x)$ and contained in $D_{2M_0}(x)$. Then, the box $\mathcal{B}(x, \mathcal{C}(x), R)$ is defined by

$$\mathcal{B}(x, \mathcal{C}(x), R) := \bigcup_{z\in[h(x)-R, h(x)[} \pi_z(\mathcal{C}(x)).$$

We will often omit the parameter $\mathcal{C}(x)$ in the notation of a box. Later we depict an appropriate choice for these spaces $\mathcal{C}(x)$. The idea of the tiling is first to distinguish layers of thickness R, then to decompose each of these layers into disjoint boxes using a tiling of disjoint cells $\mathcal{C}(x)$ as the top of these boxes. In the log model of the hyperbolic plane, when the cell $\mathcal{C}(x)$ is a segment of an horosphere, the associated box is a rectangle of $\mathbb{R}^2$. Eskin, Fisher and Whyte [8] tiled the hyperbolic plane with translates of such a rectangle. However the space we consider might not be homogeneous, therefore we will tile Gromov hyperbolic spaces with boxes which are generically not the translate of one another. We recall that $\mathcal{N}_r$ refers to the r-neighbourhood of a subspace.

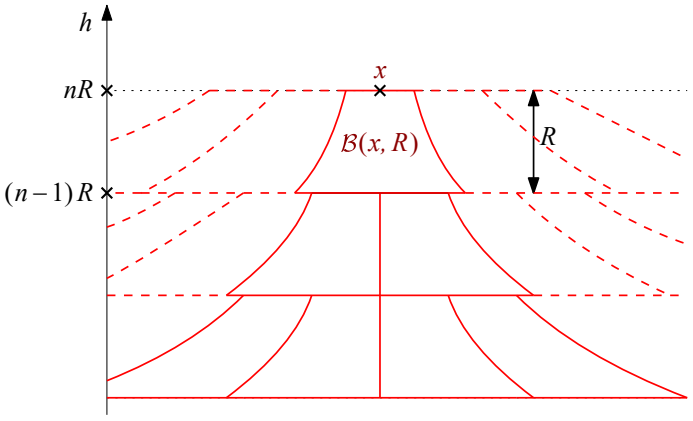


Figure 13: Box-tiling.

A subset of a metric space X is k-separated if and only if any two of its elements are at least at distance k. A maximal such set for the inclusion is called *maximal separating set*. We shall denote by $\mathcal{D}(X)$ such a set. The dependence of $\mathcal{D}(X)$ on k should be indicated; however, for simplicity and by slight abuse of notation, we will omit this dependence.

One easily sees that a maximal separated set is then a k-covering. That is the union of the metric ball of radius k centred at the points of $\mathcal{D}(X)$ cover the whole space.

To construct a box tiling, see Figure 13, of X we first fix a scale $R > 0$. Let M_0 be the constant involved in assumption (E2), then we chose a $2M_0$-maximal separating set $\mathcal{D}(X_{nR})$ of the horospheres X_{nR}, with $n \in \mathbb{Z}$. Such maximal separating sets exist since X is proper and so are X_{nR}. Let us call nuclei the points in these maximal separating sets. For every nucleus $x \in \mathcal{D}(X_{nR})$, we fix a cell $\mathcal{C}(x)$ such that $D_{M_0}(x) \subset \mathcal{C}(x) \subset D_{2M_0}(x)$. Therefore, given two different nuclei $x, x' \in \mathcal{D}(X_{nR})$, we have $D_{M_0}(x) \cap D_{M_0}(x') = \varnothing$. We choose these cells such that they are μ_{nR} measurable and such that they tile their respective horospheres:

$$\forall n \in \mathbb{Z}, \quad \bigsqcup_{x \in \mathcal{D}(X_{nR})} \mathcal{C}(x) = X_{nR}.$$

As an example, one can take Voronoi cells

$$\mathcal{VC}(x) := \{p \in X_{nR} \mid d(p,x) \leq d(p,x') \text{ for all } x' \in \mathcal{D}(X_{nR})\}.$$

These cells might not be disjoint, but a point $p \in X_{nR}$ is contained in a finite number of Voronoi cells since X is proper. Therefore, by choosing (for example thanks to an arbitrary order on $\mathcal{D}(X_{nR})$) a unique cell containing p, and removing p from the others, there exists a tiling of X_{nR} by cells $\mathcal{C}(x)$.

Now, for all $n \in \mathbb{Z}$ and for all $x \in \mathcal{D}(X_{nR})$ we define the box $\mathcal{B}(x, R)$ at scale R of nucleus x by

$$\mathcal{B}(x,R) := \bigcup_{z \in [(n-1)R; nR[} \pi_z(\mathcal{C}(x)).$$

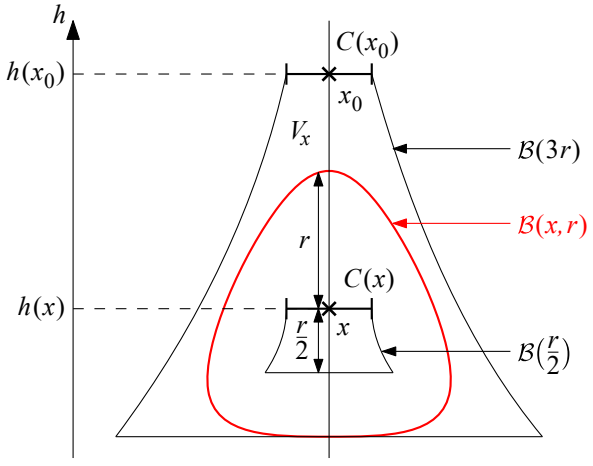


Figure 14: Proof of Lemma 3.10.

In this definition, we chose $[(n-1)R; nR[$ for the boxes' heights. It is an arbitrary choice, one could prefer to use $](n-1)R; nR]$ as these heights intervals. Moreover, to construct the horospherical product of X and Y, we will use intervals of the form $[\dots;\dots[$ for X and $]\dots;\dots]$ for Y.

We recall that the cells $\mathcal{C}(x)$ tile the horospheres X_{nR}. Furthermore there exists a unique vertical geodesic ray leaving each point of X. Consequently we have a box tiling of X, at scale R,

$$X = \bigsqcup_{n\in\mathbb{Z}} \bigsqcup_{x\in\mathcal{D}(X_{nR})} \mathcal{B}(x, R). \tag{49}$$

The next lemma explains that any box contains and is contained in metric balls of similar scales.

Lemma 3.10 *There exists a constant $M(X)$ such that for all $x \in X$ and $r > M$ there exist two boxes $\mathcal{B}\left(\frac{r}{2}\right)$ and $\mathcal{B}(3r)$ satisfying*

$$\mathcal{B}\left(\frac{r}{2}\right) \subset B(x, r) \subset \mathcal{B}(3r).$$

The proof is illustrated in Figure 14.

Proof Let $\mathcal{C}(x)$ be a subset of $X_{h(x)}$ containing $D(x, M_0)$ and contained in $D(x, 2M_0)$. Let us denote by $\mathcal{B}\left(\frac{r}{2}\right)$ the box at scale $\frac{r}{2}$ constructed from the cell $\mathcal{C}(x)$. For all $x' \in \mathcal{B}\left(\frac{r}{2}\right)$ let us denote by $x'' := V_{x'}(h(x))$ the point of $V_{x'}$ at the height $h(x)$, we have

$$d_X(x', x) \leq d_X(x', x'') + d_X(x'', x) \leq \frac{r}{2} + 2M_0 \leq r \quad \text{for } r \geq 4M_0,$$

which gives us that $x' \in B(x, r)$. To prove the second inclusion, let us denote by V_x the unique (since X is Busemann convex) vertical geodesic ray leaving x. Let $x_0 \in \mathrm{im}(V_x)$ such that $h(x_0) = h(x) + 2r$ and $\mathcal{C}(x_0)$ be a subset of $X_{h(x_0)}$ containing $D(x_0, M)$ and contained in $D(x_0, 2M)$. Then we claim that

$B(x,r)$ is included in the box at scale $3r$ constructed from the cell $\mathcal{C}(x_0)$. Let $x' \in B(x,r)$, we recall that $d_r(x',x) := d_X(x',x) - \Delta h(x',x)$. By Lemma 1.5 we have that

$$d(V_x(h(x)+2r), V_{x'}(h(x)+2r)) \leq 96\delta = M$$

since $r \geq d_X(x',x) \geq \frac{1}{2} d_r(x',x)$ and since the distance between two vertical geodesics is decreasing in the upward direction. Therefore $V_{x'}(h(x)+2r) \in \mathcal{C}(x_0)$. Furthermore

$$\Delta h(x_0,x') \leq \Delta h(x_0,x) + \Delta h(x,x') \leq 3r,$$

hence $x' \in \mathcal{B}(3r)$. □

3.3 Tiling a big box by small boxes

Let $R > 0$ and $N \in \mathbb{N}$. Our next result shows that a box at scale NR can be tiled with boxes at scale R.

Proposition 3.11 *Let M_0 be the constant of assumption* (E2)*. Let $R > 0$ and $N \in \mathbb{N}$. Let $\mathcal{B}^X$ be a box at scale NR, and let us denote by $h^- := h^-(\mathcal{B}^X)$ the lowest height of $\mathcal{B}^X$. Then there exists a box tiling at scale R of $\mathcal{B}^X$. Otherwise stated for all $k \in \{1,\dots,N\}$ there exists $\mathcal{D}_k(\mathcal{B}^X) \subset \mathcal{B}^X_{h^-+kR}$ such that*

(1) *for all $x \in \mathcal{D}_k(\mathcal{B}^X)$, there exists a cell $\mathcal{C}(x)$ such that $D_{M_0}(x) \subset \mathcal{C}(x) \subset D_{3M_0}(x)$;*

(2) *we have $\bigsqcup_{k=1}^{N} \bigsqcup_{x\in\mathcal{D}_k(\mathcal{B}^X)} \mathcal{B}^X(x,\mathcal{C}(x),R) = \mathcal{B}^X$.*

Proof To tile the box $\mathcal{B}^X$ we first tile by cells all of its level sets at height $h^- + kR$. Let $k \in \{1,\dots,N\}$, and let $\mathcal{D}_k(\mathcal{B}^X)$ be an $2M_0$-maximal separating set of $\mathrm{Int}_{M_0}(\mathcal{B}^X_{h^-+kR})$. Then

(1) for all $x, x' \in \mathcal{D}_k(\mathcal{B}^X)$ with $x \neq x'$ we have $D_{M_0}(x) \cap D_{M_0}(x') = \varnothing$;

(2) $\mathrm{Int}_{M_0}(\mathcal{B}^X_{h^-+kR}) \subset \bigcup_{x\in\mathcal{D}_k(\mathcal{B}^X)} D_{2M_0}(x)$.

Furthermore $\mathcal{N}_{M_0}(\mathrm{Int}_{M_0}(\mathcal{B}^X_{h^-+kR})) \subset \mathcal{B}^X_{h^-+kR}$, and for all $x \in \mathrm{Int}_{M_0}(\mathcal{B}^X_{h^-+kR})$ we have the inclusion $D_{M_0}(x) \subset \mathcal{B}^X_{h^-+kR}$. Therefore

$$\bigsqcup_{x\in\mathcal{D}_k(\mathcal{B}^X)} D_{M_0}(x) \subset \mathcal{B}^X_{h^-+kR} \subset \bigcup_{x\in\mathcal{D}_k(\mathcal{B}^X)} D_{3M_0}(x). \tag{50}$$

For all $x \in \mathcal{D}_k(\mathcal{B}^X)$, we define

$$\mathcal{C}(x) := \{p \in \mathcal{B}^X_{h^-+kR} \mid d(p,x) \leq d(p,x') \text{ for all } x' \in \mathcal{D}_k(\mathcal{B}^X)\}.$$

As discussed at the beginning of Section 3.2, these cells might intersect each other on their boundaries. However, a point contained in different cells can be removed in all of them except one, making them disjoint. The choice of cells on which we remove boundary points can be made thanks to an arbitrary order on the finite set $\mathcal{D}_k(\mathcal{B}^X)$.

By the inclusions (50), for all $x \in \mathcal{D}_k(\mathcal{B}^X)$ we have $D_{M_0}(x) \subset \mathcal{C}(x) \subset D_{3M_0}(x)$ and

$$\bigsqcup_{x\in\mathcal{D}_k(\mathcal{B}^X)} \mathcal{C}(x) = \mathcal{B}^X_{h^-+kR}.$$

Furthermore, since vertical geodesic rays are uniquely determined by their starting point (because X is Busemann), a tiling with cells provides us with a box tiling

$$\bigsqcup_{x\in\mathcal{D}_k(\mathcal{B}^X)} \mathcal{B}^X(x,\mathcal{C}(x),R) = \bigcup_{z\in[h^-+(k-1)R;h^-+kR[} \mathcal{B}^X_z.$$

Taking the union on $k\in\{1,\dots,N\}$ provides us with the conclusion. □

3.4 Box-tiling of $X\bowtie Y$

The boxes $\mathcal{B}$ of a horospherical product $X\bowtie Y$ are constructed as the horospherical products of boxes $\mathcal{B}^X\bowtie\mathcal{B}^Y$. Therefore they induce a tiling of $X\bowtie Y$. Such boxes are illustrated by Figure 15.

Definition 3.12 (box of $X\bowtie Y$ at scale R) Let X and Y be two admissible spaces. A set $\mathcal{B}\subset X\bowtie Y$ is called a box at scale R of $X\bowtie Y$ if there exists $\mathcal{B}^X$ a box at scale R of X and $\mathcal{B}^Y$ a box at scale R of Y such that

(1) $h^-(\mathcal{B}^X)=-h^+(\mathcal{B}^Y)$;

(2) $\mathcal{B}:=\mathcal{B}^X\bowtie\mathcal{B}^Y=\{(x,y)\in\mathcal{B}^X\times\mathcal{B}^Y \mid h_X(x)=-h_Y(y)\}$.

Let us point out that in the last definition, the box of Y is in fact defined by

$$\mathcal{B}^Y(y,R) := \bigcup_{z\in]-nR;(1-n)R]} \pi_z(\mathcal{C}(y)). \tag{51}$$

This choice on the boundaries of the height intervals allows a precise match for the height inside the two boxes. Furthermore, one can see that given a box-tiling of X and a box-tiling of Y, the natural subsequent tiling on $X\times Y$ provides the box tiling of $X\bowtie Y$ by restriction.

Proposition 3.13 (box-tiling of $X\bowtie Y$ at scale R) *Let X and Y be two admissible spaces. Let R be a positive number and let us consider the two box tilings, of X and Y,*

$$X=\bigsqcup_{n\in\mathbb{Z}}\bigsqcup_{x\in\mathcal{D}(X_{nR})}\mathcal{B}^X(x,R);\quad Y=\bigsqcup_{n\in\mathbb{Z}}\bigsqcup_{y\in\mathcal{D}(Y_{nR})}\mathcal{B}^Y(y,R).$$

Then the boxes of $X\bowtie Y$ constructed from boxes at opposite height in X and Y are a box tiling of $X\bowtie Y$. We have

$$X\bowtie Y=\bigsqcup_{n\in\mathbb{Z}}\bigsqcup_{(x,y)\in\mathcal{D}(X_{nR})\times\mathcal{D}(Y_{(1-n)R})}\mathcal{B}^X(x,R)\bowtie\mathcal{B}^Y(y,R).$$

Proof Let us consider the box tilings, of X and Y,

$$X=\bigsqcup_{n\in\mathbb{Z}}\bigsqcup_{x\in\mathcal{D}(X_{nR})}\mathcal{B}^X(x,R);\quad Y=\bigsqcup_{n\in\mathbb{Z}}\bigsqcup_{y\in\mathcal{D}(Y_{nR})}\mathcal{B}^Y(y,R).$$

We first show that the intersection of two distinct boxes is empty. Let $n_1,n_2\in\mathbb{R}$, $x_1\in\mathcal{D}(X_{n_1R})$, $x_2\in\mathcal{D}(X_{n_2R})$, $y_1\in\mathcal{D}(Y_{(1-n_1)R})$ and $y_2\in\mathcal{D}(X_{(1-n_2)R})$ such that $(x_1,y_1)\neq(x_2,y_2)$. Then we have

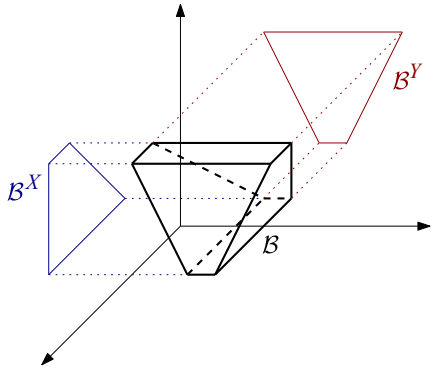


Figure 15: Box in $X \bowtie Y$.

either $x_1 \neq x_2$ or $y_1 \neq y_2$. Let us consider the case $x_1 \neq x_2$, then $\mathcal{B}^X(x_1, R) \neq \mathcal{B}^X(x_2, R)$, and since they are two tiles of the box tiling of X, we have $\mathcal{B}^X(x_1, R) \cap \mathcal{B}^X(x_2, R) = \varnothing$. Therefore

$$\forall (p_1^X, p_1^Y) \in \mathcal{B}^X(x_1, R) \bowtie \mathcal{B}^Y(y_1, R), \quad \forall (p_2^X, p_2^Y) \in \mathcal{B}^X(x_2, R) \bowtie \mathcal{B}^Y(y_2, R), \quad p_1^X \neq p_2^X.$$

Hence $(p_1^X, p_1^Y) \neq (p_2^X, p_2^Y)$, which gives us

$$(\mathcal{B}^X(x_1, R) \bowtie \mathcal{B}^Y(y_1, R)) \cap (\mathcal{B}^X(x_2, R) \bowtie \mathcal{B}^Y(y_2, R)) = \varnothing.$$

The case when $y_1 \neq y_2$ provide us with the same conclusion. Then we prove that the whole space $X \bowtie Y$ is covered by the horospherical product of boxes. Let $p = (p^X, p^Y) \in X \bowtie Y$. There exists $n \in \mathbb{Z}$ such that $(n-1)R \leq h(p) < nR$, hence there exist $x \in \mathcal{D}(X_{nR})$ and $y \in \mathcal{D}(Y_{(1-n)R})$ such that $p^X \in \mathcal{B}^X(x, R)$ and $p^Y \in \mathcal{B}^Y(y, R)$. Therefore $p \in \mathcal{B}^X(x, R) \bowtie \mathcal{B}^Y(y, R)$. □

3.5 Measure of balls, boxes and neighbourhoods

The results of this section focus on estimates on the measure μ of balls and boxes.

Lemma 3.14 *There exists $M(\bowtie)$ such that for all $r \geq M$ and all box $\mathcal{B}$ at scale r of $X \bowtie Y$ we have*

$$\mu(\mathcal{B}) \asymp_{\bowtie} e^{mr}. \tag{52}$$

Proof Without loss of generality we can assume that $h(\mathcal{B}) = [0; r[$. Let us denote by $\mathcal{C}^X$ the cell of $\mathcal{B}^X$ and $\mathcal{C}^Y$ the cell of $\mathcal{B}^Y$. Then

$$\begin{aligned}
\mu(\mathcal{B}) &= \int_0^r \mu_z(\mathcal{B}_z)\,\mathrm{d}z = \int_0^r \mu_z^X(\mathcal{B}_z^X)\mu_{-z}^Y(\mathcal{B}_{-z}^Y)\,\mathrm{d}z \quad (\text{since } \mathcal{B}_z = \mathcal{B}_z^X \times \mathcal{B}_{-z}^Y) \\
&\asymp_{\bowtie} \int_0^r e^{m(r-z)}\mu_r^X(\mathcal{C}^X)e^{nz}\mu_0^Y(\mathcal{C}^Y)\,\mathrm{d}z \quad (\text{by assumption (E3) and definition of boxes}) \\
&\asymp_{\bowtie} e^{mr}\int_0^r e^{(n-m)z}\,\mathrm{d}z \quad (\text{by Lemma 3.2}) \\
&= \frac{e^{mr} - e^{nr}}{m-n} \preceq_{\bowtie} e^{mr}.
\end{aligned}$$

However $m > n$, hence for $r \geq \frac{1}{m-n}$ we have $\frac{1}{2}e^{mr} \geq e^{nr}$. Therefore

$$\frac{e^{mr} - e^{nr}}{m-n} \geq \frac{e^{mr}}{2(m-n)} \succeq_{\bowtie} e^{mr}. \qquad \square$$

Combining Lemmas 3.10 and 3.14 we get the next corollary.

Corollary 3.15 *There exists $M(\bowtie)$ such that for any $r \geq M$ and any $p \in X \bowtie Y$ we have*

$$e^{\frac{m}{2}r} \preceq_{\bowtie} \mu(B(p,r)) \preceq_{\bowtie} e^{3mr}. \tag{53}$$

Therefore we have the following estimate between ball measures, which corresponds to [8, Lemma 5.2].

Corollary 3.16 *There exists $M(\bowtie)$ such that, for any $r_2 > 2r_1 \geq M$ and for all $p_1, p_2 \in X \bowtie Y$,*

$$\exp\left(\tfrac{1}{6}|r_2 - r_1|m\right)\mu(B(p_1,r_1)) \leq \mu(B(p_2,r_2)) \leq \exp(6|r_2 - r_1|m)\mu(B(p_1,r_1)).$$

Corollary 3.17 *There exists $M(\bowtie)$ such that, for any $r_2 > r_1 \geq M$ and for all $A \subset X \bowtie Y$,*

$$\mu(\mathcal{N}_{r_2}(A)) \preceq_{\bowtie} e^{6|r_2 - r_1|m}\mu(\mathcal{N}_{r_1}(A)).$$

Furthermore, if there exists $z \in \mathbb{R}$ such that $A \subset X_z$ we have

$$\mu(\mathcal{N}_M(A)) \succeq_{\bowtie} \mu_z(\mathcal{N}_M(A) \cap X_z).$$

In particular, for all $p \in (X \bowtie Y)_z$,

$$\mu(B(p,M)) \succeq_{\bowtie} \mu_z(D_M(p)).$$

Proof Since $X \bowtie Y$ is a proper metric space. By a covering lemma of [15], there exists a set $Z \subset A$ such that

(1) the balls $B(p,r_1)$ for $p \in Z$ are pairwise disjoint;

(2) we have the inclusions

$$\bigsqcup_{p\in Z} B(p,r_1) \subset \mathcal{N}_{r_1}(A) \subset \bigcup_{p\in Z} B(p,5r_1).$$

Therefore $\mathcal{N}_{r_2}(A) \subset \bigcup_{p\in Z} B(p, 5r_1 + (r_2 - r_1))$.

Moreover, if $A \subset X_z$, for $r_1 = M$ we have

$$\bigsqcup_{p\in Z} D_M(p) \subset \mathcal{N}_M(A) \cap X_z \subset \bigcup_{p\in Z} D_{5M}(p),$$

and for all $p \in Z$, $\mu_z(B(p,5M)) \succeq_{\bowtie} 1 \succeq_{\bowtie} \mu_z(D_{5M}(p))$. Hence

$$\begin{aligned}\mu(\mathcal{N}_M(A)) &\succeq_{\bowtie} \sum_{p\in Z} \mu(B(p,5M)) \\ &\succeq_{\bowtie} \sum_{p\in Z} \mu_z(D_{5M}(p)) \succeq_{\bowtie} \mu_z(\mathcal{N}_M(A) \cap X_z).\end{aligned} \qquad \square$$

A (k,c)-quasi-isometry $\Phi : X \bowtie Y \to X' \bowtie Y'$ "quasi"-preserves the measure μ.

Lemma 3.18 *For all (k,c)-quasi-isometry $\Phi : X \bowtie Y \to X' \bowtie Y'$ and for all measurable subsets $U \subset X \bowtie Y$ we have*

$$\mu(\mathcal{N}_{k(c+1)}(U)) \asymp_{k,c,\bowtie} \mu\big(\mathcal{N}_1(\Phi(U))\big).$$

Proof Since $X \bowtie Y$ is a proper metric space, by a classical covering lemma of [15] there exists a set $Z \subset U$ such that

(1) the balls $B(p, k(c+1))$ for $p \in Z$ are pairwise disjoint;

(2) we have the inclusions

$$\bigsqcup_{p\in Z} B(p, k(c+1)) \subset \mathcal{N}_{k(c+1)}(U) \subset \bigcup_{p\in Z} B(p, 5k(c+1)).$$

Since Φ is a (k,c)-quasi-isometry,

(1) the balls $B(q,1)$ for $q \in \Phi(Z)$ are pairwise disjoint;

(2) we have the inclusions

$$\bigsqcup_{q\in\Phi(Z)} B(q,1) \subset \mathcal{N}_1(\Phi(U)) \subset \bigcup_{q\in\Phi(Z)} B(q, 5k^2(c+1)+c).$$

Furthermore, for all $p \in Z$ we have

$$\mu(B(p,1)) \asymp_{\bowtie} 1 \asymp_{\bowtie'} \mu\big(B(\Phi(p),1)\big) \asymp_{k,c,\bowtie'} \mu\big(B(\Phi(p), 5k^2(c+1)+c)\big),$$

hence $\mu(\mathcal{N}_{k(c+1)}(U)) \asymp_{k,c,\bowtie} \#Z \asymp_{k,c,\bowtie'} \mu\big(\mathcal{N}_1(\Phi(U))\big)$. □

3.6 Set of vertical geodesics

Since X is a Gromov hyperbolic, Busemann space, for any $x \in X$, there exists a unique vertical geodesic ray starting from x in X, therefore, there is a one-to-one correspondence between portions of vertical geodesic rays in a box $\mathcal{B}^X$, and the points at the bottom of $\mathcal{B}^X$. A vertical geodesic segment of $\mathcal{B}^X$ is defined as the intersection of a vertical geodesic and $\mathcal{B}^X$. We recall that vertical geodesics are parametrised by arclength by their height.

Let $\mathcal{B}^X$ be a box at scale R of X. Let us denote by $V\mathcal{B}^X$ the set of vertical geodesic segments of $\mathcal{B}$. A geodesic segment $v \in V\mathcal{B}^X$ intersects only in one point x the bottom of $\mathcal{B}^X$, and v is the only vertical geodesic segment of $V\mathcal{B}^X$ intersecting x by the Busemann assumption on X.

Definition 3.19 (measure η on $V\mathcal{B}^X$) Let $\mathcal{B}^X$ be a box at scale R of X. The measure $\eta^X_{V\mathcal{B}^X}$ on $V\mathcal{B}^X$ is defined on all measurable subset $U \subset V\mathcal{B}^X$ by

$$\eta^X_{V\mathcal{B}^X}(U) = \lambda^X_{h^-(\mathcal{B}^X)}\big(\{\gamma(h^-(\mathcal{B}^X)) \mid \gamma \in U\}\big). \tag{54}$$

In particular, we say that U is measurable if $\{\gamma(h^-(\mathcal{B}^X)) \mid \gamma \in U\}$ is measurable. Since the measure λ is almost constant along projections, the measure on the set of vertical geodesic segment is related to the height of the boxes. Specifically we show that up to a multiplicative constant, the measure of a box is equal to the measure of its set of vertical geodesic segments multiplied by its height, as for rectangles in $\mathbb{R}^2$. In the sequel we might omit the index of the measure η^X.

Property 3.20 Let $\mathcal{B}^X$ be a box at scale R of X and let us define $h^- := h^-(\mathcal{B}^X)$ and $h^+ := h^+(\mathcal{B}^X)$. We have, for all $z \in [h^-, h^+[$,

(1) $\eta^X(V\mathcal{B}^X) \asymp_X \lambda_z^X(\mathcal{B}_z^X) \asymp_X e^{mh^+}$;

(2) $\lambda^X(\mathcal{B}^X) \asymp_X R\lambda_z^X(\mathcal{B}_z^X) \asymp_X R\eta^X(V\mathcal{B}^X) \asymp_X Re^{mh^+}$.

Proof Let $x \in X$ be such that $\mathcal{C}(x)$ is the cell of $\mathcal{B}^X$. We know that $D_{M_0}(x) \subset \mathcal{C}(x) \subset D_{2M_0}(x)$, hence by Lemma 3.2 we have

$$\mu^X_{h(x)}(\mathcal{C}(x)) \asymp_X 1. \tag{55}$$

Then

$$\begin{aligned}\eta^X(V\mathcal{B}^X) &= \lambda^X_{h^-}(\mathcal{B}^X \cap h^{-1}(h^-)) \quad \text{(by definition)}\\ &\asymp_X \lambda_z^X(\mathcal{B}_z^X) \asymp_X \lambda^X_{h^+}(\mathcal{C}(x)) \asymp_X e^{mh^+}\mu^X_{h^+}(\mathcal{C}(x)) \quad \text{(by Property 3.5)}\\ &\asymp_X e^{mh^+},\end{aligned}$$

which proves the first point. The second point follows from the fact that the measures λ_z are constant by projections on height level sets, up to the multiplicative constant $M(X)$:

$$\begin{aligned}\lambda^X(\mathcal{B}^X) &= \int_{h^-}^{h^+} \lambda_z^X(\mathcal{B}^X \cap h^{-1}(z))\,\mathrm{d}z = \int_{h^-}^{h^+} \lambda_z^X\big(\pi_z(\mathcal{C}(x))\big)\,\mathrm{d}z\\ &\asymp_X \int_{h^-}^{h^+} \lambda^X_{h^+}(\mathcal{C}(x))\,\mathrm{d}z \quad \text{(by Property 3.5)}\\ &\asymp_X R\lambda^X_{h^+}(\mathcal{C}(x)) \asymp_X Re^{mh^+}.\end{aligned}$$

□

A vertical geodesic $V = (V^X, V^Y) \subset X \bowtie Y$ is a couple of vertical geodesics of X and Y. Therefore, there is a bijection between the set of vertical geodesic segments $V\mathcal{B}$ of a box $\mathcal{B} := \mathcal{B}^X \bowtie \mathcal{B}^Y$ and $V\mathcal{B}^X \times V\mathcal{B}^Y$.

Definition 3.21 Let $\mathcal{B}$ be a box at scale R of $X \bowtie Y$. We define the measure $\eta_{V\mathcal{B}}$ on $V\mathcal{B}$ as

$$\eta_{V\mathcal{B}} := \eta^X_{V\mathcal{B}^X} \otimes \eta^Y_{V\mathcal{B}^Y}. \tag{56}$$

In the notation of measures on sets of vertical geodesic segments, we might omit the reference to the corresponding sets. The measures $\eta_{V\mathcal{B}}$, respectively $\eta^X_{V\mathcal{B}^X}$, $\eta^Y_{V\mathcal{B}^Y}$, will simply be denoted by η, respectively η^X, η^Y.

Proposition 3.22 *For each box $\mathcal{B}$ at scale R of $X \bowtie Y$ we have, for all $z_1, z_2 \in [h^-, h^+[$,*

(1) $\eta(V\mathcal{B}) \asymp_{\bowtie} e^{mh^+} e^{-nh^-} \asymp_{\bowtie} \lambda^X_{z_1}(\mathcal{B}^X_{z_1})\lambda^Y_{-z_2}(\mathcal{B}^Y_{-z_2})$;

(2) $\lambda(\mathcal{B}) \asymp_{\bowtie} R\eta(V\mathcal{B}) \asymp_{\bowtie} R\lambda^X_{z_1}(\mathcal{B}^X_{z_1})\lambda^Y_{-z_2}(\mathcal{B}^Y_{-z_2})$.

Proof The first point follows from Definition 3.21 and Property 3.20 applied on $\mathcal{B}^X$ and $\mathcal{B}^Y$. The proof of the second point is similar to the proof of Property 3.20:

$$\begin{aligned}
\lambda(\mathcal{B}) &= \int_{h^-}^{h^+} \lambda^X_z \otimes \lambda^Y_{-z}(\mathcal{B}^X_z \times \mathcal{B}^Y_{-z})\,\mathrm{d}z = \int_{h^-}^{h^+} \lambda^X_z(\mathcal{B}^X_z)\lambda^Y_{-z}(\mathcal{B}^Y_{-z})\,\mathrm{d}z \\
&\asymp_{\bowtie} \int_{h^-}^{h^+} \lambda^X_{h^-}(\mathcal{B}^X_{h^-})\lambda^Y_{-h^+}(\mathcal{B}^Y_{-h^+})\,\mathrm{d}z \quad \text{(by Property 3.5)} \\
&\asymp_{\bowtie} \int_{h^-}^{h^+} \eta^X(V\mathcal{B}^X)\eta^Y(V\mathcal{B}^Y)\,\mathrm{d}z \quad \text{(by definition of } \eta) \\
&\asymp_{\bowtie} \eta(V\mathcal{B}) \int_{h^-}^{h^+} 1\,\mathrm{d}z = R\eta(V\mathcal{B}).
\end{aligned}$$

Then applying twice Property 3.20 provides us with the result. □

Let $\mathcal{B}$ be a box at scale R. Let $z \in [h^-(\mathcal{B}); h^+(\mathcal{B})[$ and let $U \subset \mathcal{B}_z$. Then we denote by $V_{\mathcal{B}}(U)$ the set of vertical geodesic segments of $V\mathcal{B}$ intersecting U. It is in bijection with

$$\{(x, y) \in \mathcal{B}^X_0 \times \mathcal{B}^Y_{-R} \mid (\pi^X_z(x), \pi^Y_{-z}(y)) \in U\}.$$

We need the following property stating that the measure of a given subfamily of vertical geodesics can be computed on any level of our box.

Property 3.23 Let $\mathcal{B}$ be a box at scale R of $X \bowtie Y$. Then, for all $z \in [h^-(\mathcal{B}); h^+(\mathcal{B})[$ and for all measurable subsets $U_z \subset \mathcal{B}_z$,

$$\eta(V_{\mathcal{B}}(U_z)) \asymp_{\bowtie} \lambda_z(U_z).$$

Proof Without loss of generality we can assume that $[h^-(\mathcal{B}); h^+(\mathcal{B})[= [0 : R[$. By definition we have

$$\begin{aligned}
&\eta(V_{\mathcal{B}}(U_z)) \\
&:= \int_{x_0 \in \mathcal{B}^X_0} \int_{y_0 \in \mathcal{B}^Y_{-R}} \mathbb{1}_{\{(x,y) \in \mathcal{B}^X_0 \times \mathcal{B}^Y_{-R} \mid (\pi^X_z(x), \pi^X_{-z}(y)) \in U_z\}}(x_0, y_0)\, d\lambda^Y_{-R} d\lambda^X_0 \\
&= \int_{x_0 \in \mathcal{B}^X_0} \int_{y_0 \in \mathcal{B}^Y_{-R}} \mathbb{1}_{U_z}(\pi^X_z(x_0), \pi^Y_{-z}(y_0))\, d\lambda^Y_{-R} d\lambda^X_0 \\
&= \int_{x_0 \in \mathcal{B}^X_0} \Big(\int_{y \in \mathcal{B}^Y_{-z}} \mathbb{1}_{U_z}(\pi^X_z(x_0), y)\, d((\pi^Y_{-z})_* \lambda^Y_{-R}) \Big) d\lambda^X_0 \quad \text{(with a pushforward of } \lambda^Y_{-R} \text{ by } \pi^Y_{-z}) \\
&= \int_{y \in \mathcal{B}^Y_{-z}} \Big(\int_{x_0 \in \mathcal{B}^X_0} \mathbb{1}_{U_z}(\pi^X_z(x_0), y)\, d\lambda^X_0 \Big) d((\pi^Y_{-z})_* \lambda^Y_{-R}) \quad \text{(by Fubini's theorem)} \\
&= \int_{y \in \mathcal{B}^Y_{-z}} \Big(\int_{x \in \mathcal{B}^X_z} \mathbb{1}_{U_z}(x, y)\, d((\pi^Y_z)_* \lambda^X_0) \Big) d((\pi^Y_{-z})_* \lambda^Y_{-R}) \quad \text{(with a pushforward of } \lambda^X_0 \text{ by } \pi^X_z) \\
&\asymp_{\bowtie} \int_{y \in \mathcal{B}^Y_{-z}} \int_{x \in \mathcal{B}^X_z} \mathbb{1}_{U_z}(x, y)\, d\lambda^X_z d\lambda^Y_{-z} \quad \text{(by using Property 3.5 twice)} \\
&\asymp_{\bowtie} \lambda_z(U_z).
\end{aligned}$$

□

3.7 Projections of set of almost full measure

Let us denote by $p^X : X \bowtie Y \to X$, $(x, y) \mapsto x$, and by $p^Y : X \bowtie Y \to X$, $(x, y) \mapsto y$, the projections on the two coordinates of $X \bowtie Y$. We also denote by slight abuse the projection on a set of vertical geodesic segments $p^X : V\mathcal{B} \to V\mathcal{B}^X$, $(v^X, v^Y) \mapsto v^X$, and $p^X : V\mathcal{B} \to V\mathcal{B}^Y$, $(v^X, v^Y) \mapsto v^Y$. Given a subset $U \subset \mathcal{B}$, we might simply denote by U^X, respectively U^Y, its projection on X, respectively on Y, and similarly for subsets of $V\mathcal{B}$.

In this section, we show that if a subset of a box has almost full measure, then most of the fibres with respect to these projections also have almost full measure.

Let $0 < \alpha \leq 1$, let $V_1 \subset V\mathcal{B}$ be a measurable subset (it will be chosen later as a subset of small measure, containing "bad" vertical geodesics). Let us define, for all $v^X \in V\mathcal{B}^X$,

$$G^Y(v^X) := \{v^Y \in V\mathcal{B}^Y \mid (v^X, v^Y) \in V_0\} = p^Y((p^X)^{-1}(v^X) \cap (V\mathcal{B} \setminus V_1)),$$
$$G^X := \big\{v^X \in V\mathcal{B}^X \mid \eta^Y(G^Y(v^X)) \geq \big(1 - \sqrt{\alpha}\big)\eta^Y(V_1^Y)\big\}.$$

The set G^X is the set of vertical geodesics in $V\mathcal{B}^X$ whose fibres have almost full intersection with $V\mathcal{B} \setminus V_1$.

The following lemma asserts that almost all fibres have almost full intersection with $V\mathcal{B} \setminus V_1$.

Lemma 3.24 *Let $0 < \alpha \leq 1$ and let $V_1 \subset V\mathcal{B}$ be a measurable subset such that $\eta(V_1) \leq \alpha\eta(V\mathcal{B})$. Then*

$$\eta^X(G^X) \geq \big(1 - \sqrt{\alpha}\big)\eta^X(V\mathcal{B}^X).$$

Proof By construction we have

$$\bigcup_{v^X \in V\mathcal{B}^X} G^Y(v^X) = (V\mathcal{B} \setminus V_1)^Y.$$

To prove the lemma we proceed by contradiction. Let us assume that $\eta^X(G^X) < \big(1 - \sqrt{\alpha}\big)\eta^X(V\mathcal{B}^X)$. Then

$$\eta^X(V\mathcal{B}^X \setminus G^X) > \sqrt{\alpha}\eta^X(V\mathcal{B}^X).$$

Therefore

$$\begin{aligned}
\eta(V_1) &= \int_{V\mathcal{B}} \mathbb{1}_{V_1}(v)\,\mathrm{d}\eta(v) \\
&= \int_{V\mathcal{B}^X}\int_{V\mathcal{B}^Y} \mathbb{1}_{V_1}(v^X, v^Y)\,\mathrm{d}\eta^Y(v^Y)\,\mathrm{d}\eta^X(v^X) \quad \text{(by definition of } \eta) \\
&= \int_{V\mathcal{B}^X}\int_{V\mathcal{B}^Y} \mathbb{1}_{V\mathcal{B}^Y \setminus G^Y(v^X)}(v^Y)\,\mathrm{d}\eta^Y(v^Y)\,\mathrm{d}\eta^X(v^X) \quad \text{(by definition of } G^Y(v^X)) \\
&= \int_{V\mathcal{B}^X} \eta^Y(V\mathcal{B}^Y \setminus G^Y(v^X))\,\mathrm{d}\eta^X(v^X) \\
&\geq \int_{V\mathcal{B}^X \setminus G^X} \eta^Y(V\mathcal{B}^Y \setminus G^Y(v^X))\,\mathrm{d}\eta^X(v^X).
\end{aligned}$$

Furthermore, when $v^X \in V\mathcal{B}^X \setminus G^X$ we have that $\eta^Y(G^Y(v^X)) < \big(1-\sqrt{\alpha}\big)\eta^Y(V\mathcal{B}^Y)$, and hence $\eta^Y(V\mathcal{B}^Y \setminus G^Y(v^X)) \geq \sqrt{\alpha}\eta^Y(V\mathcal{B}^Y)$. Therefore

$$\begin{aligned}\eta(V_1) &\geq \int_{V\mathcal{B}^X\setminus G^X} \sqrt{\alpha}\eta^Y(V\mathcal{B}^Y)\,\mathrm{d}\eta^X(v^X)\\ &\geq \sqrt{\alpha}\eta^Y(V\mathcal{B}^Y)\eta^X(V\mathcal{B}^X\setminus G^X)\\ &\geq \sqrt{\alpha}\sqrt{\alpha}\eta^Y(V\mathcal{B}^Y)\eta^X(V\mathcal{B}^X) \quad \text{(by the contradiction assumption)}\\ &> \alpha\eta(V\mathcal{B}) \quad \text{(since } V\mathcal{B} \text{ is a product)},\end{aligned}$$

which contradicts $\eta(V_1) \leq \alpha\eta(V\mathcal{B})$. □

In the previous lemma we only used the fact that the set of vertical geodesic segments $V\mathcal{B}$ was the product of its projections endowed with a product measure η. We will use it once again on the product of two measured spaces endowed with a product measure in the proof of Proposition 4.7.

We recall that for any $U \subset X \bowtie Y$ we define $V\mathcal{B}(U) := \{v \in V\mathcal{B} \mid \mathrm{im}(v) \cap U \neq \varnothing\}$. Similarly for all $V_1 \subset V\mathcal{B}$ we define $V_1(U) := \{v \in V_1 \mid \mathrm{im}(v) \cap U \neq \varnothing\}$.

The next lemma is a local version of Lemma 3.24. Let $V_1 \subset V\mathcal{B}$. Let $M > 0$ be a constant, let $a \in \mathcal{B}$ and let us define $VD := V\mathcal{B}(D_M(a))$ and $V_1D := V_1(D_M(a))$. For all $v = (v^X, v^Y) \in V_{\mathcal{B}}$, let us define

$$\begin{aligned}E^Y(v^X) &:= \{v^Y \in VD^Y \mid (v^X, v^Y) \in V_1D\} = (p^Y)^{-1}(p^X(v^X) \cap V_1D);\\ F^X &:= \big\{v^X \in VD^X \mid \eta^Y(E^Y(v^X)) \geq \sqrt{\alpha}\eta^Y(VD^Y)\big\}.\end{aligned}$$

Lemma 3.25 *Let $0 < \alpha \leq 1$. If $\eta(V_1D) \leq \alpha\eta(VD)$ then*

$$\eta^X(F^X) \leq \sqrt{\alpha}\eta^X(VD^X). \tag{57}$$

Proof Let us proceed by contradiction. We assume that

$$\eta^X(F_i^X) > \sqrt{\alpha}\eta^X(VD_i^X). \tag{58}$$

Then we have

$$\begin{aligned}\eta(V_1D) &= \int_{v^X\in VD^X}\int_{v^Y\in VD^Y} \mathbb{1}_{V_1D}(v^X, v^Y)\,\mathrm{d}\eta^Y\,\mathrm{d}\eta^X\\ &= \int_{v^X\in VD^X}\int_{v^Y\in VD^Y} \mathbb{1}_{E^Y(v^X)}(v^Y)\,\mathrm{d}\eta^Y\,\mathrm{d}\eta^X\\ &= \int_{v^X\in VD^X} \eta^Y(E^Y(v^X))\,\mathrm{d}\eta^X \quad \text{(by the definition of } E^Y(v^X))\\ &\geq \int_{v^X\in F^X} \eta^Y(E^Y(v^X))\,\mathrm{d}\eta^X \quad \text{(since } F^X \subset VD^X)\\ &> \sqrt{\alpha}\eta^X(VD^X)\sqrt{\alpha}\eta^Y(VD^Y) > \alpha\eta(VD),\end{aligned}$$

which contradicts assumption on VD. Hence $\eta^X(F^X) \leq \sqrt{\alpha}\eta^X(VD^X)$. □

The following lemma asserts that for almost all points of the box, almost all vertical geodesics passing through the disc $D_{M_0}(x)$ do not belong to V_1.

Lemma 3.26 *There exists a constant $0<\alpha(\bowtie)\leq 1$ such that for all $0<\alpha\leq\alpha(\bowtie)$ the following statement holds. Let M_0 be the constant involved in assumption* (E2) *and let $\mathcal{B}$ be a box at scale R. If $V_1\subset V\mathcal{B}$ satisfies $\eta(V_1)\leq\alpha\eta(V\mathcal{B})$, then*

$$\lambda\left(\left\{x\in\mathcal{B}\mid\frac{\eta\big(V_1(D_{M_0}(x))\big)}{\eta\big(V\mathcal{B}(D_{M_0}(x))\big)}>\alpha^{\frac{1}{4}}\right\}\right)\leq\alpha^{\frac{1}{4}}\lambda(\mathcal{B}). \tag{59}$$

Proof Without loss of generality we may assume that $h(\mathcal{B})=[0;R[$. Let us define

$$U=\left\{x\in\mathcal{B}\mid\frac{\eta\big(V_1(D_{M_0}(x))\big)}{\eta\big(V\mathcal{B}(D_{M_0}(x))\big)}>\alpha^{\frac{1}{4}}\right\}. \tag{60}$$

We proceed by contradiction, let us assume that $\lambda(U)>\alpha^{\frac{1}{4}}\lambda(\mathcal{B})$. In this case there exists $z\in[0;R[$ such that $\lambda_z(U_z)>\alpha^{\frac{1}{4}}\lambda_z(\mathcal{B}_z)$. Let $U_z'\subset U_z$ be a $2M_0$ maximal separating set of U_z. We have that $\bigsqcup_{x\in U_z'}D_{M_0}(x)$ is a disjoint union and that $U_z\subset\bigcup_{x\in U_z'}D_{2M_0}(x)$. Then we have

$$\begin{aligned}\lambda_z\Big(\bigsqcup_{x\in U_z'}D_{M_0}(x)\Big)&=\sum_{x\in U_z'}\lambda_z(D_{M_0}(x))=\sum_{x\in U_z'}\lambda_z(D_{2M_0}(x))\frac{\lambda_z(D_{M_0}(x))}{\lambda_z(D_{2M_0}(x))}\\&\asymp_{\bowtie}\sum_{x\in U_z'}\lambda_z(D_{2M_0}(x))\quad\text{(by Lemma 3.2)}\\&\geq\lambda_z\Big(\bigcup_{x\in U_z'}D_{2M_0}(x)\Big)\geq\lambda_z(U_z)\\&\succeq_{\bowtie}\alpha^{\frac{1}{4}}\lambda_z(\mathcal{B}_\ddagger)\quad\text{(by assumption on }U_z).\end{aligned} \tag{61}$$

However $\forall x\in U_z'$ we have $\eta\big(V_1(D_{M_0}(x))\big)>\alpha^{\frac{1}{4}}\eta\big(V\mathcal{B}(D_{M_0}(x))\big)$, therefore

$$\begin{aligned}\eta\Big(V_1\Big(\bigcup_{x\in U_z'}D_{M_0}(x)\Big)\Big)&>\alpha^{\frac{1}{4}}\eta\Big(V\mathcal{B}\Big(\bigcup_{x\in U_z'}D_{M_0}(x)\Big)\Big)\\&\asymp_{\bowtie}\alpha^{\frac{1}{4}}\lambda_z\Big(\bigcup_{x\in U_z'}D_{M_0}(x)\Big)\quad\text{(by Property 3.23)}\\&\geq\alpha^{\frac{1}{4}}\alpha^{\frac{1}{4}}\lambda_z(\mathcal{B})=\sqrt{\alpha}\lambda_z(\mathcal{B})\quad\text{(by inequality (61))}\\&\asymp_{\bowtie}\sqrt{\alpha}\eta(V\mathcal{B})\quad\text{(by Property 3.23).}\end{aligned}$$

Since $\eta(V_1)\geq\eta\big(V_1\big(\bigcup_{x\in U_z'}D_{M_0}(x)\big)\big)$ and since $\sqrt{\alpha}>M(\bowtie)\alpha$ for $\alpha<\frac{1}{M^2}$, it contradicts the assumptions of the lemma. □

Let us point out that in this lemma, we first showed that on a fixed level-set, most of its point were surrounded by almost only of vertical geodesic not in V_1. This remark will be relevant in the proof of Proposition 4.7.

The next three lemmas are estimates on the quantity of Y-horospheres satisfying specific properties. They are used in Section 4.4. Let $\mathcal{B}$ be a box, $x \in \mathcal{B}^X$, let $U \subset \mathcal{B}$ and let us denote by

$$H_x := \{x\} \bowtie \mathcal{B}^Y = \{(x, y) \mid y \in \mathcal{B}^Y,\ h(y) = -h(x)\} = (p^X)^{-1}(x)$$

a Y-horosphere of $\mathcal{B}$. Let us define

$$E^Y(x) := \{y \in \mathcal{B}^Y \mid (x, y) \in U^c\} = p^Y(p^{X-1}(x) \cap U^c) = (H_x \cap U^c)^Y;$$
$$E^X := \left\{ x \in \mathcal{B}^X \mid \lambda^Y_{-h(x)}(E^Y(x)) > \sqrt{\alpha}\lambda^Y(H^Y_x) \text{ and } h(x) \geq h^-(\mathcal{B}^X) + \frac{R}{2} \right\}.$$

The set E^X is in bijection with the "bad" Y-horospheres H above the middle of $\mathcal{B}$, the ones which have more than $\sqrt{\alpha}$ fraction of their measure λ^Y in U^c.

The following lemma asserts that almost all Y-horospheres in the upper half of the box are good Y-horospheres.

Lemma 3.27 *If $\lambda(U) \geq (1-\alpha)\lambda(\mathcal{B})$ with $0 < \alpha < 1$, then we have*

$$\lambda^X(E^X) < \sqrt{\alpha}\lambda^X(\mathcal{B}^X).$$

Proof Without loss of generality we can assume that $h(\mathcal{B}) = [0; R[$. We proceed by contradiction, let us assume that $\lambda^X(E^X) \geq \sqrt{\alpha}\lambda^X(\mathcal{B}^X)$. Then we compute the measure of U^c:

$$\begin{aligned}
\lambda(U^c) &= \int_0^R \lambda^X_z \otimes \lambda^Y_{-z}(U^c_z)\,\mathrm{d}z = \int_0^R \int_{\mathcal{B}^X_z} \lambda^Y_{-z}(\{y \in Y_{-z} \mid (x, y) \in U^c_z\})\,\mathrm{d}\lambda^X_z(x)\,\mathrm{d}z \quad \text{(by definition)} \\
&= \int_0^R \int_{\mathcal{B}^X_z} \lambda^Y_{-z}((H_x \cap U^c)^Y)\,\mathrm{d}\lambda^X_z(x)\,\mathrm{d}z \\
&\geq \int_0^R \int_{E^X_z} \lambda^Y_{-z}((H_x \cap U^c)^Y)\,\mathrm{d}\lambda^X_z(x)\,\mathrm{d}z \quad \text{(since } E^X_z \subset \mathcal{B}^X_z\text{)} \\
&> \sqrt{\alpha}\int_0^R \left[\int_{E^X_z} \lambda^Y_{-z}(H^Y_x)\,\mathrm{d}\lambda^X_z(x)\right]\mathrm{d}z \quad \text{(by the definition of } E^X\text{)} \\
&= \sqrt{\alpha}\int_0^R [\lambda^Y_{-z}(\mathcal{B}^Y_{-z})\lambda^X_z(E^X_z)]\,\mathrm{d}z \quad \text{(by the definition of } H_x\text{)} \\
&\geq \sqrt{\alpha}\sqrt{\alpha}\int_0^R \lambda^Y_{-z}(\mathcal{B}^Y_{-z})\lambda^X_z(\mathcal{B}^X_z)\,\mathrm{d}z \geq \alpha\lambda(\mathcal{B}) \quad \text{(by assumption on } E^X\text{)},
\end{aligned}$$

which contradicts the assumption on U. □

For all $U \subset \mathcal{B}$ we define the shadow of U, denoted by $\mathrm{Sh}(U)$, as

$$\mathrm{Sh}(U) := \{p \in \mathcal{B} \mid \exists V \in V\mathcal{B} \text{ containing } p \text{ and intersecting } U \text{ on a point } p' \text{ such that } h(p') \geq h(p)\}.$$

For S a subset of X, we shall call the *large Y-horosphere* (see Figure 16) the subset H_S defined by

$$H_S := S \bowtie Y = (p^X)^{-1}(S).$$

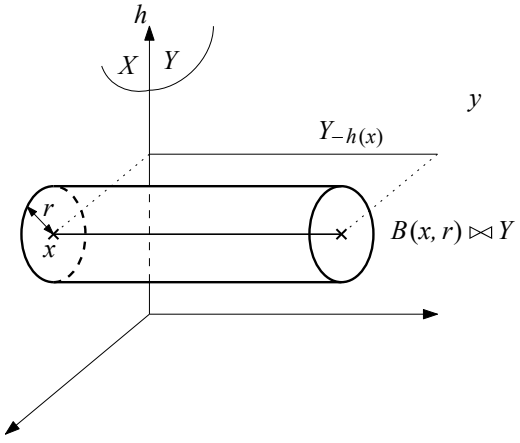


Figure 16: Large X-horosphere in $X \bowtie Y$.

Let M_0 be the constant involved in assumption (E2). Let us denote by $F^X \subset \mathcal{B}^X$ the subset

$$F^X := \left\{ x \in \mathcal{B}^X \mid \lambda(\mathrm{Sh}(H_{D_{M_0}(x)}) \cap U^c) > \alpha^{\frac{1}{4}} \lambda(\mathrm{Sh}(H_{D_{M_0}(x)})) \text{ and } h(x) \geq h^-(\mathcal{B}^X) + \frac{R}{2} \right\}.$$

The set F^X is in bijection with the "bad" Y-horospheres H that are above the middle of the box $\mathcal{B}$. By "bad" we mean the ones which have more than $\alpha^{\frac{1}{4}}$ fraction of the measure λ of their shadow in U^c.

In the following lemma, we show that the shadow of almost all the Y-horospheres in the upper half of the box have almost full measure.

Lemma 3.28 *There exists a constant $0 < \alpha(\bowtie) \leq 1$ such that for all $0 < \alpha \leq \alpha(\bowtie)$ the following statement holds. If $\lambda(U) \geq (1-\alpha)\lambda(\mathcal{B})$, then we have*

$$\lambda^X(F^X) < \alpha^{\frac{1}{4}} \lambda^X(\mathcal{B}^X).$$

Proof Without loss of generality we can assume that $h(\mathcal{B}) = [0; R[$. We proceed by contradiction, let us assume that $\lambda^X(F^X) \geq \alpha^{\frac{1}{4}} \lambda^X(\mathcal{B}^X)$. Therefore, there exists $z_0 \in \left[\frac{R}{2}, R\right[$ such that

$$\lambda^X_{z_0}(F^X_{z_0}) \geq \alpha^{\frac{1}{4}} \lambda^X_{z_0}(\mathcal{B}^X_{z_0}).$$

Let Z be a $2M_0$-maximal separating subset of $F^X_{z_0}$. Then we have

$$\begin{aligned}
&\lambda(U^c) \\
&\quad \geq \lambda\left(\mathrm{Sh}\left(\bigsqcup_{x\in Z} H_{D_{M_0}(x)}\right) \cap U^c\right) = \sum_{x\in Z} \lambda(\mathrm{Sh}(H_{D_{M_0}(x)}) \cap U^c) \quad \text{(since this is a disjoint union)} \\
&\quad \geq \alpha^{\frac{1}{4}} \sum_{x\in Z} \lambda(\mathrm{Sh}(H_{D_{M_0}(x)})) \asymp_{\bowtie} \alpha^{\frac{1}{4}} \sum_{x\in Z} z_0 \lambda_{z_0}(H_{D_{M_0}(x)}) \quad \text{(by definition of } F^X_{z_0} \text{ and Property 3.5).}
\end{aligned}$$

However $\lambda_{z_0}(H_{D_{M_0}(x)}) = \lambda^X_{z_0}(D_{M_0}(x))\lambda^Y_{-z_0}(\mathcal{B}^Y_{-z_0})$ since $H_{D_{M_0}(x)} = D_{M_0}(x) \times \mathcal{B}^Y_{-z_0}$, hence

$$\begin{aligned}
\lambda(U^c) &\succeq_{\bowtie} \alpha^{\frac{1}{4}} z_0 \sum_{x\in Z} \lambda^X_{z_0}(D_{M_0}(x))\lambda^Y_{-z_0}(\mathcal{B}^Y_{-z_0}) \\
&\asymp_{\bowtie} \alpha^{\frac{1}{4}} z_0 \lambda^Y_{-z_0}(\mathcal{B}^Y_{-z_0}) \sum_{x\in Z} \lambda^X_{z_0}(D_{2M_0}(x)) \quad \text{(by Lemma 3.2)} \\
&\geq \alpha^{\frac{1}{4}} z_0 \lambda^Y_{-z_0}(\mathcal{B}^Y_{-z_0})\lambda^X_{z_0}\Big(\bigcup_{x\in Z} D_{2M_0}(x)\Big) \geq \alpha^{\frac{1}{4}} z_0 \lambda^Y_{-z_0}(\mathcal{B}^Y_{-z_0})\lambda^X_{z_0}(F^X_{z_0}) \quad \text{(by definition of } Z) \\
&\geq \alpha^{\frac{1}{4}}\alpha^{\frac{1}{4}} z_0 \lambda^Y_{-z_0}(\mathcal{B}^Y_{-z_0})\lambda^X_{z_0}(\mathcal{B}^X_{z_0}) \quad \text{(by assumption on } F^X_{z_0}) \\
&\geq \sqrt{\alpha}\frac{R}{2}\lambda^Y_{-z_0}(\mathcal{B}^Y_{-z_0})\lambda^X_{z_0}(\mathcal{B}^X_{z_0}) \asymp_{\bowtie} \tfrac{1}{2}\sqrt{\alpha}\lambda(\mathcal{B}) \quad \Big(\text{since } z_0 \geq \frac{R}{2} \text{ and by Proposition 3.22}\Big),
\end{aligned}$$

which contradicts the assumptions on U for $\alpha < \frac{1}{M(\bowtie)^2}$. □

The following lemma asserts that the projection on a level-set of almost all the Y-horospheres have almost full measure.

Lemma 3.29 *If $\lambda(U) \geq (1-\alpha)\lambda(\mathcal{B})$, then there exists a constant $M(\bowtie)$ such that for any large Y-horosphere $H_{D_{M_0}(x)}$ with $x \in \mathcal{B}_{\S} \setminus F_X$ as in Lemma 3.28, and for $1 \geq M\rho \geq M^2\alpha^{\frac{1}{4}} > 0$, there exists P a level set of the height function in $\mathcal{B}$, such that*

$$\lambda_{h(P)}(P \cap \mathrm{Sh}(H_{D_{M_0}(x)}) \cap U^c) \preceq_{\bowtie} \alpha^{\frac{1}{4}} \lambda_{h(P)}(P \cap \mathrm{Sh}(H_{D_{M_0}(x)})).$$

Furthermore, P can be chosen such that $\rho R < d_{\bowtie}(P, H) < 2\rho R$.

Proof We proceed by contradiction. Let us assume that such a plane P does not exist. Computing the measure λ of $\mathrm{Sh}(H_{D_{M_0}(x)}) \cap U^c \cap \mathcal{B}_{[h(H)-2\rho R; h(H)-\rho R]}$ contradicts the fact that

$$\lambda(\mathrm{Sh}(H_{D_{M_0}(x)}) \cap U^c) \leq \alpha^{\frac{1}{4}} \lambda(\mathrm{Sh}(H_{D_{M_0}(x)})).$$

Indeed, we show the contradiction using Property 3.5 and because we integrate on a sufficiently large portion of $[0, R]$ ($\rho \geq M\alpha^{\frac{1}{4}}$). □

In the following lemma we show that almost all level-sets admit a point with large X-horospheres and Y-horospheres.

Lemma 3.30 *There exists a constant $0 < \alpha(\bowtie) \leq 1$ such that for all $0 < \alpha \leq \alpha(\bowtie)$ the following statement holds. Let $U \subset \mathcal{B}$ be such that $\lambda(U) \geq (1-\alpha)\lambda(\mathcal{B})$. Then there exists $U' \subset U$ such that*

(1) $\lambda(U') \geq (1-\alpha^{\frac{1}{4}})\lambda(\mathcal{B})$;

(2) *for all $z \in h(U')$ there exists $(x_{0,z}, y_{0,z}) \in U'_z$ such that for all $(x_1, y_1) \in U'_z$, we have $(x_1, y_{0,z}) \in U'_z$ and $(x_{0,z}, y_1) \in U'_z$.*

Proof We may assume without loss of generality that $h(\mathcal{B}) = [0, R[$. Let us define

$$H_U := \{z \in [0, R[\mid \lambda_z(U_z) \geq (1-\alpha^{\frac{1}{4}})\lambda_z(\mathcal{B}_z)\}.$$

Then we claim that $\mathrm{Leb}(H_U) \geq (1-\alpha^{\frac{1}{4}})R$. To prove this claim we proceed by contradiction. Let us assume that $\mathrm{Leb}(H_U) < (1-\alpha^{\frac{1}{4}})R$. Then $\mathrm{Leb}([0, R[\setminus H_U) \geq \alpha^{\frac{1}{4}} R$. Furthermore, for all $z \in [0, R[\setminus H_U$ we have $\lambda_z(U_z) < (1-\alpha^{\frac{1}{4}})\lambda_z(\mathcal{B}_z)$, hence

$$\lambda_z(\mathcal{B}_z \setminus U_z) \geq \alpha^{\frac{1}{4}} \lambda_z(\mathcal{B}_z). \tag{62}$$

Therefore, by computing the measure of $\mathcal{B} \setminus U$ we have

$$\begin{aligned}
\lambda(\mathcal{B} \setminus U) &= \int_{z\in[0,R[} \lambda_z(\mathcal{B}_z \setminus U_z)\,\mathrm{d}z \geq \int_{z\in([0,R[\setminus H_U)} \lambda_z(\mathcal{B}_z \setminus U_z)\,\mathrm{d}z \\
&\geq \int_{z\in([0,R[\setminus H_U)} \alpha^{\frac{1}{4}} \lambda_z(\mathcal{B}_z)\,\mathrm{d}z \quad \text{(by inequality (62))} \\
&\succeq_X \alpha^{\frac{1}{2}} \lambda(\mathcal{B}) \quad \text{(by the contradiction assumption and Property 3.5)},
\end{aligned}$$

which contradicts the assumption on U for α small enough. Hence $\mathrm{Leb}(H_U) \geq (1-\alpha^{\frac{1}{4}})R$.

Let us define, for $z \in [0; R[$,

$$U^y := \{x \in \mathcal{B}_z^X \mid (x, y) \in U\}; \quad H := \{z \in [0, R[\mid \exists y \in \mathcal{B}_{-z}^Y,\ \lambda_z^X(U^y) \geq (1-\alpha^{\frac{1}{4}})\lambda_z^X(\mathcal{B}_z^X)\}.$$

In particular, for all $y \in \mathcal{B}_{-z}^Y$ we have $U^y \subset U_z^X$, and, by the definition of λ,

$$\lambda(U) = \int_{z\in[0,R[} \int_{y\in\mathcal{B}_{-z}^Y} \lambda_z^X(U^y).$$

We claim that $\mathrm{Leb}(H) \geq (1-\alpha^{\frac{1}{4}})R$. To prove this claim, we also proceed by contradiction. Let us assume that $\mathrm{Leb}(H) < (1-\alpha^{\frac{1}{4}})R$. Then $\mathrm{Leb}([0, R[\setminus H) \geq \alpha^{\frac{1}{4}} R$. Furthermore for all $z \in [0, R[\setminus H$ we have that

$$\forall y \in \mathcal{B}_{-z}^Y, \quad \lambda_z^X(U^y) < (1-\alpha^{\frac{1}{4}})\lambda_z^X(\mathcal{B}_z^X).$$

Therefore, by the definition of U_y we have that, $\forall y \in \mathcal{B}_{-z}^Y$,

$$\lambda_z^X(\{x \in \mathcal{B}_z^X \mid (x, y) \notin U\}) \geq \alpha^{\frac{1}{4}} \lambda_z^X(\mathcal{B}_z^X). \tag{63}$$

Hence, by computing the measure of $\mathcal{B} \setminus U$ we have

$$\begin{aligned}
\lambda(\mathcal{B} \setminus U) &= \int_{z\in[0,R[} \int_{y\in\mathcal{B}_{-z}^Y} \lambda_z^X(\{x \in U_z^X \mid (x, y) \notin U\})\,\mathrm{d}\lambda_{-z}^Y\,\mathrm{d}z \\
&\geq \int_{z\in([0,R[\setminus H)} \int_{y\in\mathcal{B}_{-z}^Y} \lambda_z^X(\{x \in U_z^X \mid (x, y) \notin U\})\,\mathrm{d}\lambda_{-z}^Y\,\mathrm{d}z \\
&\geq \int_{z\in([0,R[\setminus H)} \int_{y\in\mathcal{B}_{-z}^Y} \alpha^{\frac{1}{4}} \lambda_z^X(\mathcal{B}_z^X)\,\mathrm{d}\lambda_{-z}^Y\,\mathrm{d}z \quad \text{(by inequality (63))} \\
&= \alpha^{\frac{1}{4}} \int_{z\in([0,R[\setminus H)} \lambda_{-z}^Y(\mathcal{B}_{-z}^Y)\lambda_z^X(\mathcal{B}_z^X)\,\mathrm{d}z \\
&\succeq_X \alpha^{\frac{1}{4}}\alpha^{\frac{1}{4}} \lambda(\mathcal{B}) = \alpha^{\frac{1}{2}} \lambda(\mathcal{B}) \quad \text{(by the contradiction assumption and Property 3.5)},
\end{aligned}$$

which contradicts the assumption $\lambda(\mathcal{B}\setminus U)<\alpha\lambda(\mathcal{B})$, for $\alpha<\frac{1}{M(\bowtie)^2}$. Let us define, for all $x\in\mathcal{B}^X_z$,

$$U^x:=\{y\in\mathcal{B}^Y_{-z}\mid(x,y)\in U\};\quad H':=\{z\in[0,R[\mid\exists x\in\mathcal{B}^X_z,\ \lambda^Y_{-z}(U^x)\geq(1-\alpha^{\frac14})\lambda^Y_{-z}(\mathcal{B}^Y_{-z})\}.$$

We show similarly that $\mathrm{Leb}(H')\geq(1-\alpha^{\frac14})R$. Therefore $\mathrm{Leb}(H\cap H'\cap H_U)\geq(1-3\alpha^{\frac14})R$.

For all $z\in H\cap H'$ there exists $(x_{0,z},y_{0,z})\in\mathcal{B}_z$ such that

$$\lambda^X_z(U^{y_0})\geq(1-\alpha^{\frac14})\lambda^X_z(\mathcal{B}^X_z);\tag{64}$$

$$\lambda^Y_{-z}(U^{x_0})\geq(1-\alpha^{\frac14})\lambda^Y_{-z}(\mathcal{B}^Y_{-z}).\tag{65}$$

Let us define for all $z\in H_U\cap H\cap H'$, $U'_z:=(U^{x_{0,z}}\times U^{y_{0,z}})$. Then we have

(1) $U'\subset U$;

(2) $\lambda_z(U'_z)=\lambda_z((U^{x_{0,z}}\times U^{y_{0,z}})\cap U_z)\geq(1-3\alpha^{\frac14})\lambda_z(\mathcal{B})$ by inequalities (64), (65) and by the definition of H_U;

(3) for all $(x_1,y_1)\in U'_z$ we have $(x_1,y_{0,z})\in U'_z$ and $(x_{0,z},y_1)\in U'_z$.

Let $(x_1,y_1)\in U'_z$. Then $(x_1,y_{0,z})\in U'$, and hence $(x_{0,z},y_{0,z})\in U'$. Furthermore we have that $\mathrm{Leb}(H_U\cap H\cap H')\geq(1-3\alpha^{\frac14})R$, and hence $\mathrm{Leb}([0,R[\setminus(H_U\cap H\cap H'))\leq3\alpha^{\frac14}R$. Therefore

$$\begin{aligned}\lambda(\mathcal{B}\setminus U')&=\int_{z\in[0,R[}\lambda_z((\mathcal{B}\setminus U')_z)\,\mathrm{d}z\\&=\int_{z\in([0,R[\setminus(H_U\cap H\cap H'))}\lambda_z(\mathcal{B}_z\setminus(U^{x_{0,z}}\times U^{y_{0,z}}))\,\mathrm{d}z\\&\leq\int_{z\in([0,R[\setminus(H_U\cap H\cap H'))}(3\alpha^{\frac14})\lambda_z(\mathcal{B}_z)\,\mathrm{d}z\quad\text{(by construction of }U'_z)\\&\preceq_X 9\alpha^{\frac12}\lambda(\mathcal{B})\quad\text{(by the measure of }[0,R[\setminus(H_U\cap H\cap H')\text{ and by Property 3.5).}\end{aligned}$$

Hence $\lambda(U')\geq(1-\alpha^{\frac14})\lambda(\mathcal{B})$, since $\alpha^{\frac14}>9M(X)\alpha^{\frac12}$ (α small enough in comparison to a constant depending only on X). □

These points $(x_{0,z},y_{0,z})$ will play a key role in the definition of the product map close to a given quasi-isometry in Theorem 4.5.

3.8 Divergence

Two distinct vertical geodesics in a δ-hyperbolic and Busemann space diverge quickly from each other. The next lemma aims at making this more precise for X an admissible horopointed space. More specifically we are going to look at a point x and at all the vertical geodesics passing by a point of the disc centred at x of radius M_0 (the (E2) constant) along the horosphere at height $h(x)$, that is, $VD_{M_0}(x)$. Let V_0 be a geodesic containing x. We want to quantify the vertical geodesics in $VD_{M_0}(x)$ which start diverging from the vertical geodesic V_0 between the heights $h(x)-l$ and $h(x)+l$. We denote this set by

$$\mathrm{Div}(V_0):=\{V\in VD_{M_0}(x)\mid|h_{\mathrm{Div}}(V_0,V)-h(x)|\leq l\}.$$

Lemma 3.31 *With the above notation we have*

$$\eta^X(VD_{M_0}(x)\backslash \mathrm{Div}(V_0)) \preceq_X e^{-ml}\eta^X(VD_{M_0}(x)).$$

Proof By slight abuse of notation, we may intersect a set of vertical geodesic segments $E \subset V\mathcal{B}$ with a subset $F \subset \mathcal{B}$. By this, we mean the intersection of F with the union of the images of the vertical geodesics of E, where each image is a vertical geodesic in $\mathcal{B}$. For example,

$$VD_{M_0}(x) \cap \mathcal{B}_{h(x)} = D_{M_0}(x).$$

Any vertical geodesic segment $V \in VD_{M_0}(x)$ did not start to diverge from the vertical geodesic V_0 at the height $h(x)$, we have $h_{\mathrm{Div}}(V, V_0) \leq h(x)$. Therefore, all the vertical geodesic segments which did not start to diverge at the height $h(x)-l$, denoted by $VD_{M_0}(x)\backslash \mathrm{Div}(V_0)$, are still M_0-close to $\pi_{h(x)-l}(x)$:

$$(VD_{M_0}(x)\backslash \mathrm{Div}(V_0)) \cap \mathcal{B}_{h(x)-l} \subset D_{M_0}(\pi_{h(x)-l}(x)). \tag{66}$$

We use Lemma 1.8 with $z_0 = h(x)$ and $z = h(x)-l$, which gives

$$D_{2l-M_0}(\pi_{h(x)-l}(x)) \subset \pi_{h(x)-l}(D_{M_0}(x)) = VD_{M_0}(x) \cap \mathcal{B}_{h(x)-l}. \tag{67}$$

Therefore

$$\begin{aligned}
\frac{\eta^X(VD_{M_0}(x)\backslash \mathrm{Div}(V_0))}{\eta^X(VD_{M_0}(x))} &\asymp_X \frac{\lambda^X_{h(x)-l}(VD_{M_0}(x)\backslash \mathrm{Div}(V_0) \cap \mathcal{B}_{h(x)-l})}{\lambda^X_{h(x)-l}(VD_{M_0}(x) \cap \mathcal{B}_{h(x)-l})} \quad \text{(by Property 3.23)} \\
&\leq \frac{\lambda^X_{h(x)-l}\big(D_{M_0}(\pi_{h(x)-l}(x))\big)}{\lambda^X_{h(x)-l}(VD_{M_0}(x) \cap \mathcal{B}_{h(x)-l})} \quad \text{(by inequality (66))} \\
&\leq \frac{\lambda^X_{h(x)-l}\big(D_{M_0}(\pi_{h(x)-l}(x))\big)}{\lambda^X_{h(x)-l}\big(D_{2l-M_0}(\pi_{h(x)-l}(x))\big)} \quad \text{(by inequality (67))}.
\end{aligned}$$

Moreover by the definition of λ^X and Lemma 3.2

$$\frac{\lambda^X_{h(x)-l}\big(D_{M_0}(\pi_{h(x)-l}(x))\big)}{\lambda^X_{h(x)-l}\big(D_{2l-M_0}(\pi_{h(x)-l}(x))\big)} = \frac{\mu^X_{h(x)-l}\big(D_{M_0}(\pi_{h(x)-l}(x))\big)}{\mu^X_{h(x)-l}\big(D_{2l-M_0}(\pi_{h(x)-l}(x))\big)} \preceq_X e^{-ml}. \tag{68}$$

Therefore

$$\frac{\eta^X(VD_{M_0}(x)\backslash \mathrm{Div}(V_0))}{\eta^X(VD_{M_0}(x))} \preceq_X e^{-ml}. \qquad \square$$

Heuristically, the previous lemma asserts that most of the vertical geodesics segments passing close to a point x, start diverging from each other close to the height $h(x)$.

We now provide an estimate on the exponential contraction of the measure μ along the vertical direction.

Lemma 3.32 *There exists $M(\bowtie)$ such that the following holds. Let $h_0 \in \mathbb{R}$ and let $U \subset (X \bowtie Y)_{h_0}$ be a measurable subset. Let $\Delta > M$ and let $A \subset (X \bowtie Y)_{h_0-\Delta}$ be a measurable subset. Suppose also that all*

vertical rays V intersecting U intersect A. Then

$$\mu_{h_0-\Delta}(A) \succeq_{\bowtie} e^{(m-n)\Delta}\mu_{h_0}(U).$$

Proof Since $\pi^{\bowtie}_{h_0-\Delta}(U) \subset A$ we have

$$\mu_{h_0-\Delta}(\pi^{\bowtie}_{h_0-\Delta}(U)) \leq \mu_{h_0-\Delta}(A),$$

where $\pi^{\bowtie}$ is defined in Definition 1.13. We recall that for all $x \in X$, $U^Y_x := \{y \in Y \mid (x,y) \in U\}$. By definition

$$\mu_{h_0}(U) = \mu^X_{h_0} \otimes \mu^Y_{-h_0}(U) = \int_{X_{h_0}} \mu^Y_{-h_0}(U^Y_x)\,\mathrm{d}\mu^X_{h_0}(x). \tag{69}$$

For all $x \in U^X$ let us define $U_x := \{(x,y) \in U \mid y \in U^Y\}$. Then

$$(U_x)^Y = U^Y_x := \{y \in Y \mid (x,y) \in U\}.$$

Furthermore $U^Y_x \subset \pi^Y_{-h_0}[\pi^Y_{\Delta-h_0}(U^Y_x)]$. Hence

$$\mu^Y_{-h_0}(U^Y_x) \leq \mu^Y_{-h_0}(\pi^Y_{-h_0}[\pi^Y_{\Delta-h_0}(U^Y_x)]) \asymp_{\bowtie} e^{n\Delta}\mu^Y_{\Delta-h_0}[\pi^Y_{\Delta-h_0}(U^Y_x)] \quad \text{(by assumption (E3))},$$

which gives us

$$\mu_{h_0}(U) \preceq_{\bowtie} e^{n\Delta}\int_{U^X} \mu^Y_{\Delta-h_0}[\pi^Y_{\Delta-h_0}(U^Y_x)]\,\mathrm{d}\mu^X_{h_0}(x) \quad \text{(by definition of } \mu_{h_0}). \tag{70}$$

However we have

$$\begin{aligned}\pi^Y_{\Delta-h_0}(U^Y_x) &= (\pi^{\bowtie}_{h_0-\Delta}(U_x))^Y = (\pi^{\bowtie}_{h_0-\Delta}(U))^Y_{\pi^X_{h_0-\Delta}(x)} \\ &= \{y \in (\pi^{\bowtie}_{h_0-\Delta}(U))^Y \mid (\pi^X_{h_0-\Delta}(x), y) \in \pi^{\bowtie}_{h_0-\Delta}(U)\}.\end{aligned} \tag{71}$$

Hence

$$\begin{aligned}\mu_{h_0}(U) &\preceq_{\bowtie} e^{n\Delta}\int_{U^X} \mu^Y_{\Delta-h_0}[(\pi^{\bowtie}_{h_0-\Delta}(U))^Y_{\pi^X_{h_0-\Delta}(x)}]\,\mathrm{d}\mu^X_{h_0}(x) \quad \text{(by (70) and (71))} \\ &= e^{n\Delta}\int_{\pi^X_{h_0-\Delta}(U^X)} \mu^Y_{\Delta-h_0}[(\pi^{\bowtie}_{h_0-\Delta}(U))^Y_{x'}]\,\mathrm{d}\pi^X_{h_0-\Delta} * \mu^X_{h_0}(x') \\ &\asymp_{\bowtie} e^{n\Delta}e^{-m\Delta}\int_{\pi^X_{h_0-\Delta}(U^X)} \mu^Y_{\Delta-h_0}[(\pi^{\bowtie}_{h_0-\Delta}(U))^Y_{x'}]\,\mathrm{d}\mu^X_{h_0-\Delta}(x') \quad \text{(by assumption (E3))} \\ &= e^{(n-m)\Delta}\mu_{h_0-\Delta}(\pi^{\bowtie}_{h_0-\Delta}(U)).\end{aligned}$$

Furthermore, as said at the beginning we have $\mu_{h_0-\Delta}(\pi^{\bowtie}_{h_0-\Delta}(U)) \leq \mu_{h_0-\Delta}(A)$. Therefore

$$\mu_{h_0-\Delta}(A) \succeq_{\bowtie} e^{(m-n)\Delta}\mu_{h_0}(U). \qquad \square$$

In the next lemma we transfer a control on the measure μ to a control on the measure η.

Lemma 3.33 *Let M_0 be the constant involved in assumption* (E2), *$\mathcal{B}$ be a box and $z \in h(\mathcal{B})$. Let $A \subset (\mathcal{B})_z$ and let $E \subset \mathcal{B}$ such that $h^+(E) \leq h(A)$. Then, if there exists $Q \geq 1$ such that $\mu(\mathcal{N}_{M_0}(E)) \leq Q^{-1}\mu(\mathcal{N}_{M_0}(A))$, we have that*

$$\eta(V\mathcal{N}_{M_0}(E)) \preceq_{\bowtie} Q^{-1}\eta(V\mathcal{N}_{M_0}(A)).$$

Proof Let $Z \subset E$ be a $2M_0$-maximal separating set.

(1) The balls $B(p, M_0)$ for $p \in Z$ are pairwise disjoint.

(2) We have the inclusions

$$\bigsqcup_{p\in Z} B(p, M_0) \subset \mathcal{N}_{M_0}(E) \subset \bigcup_{p\in Z} B(p, 3M_0).$$

The radius $3M_0$ is required since we cover all $\mathcal{N}_{M_0}(E)$ and not only E. Furthermore, all balls and disks of radius M_0 have comparable measure μ by assumption (E2) and Corollary 3.17. Therefore

$$\mu(\mathcal{N}_{M_0}(E)) \asymp_{\bowtie} \#Z \asymp_{\bowtie} \sum_{p\in Z} \mu(B(p, M_0)) \asymp_{\bowtie} \sum_{p\in Z} \mu_{h(p)}(D_{M_0}(p)). \tag{72}$$

Moreover, for all $v \in VE$, there exists $p \in Z$ such that $v \cap D_{3M_0}(p) \neq \varnothing$. Consequently we have $V\mathcal{N}_{M_0}(E) \subset \bigcup_{p\in Z} VD_{3M_0}(p)$, hence

$$\begin{aligned}\eta(V\mathcal{N}_{M_0}(E)) &\leq \sum_{p\in Z} \eta(VD_{3M_0}(p)) \asymp_X \sum_{p\in Z} \lambda_{h(p)}(D_{3M_0}(p)) \quad \text{(by Property 3.23)}\\ &\leq \sum_{p\in Z} \lambda^X_{h(p)}(D_{6M_0}(p^X))\lambda^Y_{-h(p)}(D_{6M_0}(p^Y)).\end{aligned}$$

Furthermore, disks of radius r are included in rectangles of width $2r$, hence

$$\begin{aligned}\eta(V\mathcal{N}_{M_0}(E)) &\preceq_{\bowtie} \sum_{p\in Z} e^{h(p)(m-n)}\mu_{h(p)}(D_{6M_0}(p)) \quad \text{(by the definition of } \lambda_{h(p)})\\ &\leq e^{h(a)(m-n)} \sum_{p\in Z} \mu_{h(p)}(D_{6M_0}(p)) \quad \text{(because } h^+(E) \leq h(A))\\ &\preceq_{\bowtie} e^{h(a)(m-n)}\mu(\mathcal{N}_{M_0}(E)) \quad \text{(by inequalities (72)).}\end{aligned}$$

Using similar arguments we obtain

$$\eta(V\mathcal{N}_{M_0}(A)) \asymp_{\bowtie} \lambda_{h(a)}(V\mathcal{N}_{M_0}(A)) \asymp_{\bowtie} e^{h(a)(m-n)}\mu(V\mathcal{N}_{M_0}(A)).$$

Combined with the assumption $\mu(\mathcal{N}_{M_0}(E)) \leq Q^{-1}\mu(\mathcal{N}_{M_0}(A))$ we have

$$\eta(V\mathcal{N}_{M_0}(A)) \succeq_{\bowtie} e^{h(a)(m-n)}Q\mu(\mathcal{N}_{M_0}(E)) \succeq_{\bowtie} Q\eta(V\mathcal{N}_{M_0}(E)). \qquad \square$$

Heuristically, if a set E is sufficiently small and below a set A, then the set of vertical geodesic segments intersecting E will also be small.

4 Proof of the geometric rigidity

The aim of this chapter is to present a proof of our key result. Let (X, Y) and (X', Y') be two horopointed admissible couples of parameter respectively (m, n) and (m', n'). Let us assume that $m > n$ and $m' > n'$.

Theorem 4.1 *Let $\Phi : X \bowtie Y \to X' \bowtie Y'$ be a (k, c) quasi-isometry. Then there exist two quasi-isometries $\Phi^X : X \to X'$ and $\Phi^Y : Y \to Y'$ such that*

$$d_{\bowtie}(\Phi, (\Phi^X, \Phi^Y)) \preceq_{k,c,\bowtie} 1.$$

Although this statement is similar to the statement in the case of Sol and Diestel–Leader, our broader setting of admissible spaces requires additional key arguments, such as Lemma 3.3, and therefore relies heavily on the previous sections.

To make the exposition of the various statements in this chapter smoother, we made the following abuse of notation. In a statement, when a parameter, say θ, needs to be sufficiently small, we will write it by "For $\theta \preceq_{\bowtie} 1$ we have ..." instead of "There exists a constant $M(\bowtie)$ such that if $\theta \leq \frac{1}{M}$, then ...".

From now until the end of this chapter we consider $\Phi : X \bowtie Y \to X' \bowtie Y'$ a (k, c)-quasi-isometry with fixed constants $k \geq 1$ and $c \geq 0$.

4.1 Vertical geodesics with ε-monotone image

In order to construct a product map, the key idea is to use the quadrilateral lemmas of Section 2.4 on the image by the quasi-isometry Φ of a quadrilateral in $X \bowtie Y$. To do so we need to locate which vertical geodesic segments are sent close to vertical geodesic segments. Thanks to Proposition 2.6 it is sufficient to look for vertical geodesic segments with an ε-monotone image under Φ, where $0 \leq \varepsilon < 1$ is a parameter to be determined later (depending on $\bowtie$, k and c). We call *good* these vertical geodesic segments.

Notation 4.2 We recall that we denote by $V\mathcal{B}$ the set of vertical geodesic segments of the box $\mathcal{B}$. Let us denote by $V^g\mathcal{B}$ the set of good vertical geodesic segments and $V^b\mathcal{B}$ the set of bad vertical geodesic segments, that is,

$$V^g\mathcal{B} := \{\gamma \in V\mathcal{B} \mid \Phi \circ \gamma \text{ is } \varepsilon\text{-monotone}\};$$
$$V^b\mathcal{B} := \{\gamma \in V\mathcal{B} \mid \Phi \circ \gamma \text{ is not } \varepsilon\text{-monotone}\} = V\mathcal{B} \setminus V^g\mathcal{B}.$$

In the following proposition, we prove the existence of an appropriate scale on which almost all boxes possess almost only good vertical geodesics. We define $\eta := \eta_{V\mathcal{B}}$, $\eta^X := \eta^X_{V\mathcal{B}^X}$ and $\eta^Y := \eta^Y_{V\mathcal{B}^Y}$.

Proposition 4.3 *For $0 < \theta \preceq_{\bowtie} 1$, there exist two positive constants $M(k, c, \bowtie, \varepsilon)$ and $M'(k, c, \bowtie)$ such that for all $r_0 \geq M$, $N \geq \frac{M'}{\varepsilon}$ and $S \geq \frac{M'}{\varepsilon\theta^3}$ and boxes $\mathcal{B}$ at scale $L := N^S r_0$, there exist $k_0 \in \{1, \dots, S\}$, a box tiling $\bigsqcup_{i\in I} \mathcal{B}_i = \mathcal{B}$ at scale $R = N^{k_0} r_0$ and $I_g \subset I$ such that*

(1) $\lambda\big(\bigcup_{i\in I_g} \mathcal{B}_i\big) \geq (1-\theta)\lambda(\mathcal{B})$ (*boxes indexed by I_g cover almost all $\mathcal{B}$*);

(2) $\forall i \in I_g$, $\frac{\eta_i(V^b\mathcal{B}_i)}{\eta_i(V\mathcal{B}_i)} \leq \theta$ (*almost all vertical geodesic segments in $\mathcal{B}_i$ have ε-monotone image*);

where $\eta_i := \eta_{V\mathcal{B}_i}$.

Proof We recall from Proposition 2.9 the definition of $\delta_s(\alpha)$ for a quasi-geodesic segment α:

$$A_s := \{\alpha([kN^s r_0, (k+1)N^s r_0]) \mid k \in \{0, \dots, N^{S-s} - 1\}\}.$$

Then $\delta_s(\alpha)$ is the proportion of segments in A_s which are not ε-monotone:

$$\delta_s(\alpha) := \frac{\#\{\beta \in A_s \mid \beta \text{ is not } \varepsilon\text{-monotone}\}}{\#A_s}. \tag{73}$$

Using Proposition 2.9 on every vertical geodesic segment in $\mathcal{B}$ we have that, $\forall \alpha \in V\mathcal{B}$,

$$\sum_{s=1}^{S} \delta_s(\alpha) \preceq_{\bowtie,k,c} \frac{1}{\varepsilon}. \tag{74}$$

We now integrate the inequality (74) with respect to η over $V\mathcal{B}$ to get

$$\frac{1}{\varepsilon} \succeq_{\bowtie,k,c} \frac{1}{\eta(V\mathcal{B})} \int_{\alpha \in V\mathcal{B}} \Big(\sum_{s=1}^{S} \delta_s(\alpha) \Big) \, \mathrm{d}\eta = \sum_{s=1}^{S} \Big(\frac{1}{\eta(V\mathcal{B})} \int_{\alpha \in V\mathcal{B}} \delta_s(\alpha) \, \mathrm{d}\eta \Big).$$

Consequently there exists $k_0 \in \{1, \dots, S\}$ such that

$$\frac{1}{\eta(V\mathcal{B})} \int_{\alpha \in V\mathcal{B}} \delta_{k_0}(\alpha) d\eta \preceq_{\bowtie,k,c} \frac{1}{S\varepsilon} \preceq_{\bowtie} \theta^3 \quad \text{(by assumption on } S\text{)}. \tag{75}$$

From now on we fix $R := N^{k_0} r_0$. There are $\frac{L}{R}$ layers of boxes at scale R in $\mathcal{B}$. We average $\delta_{k_0}(\alpha)$ along all $\alpha \in V\mathcal{B}$:

$$\begin{aligned} \frac{1}{\eta(V\mathcal{B})} \int_{\alpha \in V\mathcal{B}} \delta_{k_0}(\alpha) \, \mathrm{d}\eta &= \frac{1}{\eta(V\mathcal{B})} \int_{\alpha \in V\mathcal{B}} \frac{R}{L} \sum_{k=0}^{\frac{L}{R}-1} \delta_{k_0}\big(\alpha([kR; (k+1)R])\big) \, \mathrm{d}\eta \\ &= \frac{1}{\eta(V\mathcal{B})} \frac{R}{L} \sum_{k=0}^{\frac{L}{R}-1} \int_{\alpha \in V\mathcal{B}} \delta_{k_0}\big(\alpha([kR; (k+1)R])\big) \, \mathrm{d}\eta. \end{aligned} \tag{76}$$

Let us denote by $\mathcal{B}_{[k]} := \mathcal{B} \cap h^{-1}([kR; (k+1)R[)$ the k-th layer of $\mathcal{B}$. Since vertical geodesic segments of $X \bowtie Y$ are couples of vertical geodesic segments, $V\mathcal{B}_{[k]}$ is in bijection with $V\mathcal{B}^X_{[k]} \times V\mathcal{B}^Y_{[k]}$ which is itself in bijection with $\mathcal{B}^X_{kR} \times \mathcal{B}^Y_{-(k+1)R}$ as explained in Section 3.6. Let us denote by f this bijection:

$$f : \mathcal{B}_{[k]} \to \mathcal{B}^X_{kR} \times \mathcal{B}^Y_{-(k+1)R}, \quad \alpha \mapsto \big(\alpha^X(kR), \alpha^Y(-(k+1)R)\big).$$

For all $\alpha \in V\mathcal{B}$ and for all $k \in \{0, \dots, \frac{L}{R} - 1\}$ we have $\delta_{k_0}\big(\alpha([kR; (k+1)R])\big) = 0$ or 1, hence

$$\begin{aligned} \delta_{k_0}\big(\alpha([kR; (k+1)R])\big) &= \mathbb{1}_{V^b\mathcal{B}_{[k]}}\big(\alpha([kR; (k+1)R])\big) \\ &= \mathbb{1}_{f(V^b\mathcal{B}_{[k]})}\big(\alpha_X(kR), \alpha_Y(-(k+1)R)\big). \end{aligned}$$

Therefore

$$\begin{aligned}
&\int_{\alpha\in V\mathcal{B}} \delta_{k_0}\big(\alpha([kR;(k+1)R])\big)\,\mathrm{d}\eta \\
&\quad= \int_{(\alpha^X,\alpha^Y)\in V\mathcal{B}^X\times V\mathcal{B}^Y} \mathbb{1}_{f(V^b\mathcal{B}_{[k]})}\big(\alpha_X(kR),\alpha_Y(-(k+1)R)\big)\,\mathrm{d}\eta^X\,\mathrm{d}\eta^Y \\
&\quad= \int_{(x,y)\in\mathcal{B}_0^X\times\mathcal{B}_{-L}^Y} \mathbb{1}_{f(V^b\mathcal{B}_{[k]})}(\pi^X_{kR}(x),\pi^Y_{-(k+1)R}(y))\,\mathrm{d}\lambda_0^X\,\mathrm{d}\lambda_{-L}^Y \quad \text{(by definition } \eta^X \text{ and } \eta^Y\text{)} \\
(77)\qquad &\quad\asymp_{\bowtie} \int_{(x',y')\in\mathcal{B}^X_{kR}\times\mathcal{B}^Y_{-(k+1)R}} \mathbb{1}_{f(V^b\mathcal{B}_{[k]})}(x',y')\,\mathrm{d}\lambda^X_{kR}\,\mathrm{d}\lambda^Y_{-(k+1)R} \quad \text{(by Property 3.5).}
\end{aligned}$$

Let $\bigsqcup_{i\in I}\mathcal{B}_i$ be the box tiling at scale R as in Proposition 3.11, and for all $k\in\{0,\dots,\frac{L}{R}-1\}$ let us denote by $I_k\subset I$ the indices of the boxes $\mathcal{B}_i$ which tile $\mathcal{B}_{[k]}$. Then we have

$$V\mathcal{B}_{[k]}=\bigsqcup_{i\in I_k} V\mathcal{B}_i \quad\text{and}\quad V^b\mathcal{B}_{[k]}=\bigsqcup_{i\in I_k} V^b\mathcal{B}_i.$$

Therefore, for all $(x,y)\in\mathcal{B}^X_{kR}\times\mathcal{B}^Y_{-(k+1)R}$,

$$\mathbb{1}_{f(V^b\mathcal{B}_{[k]})}(x,y)=\mathbb{1}_{f(\bigsqcup_{i\in I_k}V^b\mathcal{B}_i)}(x,y)=\sum_{i\in I_k}\mathbb{1}_{f(V^b\mathcal{B}_i)}(x,y).$$

Hence from inequality (77) we have

$$\begin{aligned}
\int_{\alpha\in V\mathcal{B}}\delta_{k_0}\big(\alpha([kR;(k+1)R])\big)\,\mathrm{d}\eta &\asymp_{\bowtie} \int_{(x,y)\in\mathcal{B}^X_{kR}\times\mathcal{B}^Y_{-(k+1)R}}\sum_{i\in I_k}\mathbb{1}_{f(V^b\mathcal{B}_i)}(x,y)\,\mathrm{d}\lambda^X_{kR}\,\mathrm{d}\lambda^Y_{-(k+1)R} \\
&= \sum_{i\in I_k}\int_{(x,y)\in\mathcal{B}^X_{kR}\times\mathcal{B}^Y_{-(k+1)R}}\mathbb{1}_{f(V^b\mathcal{B}_i)}(x,y)\,\mathrm{d}\lambda^X_{kR}\,\mathrm{d}\lambda^Y_{-(k+1)R} \\
&= \sum_{i\in I_k}\int_{\alpha\in V\mathcal{B}_i}\mathbb{1}_{V^b\mathcal{B}_i}(\alpha)\,\mathrm{d}\eta_i=\sum_{i\in I_k}\eta_i(V^b\mathcal{B}_i).
\end{aligned}$$

In combination with equality (76) we have

$$\begin{aligned}
\frac{1}{\eta(V\mathcal{B})}\int_{\alpha\in V\mathcal{B}}\delta_{k_0}(\alpha)\,\mathrm{d}\eta &\succeq_{\bowtie} \frac{1}{\eta(V\mathcal{B})}\frac{R}{L}\sum_{k=0}^{\frac{L}{R}-1}\sum_{i\in I_k}\eta_i(V^b\mathcal{B}_i) \\
&\succeq_{\bowtie} \sum_{i\in I}\frac{R\eta_i(V\mathcal{B}_i)}{L\eta(V\mathcal{B})}\frac{\eta_i(V^b\mathcal{B}_i)}{\eta_i(V\mathcal{B}_i)} \\
&\succeq_{\bowtie} \sum_{i\in I}\frac{\lambda(\mathcal{B}_i)}{\lambda(\mathcal{B})}\frac{\eta_i(V^b\mathcal{B}_i)}{\eta_i(V\mathcal{B}_i)} \quad \text{(by Proposition 3.22).}
\end{aligned}$$

Let us denote by I_b the set of indices i of boxes $\mathcal{B}_i$ such that $\frac{\eta_i(V^b\mathcal{B}_i)}{\eta_i(V\mathcal{B}_i)}\geq\theta$, and $I_g:=I\setminus I_b$. By definition, I_g satisfies the second part of our proposition, we are left with proving that it also satisfies the first part.

To do so we assume by contradiction that $\lambda\big(\bigcup_{i\in I_b}\mathcal{B}_i\big)\geq\theta\lambda(\mathcal{B})$. Then

$$\begin{aligned}\frac{1}{\eta(V\mathcal{B})}\int_{\alpha\in V\mathcal{B}}\delta_{k_0}(\alpha)d\eta &\succeq_{\bowtie}\sum_{i\in I_b}\frac{\lambda(\mathcal{B}_i)}{\lambda(\mathcal{B})}\frac{\eta_i(V^b\mathcal{B}_i)}{\eta_i(V\mathcal{B}_i)}\quad(\text{since } I_b\subset I)\\ &\succeq_{\bowtie}\theta\frac{\sum_{i\in I_b}\lambda(\mathcal{B}_i)}{\lambda(\mathcal{B})}\quad(\text{by the definition of } I_b)\\ &\succeq_{\bowtie}\theta^2\quad(\text{by the contradiction assumption}),\end{aligned}$$

which contradicts inequality (75) for $\theta\preceq_{\bowtie}1$. Therefore $\lambda\big(\bigcup_{i\in I_b}\mathcal{B}_i\big)<\theta\lambda(\mathcal{B})$, hence

$$\lambda\Big(\bigcup_{i\in I_g}\mathcal{B}_i\Big)\geq(1-\theta)\lambda(\mathcal{B}).$$ □

Let $\mathcal{B}$ be a box at scale R. Let us denote the upward and downward oriented vertical geodesic segments by

$$\begin{aligned}V^{\uparrow}\mathcal{B}&:=\big\{V\in V^g\mathcal{B}\mid h(\Phi\circ V(0))\leq h(\Phi\circ V(R))\big\};\\ V^{\downarrow}\mathcal{B}&:=\big\{V\in V^g\mathcal{B}\mid h(\Phi\circ V(0))\geq h(\Phi\circ V(R))\big\}.\end{aligned}$$

We are now going to show that in a given box $\mathcal{B}_i$ with $i\in I_g$, almost all vertical geodesic segments share the same orientation.

Lemma 4.4 *For $0<\varepsilon^2\preceq_{k,c,\bowtie}\theta\preceq_{k,c,\bowtie}1$, and for $R\succeq_{k,c,\bowtie}\frac{1}{\varepsilon}$ we have that if $\mathcal{B}$ is a box at scale R such that $\eta(V^b\mathcal{B})\leq\theta\eta(V\mathcal{B})$, then one of the two following statements holds*:

(1) $\eta(V^{\uparrow}\mathcal{B}\cap V^g\mathcal{B})\geq\big(1-3\sqrt{\theta}\big)\eta(V\mathcal{B})$;

(2) $\eta(V^{\downarrow}\mathcal{B}\cap V^g\mathcal{B})\geq\big(1-3\sqrt{\theta}\big)\eta(V\mathcal{B})$.

In the proof, we first characterise a set of vertical geodesic segment whose images share the same orientation, and then we show that this set has almost full measure.

Proof Without loss of generality we can assume that $h(\mathcal{B})=[0,R[$. Let us define

$$\begin{aligned}G^Y(v^X)&:=\{v^Y\in V\mathcal{B}^Y\mid(v^X,v^Y)\in V^g\mathcal{B}\};\\ G^X&:=\big\{v^X\in V\mathcal{B}^X\mid\eta^Y(G^Y(v^X))\geq\big(1-\sqrt{\theta}\big)\eta^Y(V\mathcal{B}^Y)\big\}.\end{aligned}$$

By construction we have

$$\bigcup_{v^X\in V\mathcal{B}^X}G^Y(v^X)=(V^g\mathcal{B})^Y.$$

Applying Lemma 3.24 with $V_1:=V^g(\mathcal{B})$ and $\alpha:=\theta$ we get

$$\eta^X(G^X)\geq\big(1-\sqrt{\theta}\big)\eta^X(V\mathcal{B}^X).\tag{78}$$

Let $v_1^X:[0,R]\to X$ and $v_2^X:[0,R]\to X$ be two vertical geodesic segments of G^X. Then

$$\eta^Y(G^Y(v_1^X))\geq\big(1-\sqrt{\theta}\big)\eta^Y(V\mathcal{B}^Y);\quad\eta^Y(G^Y(v_2^X))\geq\big(1-\sqrt{\theta}\big)\eta^Y(V\mathcal{B}^Y).$$

Hence

$$\eta^Y(G^Y(v_1^X)\cap G^Y(v_2^X)) \geq \big(1-2\sqrt{\theta}\,\big)\eta^Y(V\mathcal{B}^Y). \tag{79}$$

Let $v_1^Y, v_2^Y \in G^Y(v_1^X)\cap G^Y(v_2^X)$ and let us define $V_{i,j} := (v_i^X, v_j^Y)$ with $i, j = 1, 2$. By definition of v_1^Y and v_2^Y, the quasigeodesic segments $\Phi(V_{i,j})$ are ε-monotone.

Two cases occur. As a first case let us assume that

$$d_X(v_1^X(0), v_2^X(0)) > \sqrt{\theta}R;$$
$$d_Y(v_1^Y(0), v_2^Y(0)) > \sqrt{\theta}R.$$

Let M be the constant involved in Proposition 2.13. For $R \geq 4kc$ and $\varepsilon \leq \frac{\sqrt{\theta}}{20kM}$ we have that $\sqrt{\theta}R \geq 10kM\varepsilon R + 2kc$, hence we can apply Proposition 2.13 on $V_{1,1}$ and $V_{2,2}$, which gives us that they share the same orientation.

The second case, that is, when either $d_X(v_1^X(0), v_2^X(0)) \leq \sqrt{\theta}R$ or $d_Y(v_1^Y(0), v_2^Y(0)) \leq \sqrt{\theta}R$, is treated thanks to an auxiliary geodesic segment. Hence without loss of generality we focus on the case $d_X(v_1^X(0), v_2^X(0)) \leq \sqrt{\theta}R$ and consider a geodesic segment $v_3^X \in G^X$ satisfying $d_X(v_1^X(0), v_3^X(0)) > \sqrt{\theta}R$ and $d_X(v_2^X(0), v_3^X(0)) > \sqrt{\theta}R$. To prove its existence, we consider the measure of

$$G^X \setminus V_{\mathcal{B}^X}\big(D_{\sqrt{\theta}R}(v_1^X(0)) \cup D_{\sqrt{\theta}R}(v_2^X(0))\big). \tag{80}$$

Let M_0 be the constant of assumption (E2). By Lemma 3.2 we have for all $r_1 \geq r_2 > M_0$ and for all $x \in X_0$ that $\mu_0(D_{r_1}(x)) \asymp_{\bowtie} e^{m\frac{|r_1-r_2|}{2}}\mu_0(D_{r_2}(x))$. Therefore

$$\lambda_0\big(D_{\sqrt{\theta}R}(v_1^X(0))\big) \preceq_{\bowtie} e^{m\frac{\sqrt{\theta}R-R}{2}}\lambda_0\big(D_R(v_1^X(0))\big) \leq e^{-m\frac{R}{4}}\lambda_0\big(D_R(v_1^X(0))\big) \quad \big(\text{since } \theta \leq \tfrac{1}{4}\big). \tag{81}$$

Furthermore, by Lemma 1.8 the bottom of $\mathcal{B}$ contains a disk of radius $2R - M_0$, hence by Lemma 3.2 we have $\eta^X(V\mathcal{B}^X) \asymp_X \lambda_0\big(D_{2R}(v_1^X(0))\big)$. Combined with inequality (81) we have

$$\lambda_0\big(D_{\sqrt{\theta}R}(v_1^X(0))\big) \preceq_{\bowtie} e^{-m\frac{R}{4}}\eta^X(V\mathcal{B}^X).$$

The same formula holds for v_2^X instead of v_1^X. By inequality (78) we have that

$$\eta^X(G^X) \geq \big(1-\sqrt{\theta}\,\big)\eta^X(V\mathcal{B}^X) \geq \tfrac{1}{2}\eta^X(V\mathcal{B}^X),$$

hence there exists $M(\bowtie)$ such that

$$\eta^X\big(G^X \setminus V_{\mathcal{B}^X}\big(D_{\sqrt{\theta}R}(v_1^X(0)) \cup D_{\sqrt{\theta}R}(v_2^X(0))\big)\big) \geq \big(\tfrac{1}{2} - 2Me^{-m\frac{R}{4}}\big)\eta^X(V\mathcal{B}^X)$$
$$> 0 \quad \left(\text{for } R \geq \frac{4}{m}\ln(4M+1)\right).$$

Therefore there exists $v_3^X \in G^X$ such that

$$d_X(v_1^X(0), v_3^X(0)) > \sqrt{\theta}R;$$
$$d_X(v_2^X(0), v_3^X(0)) > \sqrt{\theta}R.$$

Applying twice Proposition 2.13, first on $V_{1,1}$ and $V_{3,3}$, then on $V_{2,2}$ and $V_{3,3}$, we get that the $\Phi(V_{1,1})$ has the same orientation as $\Phi(V_{3,3})$ which has the same orientation as $\Phi(V_{2,2})$. Therefore $\Phi(V_{1,1})$ and $\Phi(V_{2,2})$ share the same orientation.

Let us fix $v_0^X \in G^X$ and $v_0^Y \in G^Y(v_0^X)$. Then the image of every vertical geodesic segment $V \in \bigcup_{v^X \in G^X} \{v^X\} \times (G^Y(v_0^X) \cap G^Y(v^X))$ shares the same orientation as the image of (v_0^X, v_0^Y). Furthermore

$$\begin{aligned}
\eta\Big(\bigcup_{v^X \in G^X} \{v^X\} \times (G^Y(v_1^X) \cap G^Y(v^X)) \Big) & \\
&= \int_{v^X \in G^X} \eta^Y(G^Y(v_1^X) \cap G^Y(v^X)) d\eta^X \\
&\geq \int_{v^X \in G^X} \big(1 - 2\sqrt{\theta}\big) \eta^Y(V\mathcal{B}^Y) d\eta^X \quad \text{(by inequality (79))} \\
&= \big(1 - 2\sqrt{\theta}\big) \eta^Y(V\mathcal{B}^Y) \eta^X(G^X) \\
&\geq \big(1 - 2\sqrt{\theta}\big) \eta^Y(V\mathcal{B}^Y) \big(1 - \sqrt{\theta}\big) \eta^X(V\mathcal{B}^X) \quad \text{(by inequality (78))} \\
&\geq \big(1 - 3\sqrt{\theta}\big) \eta(V\mathcal{B}),
\end{aligned}$$

which proves the lemma. □

4.2 Factorisation of a quasi-isometry in small boxes

Proposition 4.3 gives us two scales R and L such that all boxes at scale L can be tiled with boxes at scale R. Moreover, almost all of them, that is, the $\mathcal{B}_i$ for $i \in I_g$, contained almost only vertical geodesic segments with ε-monotone image under Φ.

A map $f : X \bowtie Y \to X' \bowtie Y'$ is called a *product map* if there exist two maps f^X and f^Y such that one of the two following holds:

(1) We have $f^X : X \to X'$, $f^Y : Y \to Y'$ and $\forall p = (p^X, p^Y) \in X \bowtie Y$, $f(p) = (f^X(p^X), f^Y(p^Y))$.

(2) We have $f^X : X \to Y'$, $f^Y : Y \to X'$ and $\forall p = (p^X, p^Y) \in X \bowtie Y$, $f(p) = (f^Y(p^Y), f^X(p^X))$.

In particular, when we denote by (f^X, f^Y) a product map on a horospherical product, it implies that when $h(x) + h(y) = 0$, we have $h(f^X(x)) + h(f^Y(y)) = 0$. Therefore a product map is height respecting.

Theorem 4.5 *For $0 < \theta \leq \varepsilon \preceq_{\bowtie} 1$, $r_0 \succeq_{\bowtie} \frac{\sqrt[\varepsilon]{2}}{\varepsilon}$, $N \succeq_{\bowtie} 1$ and for $S \succeq_{\bowtie} \frac{1}{\varepsilon\theta^2}$, we have that for any $i \in I_g$, there exists a product map $\widehat{\Phi}_i$, and $U_i' \subset \mathcal{B}_i$ such that*

(1) $\lambda(U_i') \geq (1 - \theta^{\frac{1}{8}})\lambda(\mathcal{B}_i)$;

(2) *for all $(x, y) \in U_i'$, $d_{\bowtie'}(\Phi(x, y), \widehat{\Phi}_i(x, y)) \preceq_{k,c,\bowtie} \varepsilon R$.*

In particular we have $\Delta h(\Phi(x, y), \widehat{\Phi}_i(x, y)) \preceq_{k,c,\bowtie} \varepsilon R$.

This proposition corresponds to [8, Proposition 4.14].

Since almost all the points in a good box are surrounded by almost only good vertical geodesic segment (Lemma 3.26), we show that given two points sharing the same X coordinates, we can almost always construct a quadrilateral satisfying the hypotheses of Proposition 2.11.

Lemma 4.6 *Let M_0 be the constant of assumption* (E2)*. For $0 < \theta \preceq_{\bowtie} 1$ and for $R \succeq_{\bowtie} \frac{1}{\theta}$, let $\mathcal{B}$ be a box at scale R of $X \bowtie Y$. Let us assume the existence of a subset U of $\mathcal{B}$ such that*

(a) $\lambda(U) \geq (1-\theta)\lambda(\mathcal{B})$;

(b) *for all $x \in U$, $\eta\big(V^b_{\mathcal{B}}(D_{M_0}(x))\big) \leq \sqrt{\theta}\,\eta\big(V_{\mathcal{B}}(D_{M_0}(x))\big)$.*

Then we have:

(1) *For all $a_1, a_2 \in U$ such that $a_1^X = a_2^X$, there exist $b_1, b_2 \in \mathcal{B}$ and four vertical geodesic segments $\gamma_{i,j}$ linking a_i to b_j such that a_1, a_2, b_1 and b_2 form a coarse vertical quadrilateral with nodes of scale $D = \theta R$, meaning that the configuration satisfies the assumptions of Proposition 2.11.*

(2) *For $i, j \in \{1,2\}$, $\Phi(\gamma_{i,j})$ has ε-monotone image under Φ.*

By Lemma 3.26, the boxes $\mathcal{B}_i$, with $i \in I_g$, satisfy the assumptions of this lemma. Moreover, we recall that a vertical quadrilateral satisfies the assumptions of Proposition 2.11.

Proof of Lemma 4.6 Let M_0 be the constant of assumption (E2). Let $a_1, a_2 \in U$. For $i \in \{1,2\}$ let us define $VD_i := V_{\mathcal{B}}(D_{M_0}(a_i))$ and $V^bD_i := V^b_{\mathcal{B}}(D_{M_0}(a_i))$. For all $v = (v^X, v^Y) \in V_{\mathcal{B}}$ and all $i \in \{1,2\}$ let us define

(1) $E_i^Y(v^X) := \{v^Y \in VD_i^Y \mid (v^X, v^Y) \in V^bD_i\}$;

(2) $F_i^X := \{v^X \in VD_i^X \mid \eta^Y(E_i^Y(v^X)) \geq \theta^{\frac{1}{4}}\eta^Y(VD_i^Y)\}$.

Thanks to Lemma 3.25, applied with $V_1 := V^b\mathcal{B}$, $\alpha := \sqrt{\theta}$ and $a = a_i$, we have that

$$\eta^X(F_i^X) < \theta^{\frac{1}{4}}\eta^X(VD_i^X). \tag{82}$$

Let us take a_1 and a_2 in U such that $a_1^X = a_2^X$. Then $VD_1^X = VD_2^X$, that is,

(1) $\eta^X(VD_i^X \setminus (F_1^X \cup F_2^X)) \geq (1 - 2\theta^{\frac{1}{4}})\eta^X(VD_i^X)$;

(2) for all $v^X \in VD_i^X \setminus (F_1^X \cup F_2^X)$ and $i \in \{1,2\}$ we have $\eta^Y(E_i^Y(v^X)) < \theta^{\frac{1}{4}}\eta^Y(VD_i^Y)$.

The sets $VD_i^X \setminus (F_1^X \cup F_2^X)$ enclose the vertical geodesic segments in $\mathcal{B}^X$ passing close to $a_1^X = a_2^X$ such that almost all the induced vertical geodesic segments around a_1 and a_2 in $\mathcal{B}$ are good (i.e., have ε-monotone images under the quasi-isometry Φ).

Since we have a sufficient proportion of good vertical geodesic segments, we will be able to find several of them that intersect the same neighbourhood in two different points sufficiently far from each other. If $h(a_1^X) < \theta R$, the construction of the quadrilateral of Proposition 2.11 with $D = \theta R$ is straightforward since the four points a_1, a_2, b_1 and b_2 would be θR close, hence without loss of generality we may assume that $h(a_1^X) \geq \theta R$. Moreover, as we did before we can also suppose that $h(\mathcal{B}) = [0, R[$.

We apply Lemma 1.8 with $z_0 = h(a_1)$ and $z = h(a_1) - \theta R$ to get the inclusions

$$D^X_{2\theta R - M_0}(\pi_{h(a_1)-\theta R}(a_1^X)) \subset \pi_{h(a_1)-\theta R}(D_{M_0}(a_1^X)) \subset D^X_{2\theta R + M_0}(\pi_{h(a_1)-\theta R}(a_1^X)). \tag{83}$$

We now suppose by contradiction that any couple of good vertical geodesic segments does not diverge quickly. This means that they stay M_0-close until they attain a height lower than $h(a_1^X)-\theta R$. Therefore

$$\pi_{h(a_1)-\theta R}(VD_i^X\backslash(F_1^X\cup F_2^X))\subset D^X_{M_0}(\pi_{h(a_1)-\theta R}(a_1^X)).$$

Thanks to the inclusions (83) we have $VD^X_{2\theta R-M_0}(\pi_{h(a_1)-\theta R}(a_1^X))\subset VD_1^X$, hence, combined with Property 3.23 we obtain

$$\begin{aligned}\frac{\eta^X(VD_1^X\backslash(F_1^X\cup F_2^X))}{\eta^X(VD_1^X)} &\preceq_{\bowtie} \frac{\lambda^X_{h(a_1)-\theta R}\big(D_{M_0}(\pi_{h(a_1)-\theta R}(a_1^X))\big)}{\lambda^X_{h(a_1)-\theta R}\big(D_{2\theta R}(\pi_{h(a_1)-\theta R}(a_1^X))\big)}\\ &\preceq_{\bowtie} e^{\frac{m(M_0-2\theta R)}{2}} \quad \text{(by Lemma 3.2)},\end{aligned}$$

which, for R large enough in comparison to $\frac{1}{\theta}$, contradicts the fact that

$$\eta^X(VD_1^X\backslash(F_1^X\cup F_2^X))\geq(1-2\theta^{\frac14})\eta^X(VD_1^X),$$

the first conclusion of the previously used Lemma 3.25. Hence there exists a couple of vertical geodesic segments V_1^X and V_2^X of $VD_i^X\backslash(F_1^X\cup F_2^X)$ diverging quickly from each other. Furthermore we have $\eta^Y(E_i^Y(v^X))<\theta^{\frac14}\eta^Y(VD_i^Y)$, hence there are segments V_1^Y and V_2^Y such that $(V_1^X,V_1^Y)\in V^g_{\mathcal{B}}(D_M(a_1))$ and $(V_2^X,V_2^Y)\in V^g_{\mathcal{B}}(D_M(a_2))$.

Let us define $b_i^X=V_i^X\big(h(a_1)-\frac12 d(a_1^X,a_2^X)\big)$, so that b_1^X and b_2^X are at the height where V_1^X and V_2^X diverge. Similarly, let us define $b_1^Y=b_2^Y=V_1^Y\big(-h(a_1)+\frac12 d(a_1^X,a_2^X)\big)$ such that V_1^Y and V_2^Y diverge, and $\gamma_{ij}=(V_i^X,V_j^Y)$ to ensure that the vertical geodesic segments of the quadrilateral $\gamma_{11}\cup\gamma_{12}\cup\gamma_{22}\cup\gamma_{21}$ have close endpoints. Furthermore by construction, they diverge from each other and have ε-monotone image under Φ. □

In the next proofs, we will be using Proposition 2.6 on each of the four images $\Phi(\gamma_{ij})$, which will provide us with a new quadrilateral $(\varepsilon+\theta)R$ close to $\Phi(\gamma_{11}\cup\gamma_{12}\cup\gamma_{22}\cup\gamma_{21})$ on which the assumptions of Proposition 2.11 are satisfied.

Finally we deduce that on a good box, the quasi-isometry Φ is close to a product map.

Proof of Theorem 4.5 Let $i\in I_g$ and $\mathcal{B}_i$ a good box (defined in Proposition 4.3). Then following Proposition 4.3, we have $\eta_i(V^b\mathcal{B}_i)\leq\theta\eta_i(V\mathcal{B}_i)$. Therefore by Lemma 4.4, one of the two following statements hold:

(1) $\eta(V^{\uparrow}\mathcal{B}\cap V^g\mathcal{B})\geq\big(1-3\sqrt{\theta}\,\big)\eta(V\mathcal{B})$;

(2) $\eta(V^{\downarrow}\mathcal{B}\cap V^g\mathcal{B})\geq\big(1-3\sqrt{\theta}\,\big)\eta(V\mathcal{B})$.

Let us first assume that the dominant orientation is upward. Let us choose $V_1=V\mathcal{B}\setminus(V^{\uparrow}\mathcal{B}\cap V^g\mathcal{B})$, the vertical geodesics which have neither dominant orientation nor ε-monotone image by Φ. By Lemma 3.26, used with $\alpha:=\theta^2$, we have that there exists $U_i\subset\mathcal{B}_i$ such that

(1) $\lambda(U_i)\geq(1-\sqrt{\theta})\lambda(\mathcal{B}_i)$;

(2) for $p\in U_i$ we have $\eta\big(V_1(D_{M_0}(x))\big)<\eta\big(V\mathcal{B}(D_{M_0}(x))\big)\sqrt{\theta}$.

Let us apply Lemma 3.30, with $U := U_i$ and $\alpha := \sqrt{\theta}$, then there exists $U' \subset U_i$ of almost full measure such that $\forall z \in h(U')$, $\exists (x_{0,z}, y_{0,z}) \in U'_z$ such that $\forall (x_1, y_1) \in U'_z$, we have $(x_1, y_{0,z}) \in U'$ and $(x_{0,z}, y_1) \in U'$. Let $a, a_0 \in U'$ such that $a^X = a_0^X$. By Lemma 4.6 applied on a_0 and a, there exist $b_1, b_2 \in \mathcal{B}_i$ and four vertical geodesics V_{ij} in $V^{\uparrow}\mathcal{B} \cap V^g\mathcal{B}$ such that b_1 and b_2 form a coarse vertical quadrilateral T with a_0 and a, where V_{ij} are the edges of T. Proposition 2.6 gives a constant $M(k, c, \bowtie)$ and four vertical geodesic segments $M\varepsilon R$-close to the four sides of $\Phi(T)$. Furthermore we assumed that the dominant orientation is upward, hence the images of the four sides are all upward oriented. Hence thanks to Proposition 2.11 we get

$$d_{X'}(\Phi(a_0)^{X'}, \Phi(a)^{X'}) \preceq_{k,c,\bowtie} \varepsilon R.$$

Then, for all $a \in U'$ such that $a^X = a_0^X$,

$$d_{X'}(\Phi(a_0)^{X'}, \Phi(a)^{X'}) \preceq_{k,c,\bowtie} \varepsilon R. \tag{84}$$

We show similarly that for all $a \in U'$ such that $a^Y = a_0^Y$ we have

$$d_{Y'}(\Phi(a_0)^{Y'}, \Phi(a)^{Y'}) \preceq_{k,c,\bowtie} \varepsilon R. \tag{85}$$

Let us define the product map $\widehat{\Phi}_i := (\widehat{\Phi}_i^X, \widehat{\Phi}_i^Y) : X \bowtie Y \to X' \bowtie Y'$. For all $z \in h(U')$, let $(x_{0,z}, y_{0,z}) \in U'_z$ be the points involved in Lemma 3.30, and for all $z \in [0, R[\setminus h(U')$, let us fix an arbitrary point $(x_{0,z}, y_{0,z}) \in (\mathcal{B}_i)_z$. We can therefore define, for all $x \in X$,

$$\widehat{\Phi}_i^X(x) := V^{X'}_{\Phi(x, y_{0,z})}(h \circ \Phi(x_{0,z}, y_{0,z})).$$

Then for all $(x, y) \in U'$ the triangle inequality gives

$$\begin{aligned} &d_{X'}(\widehat{\Phi}_i^X(x), \Phi(x, y)^{X'}) \\ &\qquad = d_{X'}\big(V^{X'}_{\Phi(x, y_{0,z})}(h \circ \Phi(x_{0,z}, y_{0,z})), \Phi(x, y)^{X'}\big) \\ &\qquad \le d_{X'}\big(V^{X'}_{\Phi(x, y_{0,z})}(h \circ \Phi(x_{0,z}, y_{0,z})), \Phi(x, y_{0,z})^{X'}\big) + d_{X'}(\Phi(x, y_{0,z})^{X'}, \Phi(x, y)^{X'}). \end{aligned} \tag{86}$$

Furthermore, as the distance between two points of the same vertical geodesics is equal to their difference of height, we can write

$$\begin{aligned} d_{X'}\big(V^{X'}_{\Phi(x, y_{0,z})}(h \circ \Phi(x_{0,z}, y_{0,z})), \Phi(x, y_{0,z})^{X'}\big) &= \Delta h(\Phi(x, y_{0,z})^{X'}, \Phi(x_{0,z}, y_{0,z})^{X'}) \\ &= \Delta h(\Phi(x, y_{0,z})^{Y'}, \Phi(x_{0,z}, y_{0,z})^{Y'}). \end{aligned}$$

We combine it with inequality (86), and then use the Lipschitz property of h to get

$$\begin{aligned} d_{X'}(\widehat{\Phi}_i^X(x), \Phi(x, y)^{X'}) &\le \Delta h(\Phi(x, y_{0,z})^{Y'}, \Phi(x_{0,z}, y_{0,z})^{Y'}) + d_{X'}(\Phi(x, y_{0,z})^{X'}, \Phi(x, y)^{X'}) \\ &\le d_{Y'}(\Phi(x, y_{0,z})^{Y'}, \Phi(x_{0,z}, y_{0,z})^{Y'}) + d_{X'}(\Phi(x, y_{0,z})^{X'}, \Phi(x, y)^{X'}) \\ &\preceq_{k,c,\bowtie} 2\varepsilon R \quad \text{(by inequalities (84) and (85)).} \end{aligned}$$

Similarly, we define $\widehat{\Phi}_i^Y(y)$ by

$$\widehat{\Phi}_i^Y(y) := V^{Y'}_{\Phi(x_{0,z},y)}(h \circ \Phi(x_{0,z}, y_{0,z})),$$

and we show that $d_{Y'}(\widehat{\Phi}_i^Y(y), \Phi(x,y)^{Y'}) \preceq_{k,c,\bowtie} \varepsilon R$. Furthermore for all $(x,y) \in U_i$ we have the equality $h(\widehat{\Phi}_i^X(x)) = -h(\widehat{\Phi}_i^Y(y))$, hence $\widehat{\Phi}_i := (\widehat{\Phi}_i^X, \widehat{\Phi}_i^Y) : X \bowtie Y \to X' \bowtie Y'$ is a well-defined product map. Then we chose $U_i' := U'$ to conclude the proof.

The downward orientation case is dealt in the same way by switching the definitions of $\widehat{\Phi}_i^X$ and $\widehat{\Phi}_i^Y$. □

4.3 Shadows and orientation

We use the fact that $m > n$ to prove that Φ is orientation preserving, hence the upward orientation is dominant on each good box at scale R.

Proposition 4.7 *Assume that $m > n$ and that $m' > n'$. For $R \succeq_\bowtie \frac{1}{\theta}$ the product map $\widehat{\Phi}_i$ of Theorem 4.5 is orientation preserving for each $i \in I_g$.*

We recall that given a box $\mathcal{B}$, the shadow of a subset $U \subset \mathcal{B}$, denoted by $\mathrm{Sh}(U)$, is the set of points of $\mathcal{B}$ below U in the sense

$$\mathrm{Sh}(U) := \{p \in \mathcal{B} \mid \exists V \in V\mathcal{B} \text{ containing } p \text{ and intersecting } U \text{ on a point } p' \text{ such that } h(p') \geq h(p)\}.$$

And we remind the reader that given a subset $S \subset X$, the large Y-horosphere given by S and denoted by $H_S \subset X \bowtie Y$ is the set

$$H_S := S \bowtie Y.$$

Let us define $\mathcal{B} = \mathcal{B}_i$ for $i \in I_g$. Thanks to Theorem 4.5, there exist $U = U_i$ with $\lambda(U) \geq (1 - \theta^{\frac{1}{4}})\lambda(\mathcal{B})$ such that Φ is close to a product map on U. We consider two parameters ρ_1 and ρ_2 with $1 \succeq_\bowtie \rho_2 \succeq_\bowtie \rho_1 \succeq_\bowtie \theta^{\frac{1}{16}}$. The relations between them will be specified later. Hence Lemma 3.28 applies with $\alpha = \theta^{\frac{1}{4}}$, and it gives us a Y-horosphere H_{x_0} such that

$$\lambda(\mathrm{Sh}(H_{D_{M_0}(x_0)}) \cap U^c) < \theta^{\frac{1}{16}} \lambda(\mathrm{Sh}(H_{D_{M_0}(x0)})).$$

Then we apply twice Lemma 3.29 with $\alpha = \theta^{\frac{1}{4}}$, and $\rho = \rho_i$ for $i \in \{1,2\}$ to get two level sets of h in $\mathcal{B}$, P_1 and P_2 (see Figure 17), such that, for $i \in \{1,2\}$,

$$\lambda_{h(P_i)}(P_i \cap \mathrm{Sh}(H_{D_{M_0}(x_0)}) \cap U^c) \preceq_\bowtie \theta^{\frac{1}{16}} \lambda_{h(P_i)}(P_i \cap \mathrm{Sh}(H_{D_{M_0}(x_0)})),$$

and such that $\rho_i R < \Delta h(P_i, H_{x_0}) < 2\rho_i R$.

The next lemma will gives us the existence of two subsets below a Y-horosphere H, which are sufficiently big (for the measure μ in comparison to the horosphere) and sufficiently apart from each other so that any path linking them must get close to H.

This lemma is strongly inspired from [8, Lemma 5.9].

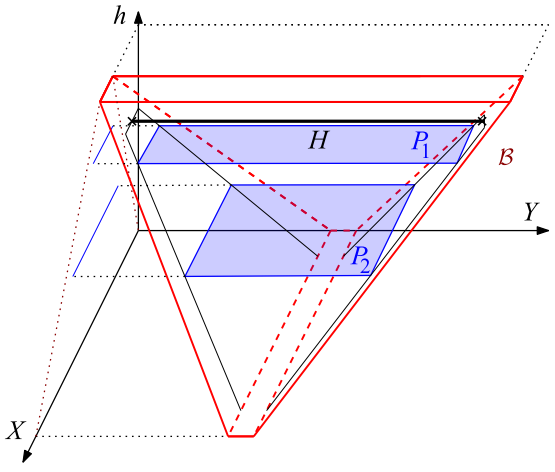


Figure 17: Configuration of Lemma 4.8.

Lemma 4.8 *There exists a constant $M_1(k, c, \bowtie)$ depending on k, c and on the metric measured spaces $X \bowtie Y$ with the following property. In the settings above, for $R \succeq_\bowtie \frac{1}{\rho_2}$, there exist S_1 and S_2, two subsets of $P_2 \cap \mathcal{B}$ such that for $j \in \{1, 2\}$ we have*

(1) $\forall s_1 \in S_1, s_2 \in S_2, d_X(s_1^X, s_2^X) \geq \rho_2 R$;

(2) $\lambda_{h(P_2)}(S_j \cap U^c) \preceq_\bowtie \theta^{\frac{1}{32}} \lambda_{h(P_2)}(S_j)$;

(3) $\mu_{h(P_2)}(S_j) \succeq_\bowtie \exp\big(\frac{m-n}{2}\rho_2 R\big)\mu_{h(H)}(\mathcal{N}_{M_0}(H))$;

(4) *any path γ joining S_1 and S_2 of length $l(\gamma) \leq M_1 \rho_2 R$ intersects $\mathcal{N}_{6\rho_1 R}(H)$.*

Proof For $j \in \{1, 2\}$, let us define $Q_j := P_j \cap \mathrm{Sh}(H_{D_{M_0}(x_0)})$. We tile Q_1^X with the top of boxes as in a box tiling. More precisely, let M_0 be the constant involved in assumption (E2), and let $Z \subset Q_1^X$ be an $2M_0$-maximal separating set of Q_1^X. Then there exists a set of disjoint cells $\{\mathcal{C}(x) \mid x \in Z\}$ such that

(1) $\forall x \in Z$, $D(x, M_0) \subset \mathcal{C}(x) \subset D(x, 2M_0)$;

(2) $Q_1^X = \bigcup_{x \in Z} \mathcal{C}(x)$.

Thanks to this tessellation, we tile Q_1 with the large horosphere $H_{\mathcal{C}(x)} := \mathcal{C}(x) \times \mathcal{B}^Y_{-h(P_1)} = \mathcal{C}(x) \times Q_1^Y$. Furthermore, for any two points $x_1, x_2 \in Z$,

$$\begin{aligned}\lambda_{h(P_1)}(H_{\mathcal{C}(x_1)}) &= \lambda^X_{h(P_1)}(\mathcal{C}(x_1))\lambda^Y_{-h(P_1)}(\mathcal{B}^Y_{-h(P_1)}) \\ &\asymp_\bowtie \lambda^X_{h(P_1)}(\mathcal{C}(x_2))\lambda^Y_{-h(P_1)}(\mathcal{B}^Y_{-h(P_1)}) \quad \text{(by Lemma 3.2)} \\ &= \lambda_{h(P_1)}(H_{\mathcal{C}(x_2)}).\end{aligned}$$

Therefore $\lambda_{h(P_1)}(Q_1) \asymp_\bowtie \lambda^Y_{-h(P_1)}(Q_1^Y)\#Z$. We tile Q_2 by projections of the tessellation of Q_1. These projections look like stripes on Q_2:

$$Q_2 = \bigsqcup_{x \in Z} \pi^X_{h(P_2)}(\mathcal{C}(x)) \times \mathcal{B}^Y_{-h(P_2)}. \tag{87}$$

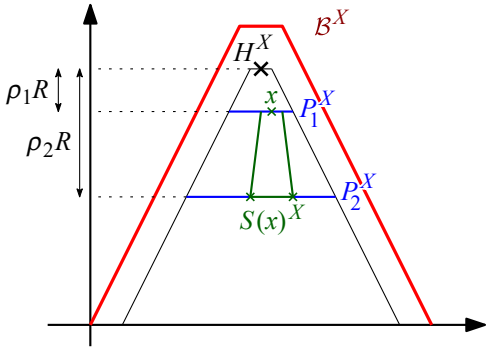


Figure 18: Construction of $S(x)^X$ in Lemma 4.8.

Let us denote these stripes by $S(x) := \pi^X_{h(P_2)}(\mathcal{C}(x)) \times \mathcal{B}^Y_{-h(P_2)}$ for all $x \in Z$ (see Figure 18). For all $x_1, x_2 \in Z$, $d_X(x_1, x_2) \geq M_0$, hence by Lemma 1.3, for all $(s_1^X, s_1^Y) \in \mathrm{Int}_{M_0}(S(x_1))$ and for all $(s_2^X, s_2^Y) \in \mathrm{Int}_{M_0}(S(x_2))$, we have

$$d_X(s_1^X, s_2^X) \geq 2\Delta h(P_1, P_2) - M_0 = 2\rho_2 R - 2\rho_1 R - M_0 - M \tag{88}$$

$$\geq 2(\rho_2 - 2\rho_1)R \quad \left(\text{for } R \geq \frac{2(M_0 + M)}{\rho_1}\right). \tag{89}$$

Furthermore we have by construction that

$$\lambda^X_{h(P_2)}\big(\pi^X_{h(P_2)}(\mathcal{C}(x_1))\big) \asymp_{\bowtie} \lambda^X_{h(P_2)}\big(\pi^X_{h(P_2)}(\mathcal{C}(x_2))\big).$$

Hence, combined with Lemma 3.3, we get

$$\lambda_{h(P_2)}\big(\mathrm{Int}_{M_0}(S(x_1))\big) \asymp_{\bowtie} \lambda_{h(P_2)}(S(x_1)) \asymp_{\bowtie} \lambda_{h(P_2)}(S(x_2)) \asymp_{\bowtie} \lambda_{h(P_2)}\big(\mathrm{Int}_{M_0}(S(x_2))\big).$$

Therefore, by the tessellation (88), $\lambda_{h(P_2)}(Q_2) \asymp_{\bowtie} \lambda^Y_{-h(P_2)}(Q_2^Y)\#Z$. By Lemma 3.29, used with $\alpha := \theta^{\frac{1}{4}}$, we get

$$\lambda_{h(P_2)}(Q_2 \cap U^c) \preceq_{\bowtie} \theta^{\frac{1}{16}} \lambda_{h(P_2)}(Q_2).$$

Moreover, for all $x_1, x_2 \in Z$ we have $\lambda_{h(P_2)}(S(x_1)) \asymp_{\bowtie} \lambda_{h(P_2)}(S(x_2))$ and the set of stripes $S(x)$ for $x \in Z$ tile the set Q_2. Therefore there exists $Z' \subset Z$ such that $\#Z' \geq (1 - \theta^{\frac{1}{32}})\#Z$ and such that for all $x \in Z'$ we have $\lambda_{h(P_2)}(S(x) \cap U^c) \leq \theta^{\frac{1}{32}} \lambda_{h(P_2)}(S(x))$.

We are now able to define S_1 and S_2. Let $x_1, x_2 \in Z$ be distinct and, for $j \in \{1, 2\}$, let us denote by S_j the following subset of $S(x_j)$:

$$S_j := \pi^X_{h(P_2)}(\mathcal{C}(x_j)) \times \mathrm{Int}_{M\rho_2 R}(\mathcal{B}^Y_{-h(P_2)}). \tag{90}$$

By Lemma 3.3, applied with $r = M\rho_2 R$, $z_0 = -h^-(\mathcal{B})$ and $z_1 = -h(P_2)$, we have

$$\mu^Y_{h(P_2)}(\mathcal{B}^Y_{-h(P_2)}) \asymp_{\bowtie} \mu^Y_{h(P_2)}(\mathrm{Int}_{M\rho_2 R}(\mathcal{B}^Y_{-h(P_2)})).$$

Therefore

$$\mu_{h(P_2)}(S_j) \asymp_{\bowtie} \mu_{h(P_2)}(S(x_j)). \tag{91}$$

The first point of the lemma holds by inequality (89), and the second point holds because we choose x_1 and x_2 in Z'.

Let us now prove the third point. Let $y_0 \in Y$ be the nucleus of the cell of $\mathcal{B}^Y$. We have that $\mathcal{B}^Y_{-z} := \pi^Y_{-z}(\mathcal{C}(y_0))$. Let us define $h^- := h^-(\mathcal{B})$. By Lemma 1.8 applied with $p = y_0$, $z_0 = -h^-$ and $z = -(h(H) - \rho_2 R) = -h(P_2)$ we have

$$D_{2|h^- - h(P_2)| - M_0}(\pi^Y_{-h(P_2)}(y_0)) \subset \mathcal{B}^Y_{-h(P_2)} \subset D_{2|h^- - h(P_2)| + M_0}(\pi^Y_{-h(P_2)}(y_0)).$$

It follows that, for $x \in Z$,

$$\pi^X_{h(P_2)}(\mathcal{C}(x)) \times D^Y_{2(|h^- - h(H)| + \rho_2 R) - M_0}(\pi^Y_{-h(P_2)}(y_0)) \subset S(x) \subset \pi^X_{h(P_2)}(\mathcal{C}(x)) \times D^Y_{2(|h^- - h(H)| + \rho_2 R) + M_0}(\pi^Y_{-h(P_2)}(y_0)).$$

By Lemma 1.8, $\pi^X_{h(P_2)}(\mathcal{C}(x))$ resembles a disk of radius $2|h(P_1) - h(P_2)| \pm M_0 = 2(\rho_2 - \rho_1)R \pm M_0$. Lemma 3.2 gives $\mu^X_{h(P_2)}\big(\pi^X_{h(P_2)}(\mathcal{C}(x))\big) \asymp e^{m(\rho_2 - \rho_1)R}$. Again by Lemma 3.2 applied on

$$D^Y_{2(|h^- - h(H)| + \rho_2 R) + M_0}(\pi^Y_{-h(P_2)}(y_0)),$$

we have

$$\mu_{h(P_2)}(S(x)) \asymp_{\bowtie} e^{m(\rho_2 - \rho_1)R} e^{n(|h^- - h(H)| + \rho_2 R)}.$$

Similarly Q_2 resembles a product $D_{2\rho_2 R \pm M_0} \times B^Y_{-h(P_2)}$, hence

$$\mu_{h(P_2)}(Q_2) \asymp_{\bowtie} e^{m\rho_2 R} e^{n(|h^- - h(H)| + \rho_2 R)}.$$

Therefore we obtain an estimate, of $\#Z$,

$$\frac{\mu_{h(P_2)}(Q_2)}{\mu_{h(P_2)}(S(x))} \asymp_{\bowtie} e^{m\rho_1 R}. \tag{92}$$

Applying Lemma 3.32 with $A = Q_2$, $U = \mathcal{N}_{M_0}(H)$ and $\Delta = \rho_2 R$ gives

$$\mu_{h(P_2)}(Q_2) \succeq_{\bowtie} \exp((m-n)\rho_2 R)\mu_{h(H)}(\mathcal{N}_{M_0}(H)).$$

In combination with inequalities (91) and (92) we have, for $j \in \{1, 2\}$,

$$\begin{aligned}\mu_{h(P_2)}(S_j) &\succeq_{\bowtie} \exp((m-n)\rho_2 R - m\rho_1 R)\mu_{h(H)}(\mathcal{N}_{M_0}(H)) \\ &\succeq_{\bowtie} \exp\left(\frac{m-n}{2}\rho_2 R\right)\mu_{h(H)}(\mathcal{N}_{M_0}(H)),\end{aligned}$$

where the last inequality holds since $(m-n)\rho_2 - m\rho_1 \geq \frac{m-n}{2}\rho_2$ when $\rho_1 \leq \frac{m-n}{m}\rho_2$. Therefore the third conclusion of this lemma holds.

Let us prove the fourth conclusion. Let γ be a path joining $s_1 \in S_1$ and $s_2 \in S_2$ such that $l(\gamma) \leq M\rho_2 R$. By inequality (89), $d_X(s_1^X, s_2^X) \geq 2\rho_2 R - 4\rho_1 R$. By Lemma 1.5 there exists a constant $M'(\delta)$ such that the geodesic segment $[s_1^X, s_2^X]$ contains a point s_3^X within $4\rho_1 R - M'(\delta) \leq 5\rho_1 R$ of $H^X = \{x_0\}$, for $R \geq \frac{M'(\delta)}{\rho_1}$. Therefore by Proposition 1.12

$$l(\gamma^X) \geq 2^{\delta d_X(\gamma^X, s_3^X)}.$$

However, every δ-hyperbolic space with $\delta \leq 1$ is also 1-hyperbolic. Therefore we can assume without loss of generality that $\delta \geq 1$. Then we have

$$l(\gamma^X) \geq 2^{d_X(\gamma^X, s_3^X)} \geq 2^{d_X(\gamma^X, H^X) - 5\rho_1 R}.$$

Hence $\log_2(M\rho_2 R) \geq d(\gamma^X, H^X) - 5\rho_1 R$. Furthermore, there exists $M'(k, c, \bowtie)$ such that for $R \geq \frac{M'}{\rho_2}$ we have $\log_2(M\rho_2 R) \leq \rho_1 R$. In this case

$$d(\gamma^X, H^X) \leq 6\rho_1 R.$$

Therefore there exists $t \in \mathbb{R}$ such that $\Delta h(\gamma(t), H) \leq 6\rho_1 R$. Let us now look at γ^Y. Two cases arise, we have either $\gamma^Y(t) \in \mathrm{Sh}(\mathcal{B}^Y_{-h(P_2)})$ or $\gamma^Y(t) \notin \mathrm{Sh}(\mathcal{B}^Y_{-h(P_2)})$.

In the first case, there exists $y \in H^Y$ such that $\gamma^Y(t) \in V_y$. Furthermore $\Delta h(\gamma(t), H) \leq 6\rho_1 R$, hence $d_Y(\gamma^Y(t), H^Y) = \Delta h(\gamma^Y(t), H^Y) \leq 6\rho_1 R$ and consequently $d_Y(\gamma^Y, H^Y) \leq 6\rho_1 R$. Which proves $d(\gamma, H) \leq 6\rho_1 R$.

In the second case, when $\gamma^Y(t) \notin \mathcal{B}^Y_{-h(P_2)}$, by our claim (90) we have that the vertical geodesic ray $V_{\gamma^Y(t)}$ starting at $\gamma^Y(t)$ intersect $Y_{-h(P_2)}$ in a point y such that $d_Y(y, S_1^Y \cup S_2^Y) > M\rho_2 R$. Therefore

$$\begin{aligned} M\rho_2 R \geq l(\gamma) \geq \tfrac{1}{2} l(\gamma^Y) &\geq \tfrac{1}{2}\big(d(s_1, \gamma(t)) + d(\gamma(t), s_2)\big) \\ &> \frac{2M\rho_2 R}{2} > M\rho_2 R, \end{aligned}$$

which is absurd, hence the second case when $\gamma^Y(t) \notin \mathcal{B}^Y_{-h(P_2)}$ does not occur. Therefore we always have that γ intersect the $6\rho_1 R$-neighbourhood of H. □

Proof of Proposition 4.7 Let us be in the settings above. Let us assume by contradiction that $\widehat{\Phi}$ is orientation reversing, which means that there exist $\widehat{\Phi}^X : X \to Y'$ and $\widehat{\Phi}^X : Y \to X'$ such that for all $(x, y) \in \mathcal{B}$ we have $\widehat{\Phi}(x, y) = (\widehat{\Phi}^Y(y), \widehat{\Phi}^X(x))$.

For all $p \in X' \bowtie Y'$ such that $d_{\bowtie'}(p, \widehat{\Phi}(H \cap U)) \leq \rho_1 R$ there exists an element $q \in H \cap U$ such that $d_{\bowtie'}(p, \widehat{\Phi}(q)) \leq \rho_1 R$. Therefore by the triangle inequality

$$\begin{aligned} d_{\bowtie'}(p, \Phi(q)) &\leq d_{\bowtie'}(p, \widehat{\Phi}(q)) + d_{\bowtie'}(\widehat{\Phi}(q), \Phi(q)) \preceq_{k,c,\bowtie} \rho_1 R + \varepsilon R \quad \text{(by Theorem 4.5 since } q \in U) \\ &\preceq_{k,c,\bowtie} \rho_1 R \quad \text{(since } \varepsilon \leq \rho_1). \end{aligned}$$

Hence there exists $M(k,c,\bowtie)$ such that $\mathcal{N}_{\rho_1 R}(\widehat{\Phi}(H\cap U))\subset\mathcal{N}_{M\rho_1 R}(\Phi(H\cap U))$. We show similarly that, for $j\in\{1,2\}$,

$$\mathcal{N}_{\rho_1 R}(\Phi(S_j\cap U))\subset\mathcal{N}_{M\rho_1 R}(\widehat{\Phi}(S_j\cap U)). \tag{93}$$

Let $M'(\bowtie)$ be the constant involved in Corollary 3.17. Then

$$\begin{aligned}
\mu\big(\mathcal{N}_{8k\rho_1 R}(\Phi(H))\big) &\preceq_{k,c,\bowtie} e^{48k\rho_1 Rm'}\mu\big(\mathcal{N}_{kc+c}(\Phi(H))\big) \quad \text{(by Corollary 3.16)}\\
&\preceq_{k,c,\bowtie} e^{48k\rho_1 Rm'}\mu(\mathcal{N}_1(H)) \quad \text{(by Lemma 3.18)}\\
&\le e^{48k\rho_1 Rm'}\mu(\mathcal{N}_{M'}(H))\\
&\asymp_{\bowtie} e^{48k\rho_1 Rm'}\mu_{h(H)}(\mathcal{N}_{M'}(H)) \quad \text{(by the second part of Corollary 3.17)}\\
&\preceq_{\bowtie} e^{48k\rho_1 Rm'}\mu_{h(H)}(\mathcal{N}_{M_0}(H)) \quad \text{(by the first part of Corollary 3.17).}
\end{aligned}$$

Combined with point 3 of Lemma 4.8 we have

$$\begin{aligned}
\mu(\mathcal{N}_{8k\rho_1 R}(\Phi(H))) &\preceq_{\bowtie} e^{-(m-n)\frac{\rho_2}{2}R}e^{48k\rho_1 Rm'}\mu_{h(P_2)}(S_j)\\
&\preceq_{\bowtie} e^{-(m-n)\frac{\rho_2}{2}R}e^{48k\rho_1 Rm'}\mu_{h(P_2)}(S_j\cap U) \quad \text{(thanks to Lemma 4.8(2))}\\
&\preceq_{\bowtie} e^{-(m-n)\frac{\rho_2}{4}R}\mu_{h(P_2)}(\mathcal{N}_1(S_j\cap U)) \quad \left(\text{since } \rho_1\le\frac{m-n}{96km'}\rho_2\right)\\
&\asymp_{\bowtie} e^{-(m-n)\frac{\rho_2}{4}R}\mu(\mathcal{N}_{M'}(S_j\cap U)) \quad \text{(by Corollary 3.17)}\\
&\le e^{-(m-n)\frac{\rho_2}{4}R}\mu(\mathcal{N}_{M'+kc+c}(S_j\cap U)).
\end{aligned}$$

Hence, using Lemma 3.18 on $\mathcal{N}_{M'}(S_j\cap U)$,

$$\begin{aligned}
&\mu\big(\mathcal{N}_{8k\rho_1 R}(\Phi(H))\big)\\
&\quad\preceq_{k,c,\bowtie} e^{-(m-n)\frac{\rho_2}{4}R}\mu\big(\mathcal{N}_{M'+1}(\Phi(S_j\cap U))\big)\\
&\quad\le e^{-(m-n)\frac{\rho_2}{4}R}\mu\big(\mathcal{N}_{\rho_1 R}(\Phi(S_j\cap U))\big) \quad \left(\text{for } R\ge\frac{M'}{\rho_1}\right)\\
&\quad\le e^{-(m-n)\frac{\rho_2}{4}R}\mu\big(\mathcal{N}_{M\rho_1 R}(\widehat{\Phi}(S_j\cap U))\big) \quad \text{(by inequality (93))}\\
&\quad\preceq_{k,c,\bowtie} e^{-(m-n)\frac{\rho_2}{4}R}e^{6M\rho_1 Rm'}\mu\big(\mathcal{N}_{M'}(\widehat{\Phi}(S_j\cap U))\big) \quad \text{(by Corollary 3.17)}\\
&\quad\preceq_{k,c,\bowtie} e^{-(m-n)\frac{\rho_2}{8}R}\mu\big(\mathcal{N}_{M'}(\widehat{\Phi}(S_j\cap U))\big) \quad \left(\text{since } \rho_1\le\frac{m-n}{48Mm'}\rho_2\right)\\
&\quad\asymp_{k,c,\bowtie} e^{-(m-n)\frac{\rho_2}{8}R}\mu_{\widehat{z_0}}\big(\mathcal{N}_{M'}(\widehat{\Phi}(S_j\cap U))\cap X'_{\widehat{z_0}}\big) \quad \text{(by the second part of Corollary 3.17),}
\end{aligned}$$

where $\widehat{z_0}:=\widehat{\Phi}(P_2)$. Since $\widehat{\Phi}$ is orientation reversing, we can now apply Lemma 3.33 with $A_j=\widehat{\Phi}(S_j\cap U)$, $E=\mathcal{N}_{8k\rho_1 R}(\Phi(H))$ and $Q=e^{(m-n)\frac{\rho_2}{8}R}$ we have that

$$\eta\big(V\mathcal{N}_{M_0}(\widehat{\Phi}(S_j\cap U))\big)\succeq_{k,c,\bowtie} e^{(m-n)\frac{\rho_2}{8}R}\eta(V\mathcal{N}_{M_0}(E)).$$

Then, as pointed out below Lemma 3.26, we can apply it on a A_j with $V_1 = VE$. Hence let us take $U_{A_j} \subset A_j$ maximal for the inclusion such that

- $\lambda_{\widehat{z_0}}(U_{A_j}) \geq (1 - e^{(m-n)\frac{\rho_2}{8}R})\lambda_{\widehat{z_0}}(A_j)$;
- for all $p \in U_{A_j}$, most of the vertical geodesic in $D_{M_0}(p)$ do not intersect E.

By Property 3.23 we have

$$\lambda_{\widehat{z_0}}\big(\mathcal{N}_{M_0}(\widehat{\Phi}(S_j \cap U))\big) \succeq_{k,c,\bowtie} e^{(m-n)\frac{\rho_2}{8}R}\lambda_{\widehat{z_0}}\big(\pi^{\bowtie}_{\widehat{z_0}}(\mathcal{N}_{M_0}(E))\big).$$

Hence, by the definition of $\lambda_{\widehat{z_0}}$,

$$\mu_{\widehat{z_0}}\big(\mathcal{N}_{M_0}(\widehat{\Phi}(S_j \cap U))\big) \succeq_{k,c,\bowtie} e^{(m-n)\frac{\rho_2}{8}R}\mu_{\widehat{z_0}}\big(\pi^{\bowtie}_{\widehat{z_0}}(\mathcal{N}_{M_0}(E))\big). \tag{94}$$

Let us define $E' := \mathcal{N}_{M_0}(\widehat{\Phi}(S_j \cap U) \setminus U_{A_j})$. By the construction of U_{A_j}, $E' \cap X'_{z_0}$ is of almost full measure in $\pi^{\bowtie}_{\widehat{z_0}}(\mathcal{N}_{M_0}(E))$. Furthermore, by Theorem 4.5 $\widehat{\Phi}$, is $M\varepsilon R$-close to Φ on U, hence we have (similarly as in inequality (93)) that

$$\mathcal{N}_{\rho_1 R}(\widehat{\Phi}^{-1}(E')) \subset \mathcal{N}_{M\rho_1 R}(\Phi^{-1}(E')).$$

Therefore

$$\begin{aligned}
\mu\big(\mathcal{N}_{\rho_1 R}(\widehat{\Phi}^{-1}(E'))\big) &\preceq_{k,c,\bowtie} \mu\big(\mathcal{N}_{M\rho_1 R}(\Phi^{-1}(E'))\big)\\
&\preceq_{k,c,\bowtie} e^{6M\rho_1 Rm}\mu\big(\mathcal{N}_{kc+c}(\Phi^{-1}(E'))\big) \quad \text{(by the first part of Corollary 3.17)}\\
&\asymp_{k,c,\bowtie} e^{6M\rho_1 Rm}\mu(\mathcal{N}_1(E')) \quad \text{(by Lemma 3.18)}\\
&\asymp_{k,c,\bowtie} e^{6M\rho_1 Rm}\mu_{\widehat{z_0}}(\mathcal{N}_{M_0}(E')) \quad \text{(by the second part of Corollary 3.17)}\\
&\asymp_{k,c,\bowtie} e^{6M\rho_1 Rm}\mu_{\widehat{z_0}}\big(\pi^{\bowtie}_{\widehat{z_0}}(\mathcal{N}_{M_0}(E))\big)\\
&\preceq_{k,c,\bowtie} e^{-(m-n)\frac{\rho_2}{8}R}e^{6M\rho_1 Rm}\mu_{\widehat{z_0}}\big(\mathcal{N}_{M_0}(\widehat{\Phi}(S_j \cap U))\big) \quad \text{(by the definition of } U_{A_j})\\
&\preceq_{k,c,\bowtie} e^{-(m-n)\frac{\rho_2}{16}R}\mu_{h(P_2)}(\mathcal{N}_{M_0}(S_j \cap U)) \quad \left(\text{since } \rho_1 \leq \frac{\rho_2}{M}\right)\\
&\leq e^{-(m-n)\frac{\rho_2}{16}R}\mu_{h(P_2)}(\mathcal{N}_{M_0}(S_j)).
\end{aligned}$$

Following the second conclusion of Lemma 4.8, there exists a constant $M(\bowtie)$ such that $\lambda_{h(P_2)}(S_j \cap U^c) \leq M\theta^{\frac{1}{32}}\lambda_{h(P_2)}(S_j)$.

Next, we apply twice Lemma 3.24 for $j = 1, 2$ with $(V_1, \eta) = (\mathcal{N}_{M_0}(S_j^X) \times \mathcal{N}_{M_0}(S_j^Y), \mu_{h(P_2)})$, $V_0 = U^c \cap \mathcal{N}_{\rho_1 R}(\widehat{\Phi}^{-1}(E'))$ and $\alpha := e^{(m-n)\frac{\rho_2}{16}R}\mu_{h(P_2)} + M\theta^{\frac{1}{32}}$. Let us define

$$G^Y(p^X) := \{p^Y \in V_1^Y \mid (p^X, p^Y) \in V_0\}.$$

We have that

$$\mu^X_{h(P_2)}\big(\{p^X \in V_1^X \mid \mu^Y_{-h(P_2)}(G^Y(p^X))\}\big) \geq (1 - e^{-(m-n)\frac{\rho_2}{32}R})\mu^Y_{h(P_2)}(V_1^Y).$$

Since $e^{-(m-n)\frac{\rho_2}{32}R} + M\theta^{\frac{1}{32}} < \frac{1}{2}$, there exists $s_1 \in (S_1 \cap U) \setminus \widehat{\Phi}^{-1}(E')$ and $s_2 \in (S_2 \cap U) \setminus \widehat{\Phi}^{-1}(E')$ such that $s_1^Y = s_2^Y$.

Let us define $\hat{s}_j := \widehat{\Phi}(s_j)$ for $j \in \{1,2\}$. By construction we have $\hat{s}_j \in A_j$. Then $VD_{M_0}(\widehat{s_j})$ contains almost only vertical geodesic segments which do not intersect E. Since $\hat{s}_1^{X'} = \hat{s}_2^{X'}$, and by Lemma 3.25, we can find two vertical geodesics $v_1 \in VD_{M_0}(\widehat{s_1})$ and $v_2 \in VD_{M_0}(\widehat{s_2})$ which do not intersect $E = \mathcal{N}_{8k\rho_1 R}(\Phi(H))$, and such that $v_1^X = v_2^X$. Since v_1^Y and v_2^Y meet (up to an additive constant) at the height $-\hat{z}_0 + \frac{1}{2}d_{Y'}(\hat{s}_1^{Y'}, \hat{s}_2^{Y'})$, there exist $M(\delta)$ such that the concatenation of v_1 and v_2 is $(1, M(\delta))$-quasigeodesic linking $\hat{s}_1$ to $\hat{s}_2$.

Let us define $\gamma := \Phi^{-1}(v_1 \cup v_2)$. Then γ is a $(k, c+M)$-quasigeodesic. By [13, Lemma 2.1], there exists a $2k$-Lipschitz, $(k, 4(M+c))$-quasi-geodesic γ' in the $2(M+c)$-neighbourhood of γ, linking $\Phi^{-1}(\hat{s}_1)$ to $\Phi^{-1}(\hat{s}_2)$. Let us define $s'_1 = \Phi^{-1}(\hat{s}_1)$ and $s'_2 = \Phi^{-1}(\hat{s}_2)$. Because γ' is $2k$-Lipschitz, and since Φ^{-1} is a (k,c)-quasi-isometry we have

$$l(\gamma') \leq 2k d_{\bowtie'}(\hat{s}_1, \hat{s}_2) \leq k^2 d_{\bowtie}(s'_1, s'_2) + c. \tag{95}$$

Furthermore, γ' does not intersect the $\frac{1}{k}(7k\rho_1 R - 2c) - c$-neighbourhood of H since Φ^{-1} is a quasi-isometry. Moreover s'_j and s_j are εR close to each other, that is,

$$\begin{aligned} d_{\bowtie}(s'_j, s_j) &= d_{\bowtie}\big(\Phi^{-1}(\widehat{\Phi}(s_j)), s_j\big) \\ &\leq k d_{\bowtie'}(\widehat{\Phi}(s_j), \Phi(s_j)) \preceq_{k,c,\bowtie} \varepsilon R \quad (\text{since } s_j \in U). \end{aligned} \tag{96}$$

Consequently by the triangle inequality we get

$$\begin{aligned} d_{\bowtie}(s'_1, s'_2) &\leq d_{\bowtie}(s'_1, s_1) + d_{\bowtie}(s_1, s_2) + d_{\bowtie}(s_2, s'_2) \\ &\preceq_{k,c,\bowtie} \varepsilon R + d_{\bowtie}(s_1, s_2) \quad (\text{since } \widehat{\Phi}^{-1}(s_j) \in U). \end{aligned} \tag{97}$$

Furthermore $s_1^Y = s_2^Y$. Therefore by Theorem 1.11, with $M = 15C_0$ we obtain

$$d_{\bowtie}(s_1, s_2) \leq d_X(s_1^X, s_2^X) + M \leq 2\rho_2 R + M \quad (\text{by the first point of Lemma 4.8}).$$

Combined with inequalities (95) and (97) we get

$$l(\gamma') \preceq_{k,c,\bowtie} 2k^2(2\rho_2 R + M + 2\varepsilon R) + c \preceq_{k,c,\bowtie} \rho_2 R \quad \left(\text{for } R \geq \frac{M+c}{\rho_2}\right).$$

For $j \in \{1,2\}$, let $\gamma_j := [s_j, s'_j]$, by inequality (96) we have $l(\gamma_j) \preceq_{k,c,\bowtie} \varepsilon R$. Hence the path γ'', constructed as the concatenation of γ_1, γ' and γ_2, is a path linking $s_1 \in S_1$ to $s_2 \in S_2$, of length $l(\gamma) \preceq_{k,c,\bowtie} \rho_2 R$ since $\varepsilon \leq \rho_2$. Furthermore, by construction, γ'' does not intersect the $7\rho_1 R - 3c - 2M\varepsilon R > 6\rho_1 R$-neighbourhood of H. This contradicts the fourth point of Lemma 4.8, therefore Φ is orientation preserving. □

4.4 Factorisation of a quasi-isometry in big boxes

In Section 4.2 we proved that for all $i \in I_g$, $\Phi_{|\mathcal{B}_i}$ is close to a quasi-isometry product $\widehat{\Phi}_i = (\widehat{\Phi}_i^X, \widehat{\Phi}_i^Y)$ on a set of almost full measure $U_i \subset \mathcal{B}_i$. In this section we prove that Φ is close to $\widehat{\Phi}$ on all boxes at scale L on a set of almost full measure. This is a step forward since this is true on all boxes at scale L and not only a significant number of them.

Theorem 4.9 *For $0 < \theta \preceq_{k,c,\bowtie} 1$ there exists $L_0(k,c,\bowtie,\theta) > 0$ such that for all $L \geq L_0$ and for all box $\mathcal{B}$ at scale L, there exists $M(k,c,\bowtie)$, $U \subset \mathcal{B}$ and a $\big(k, M\sqrt{\theta}L\big)$-quasi-isometry product map $\widehat{\Phi} = (\widehat{\Phi}^X, \widehat{\Phi}^Y)$, with $\widehat{\Phi}^X : X \to X'$ and $\widehat{\Phi}^Y : Y \to Y'$, such that*

(1) $\lambda(U) \geq (1-\theta)\lambda(\mathcal{B})$;

(2) $d_{\bowtie'}(\Phi_{|U}, \widehat{\Phi}_{|U}) \preceq_{\bowtie} \theta L$.

Let $\mathcal{B}$ be a box at scale L, let $i \in I_g$ and for all $i \in I_g$ let $U_i \subset \mathcal{B}_i$ be as in Theorem 4.5, where U_i is the subset of $\mathcal{B}_i$ on which Φ is close to a product map $\widehat{\Phi}_i$. Let us denote by $W \subset \mathcal{B}$ the "good" set of $\mathcal{B}$,

$$W := \bigsqcup_{i \in I_g} U_i,$$

where "good" means the set on which Φ is close to a product map on boxes at scale R. We introduce the function P which quantifies the portion of a geodesic segment which is not in W.

Definition 4.10 Let $\gamma : [0, L] \to X \bowtie Y$ be a vertical geodesic segment of $\mathcal{B}$. We denote the measure of points in $\gamma \cap W^c$ by

$$P(\gamma) := \mathrm{Leb}(\gamma^{-1}(W^c)). \tag{98}$$

The value of $P(\gamma)$ is related to γ being ε-monotone. The following lemma is mostly inspired from Lemma 5.10 of [8].

Lemma 4.11 *For $0 \leq \varepsilon \preceq_{k,c,\bowtie} \sqrt{\theta} \preceq_{k,c,\bowtie} 1$, there exists $M(\bowtie, k, c)$ such that for all vertical geodesic segments $\gamma : [0, L] \to X \bowtie Y$ we have*

$$P(\gamma) \leq \sqrt{\theta}L \implies \Phi \circ \gamma \text{ is } M\sqrt{\theta}\text{-monotone.}$$

Proof Let $t_1, t_2 \in [0, L]$ such that $h\big(\Phi(\gamma(t_1))\big) = h\big(\Phi(\gamma(t_2))\big)$ and such that $t_2 \geq t_1$. Let us decompose $[t_1, t_2]$ into segments of length $\sqrt{\theta}R$. Without loss of generality we can assume that $t_2 - t_1 \geq \sqrt{\theta}L$. Let us define $N := \big\lfloor (t_2 - t_1)/\sqrt{\theta}R \big\rfloor$, $I_i := \big[t_1 + i\sqrt{\theta}R, t_1 + (i+1)\sqrt{\theta}R\big[$ for any $i \in \{0, \dots, N-1\}$ and $I_N := \big[t_1 + (N-1)\sqrt{\theta}R, t_2\big]$. We have

$$[t_1, t_2] := \bigsqcup_{i=0}^{N} I_i.$$

Then for all $i \in \{0, \dots, N\}$ let us choose $s_i \in I_i$ such that $\gamma(s_i) \in W$ if possible, and any $s_i \in I_i$ otherwise. Let us denote by J the set of odd indexes in $\{0, \dots, N\}$. We split J into the sets

$$\begin{aligned}
J_0 &:= \{j \in J \mid \gamma(s_j) \text{ and } \gamma(s_{j+2}) \text{ are both in the same box and in } W\};\\
J_1 &:= \{j \in J \mid \gamma(s_j) \text{ and } \gamma(s_{j+1}) \text{ are in different boxes}\};\\
J_1' &:= \{j \in J \mid \gamma(s_{j+1}) \text{ and } \gamma(s_{j+2}) \text{ are in different boxes}\};\\
J_2 &:= \{j \in J \mid I_j \subset W^c\};\\
J_2' &:= \{j \in J \mid I_{j+2} \subset W^c\}.
\end{aligned}$$

We claim that

$$J = J_0 \sqcup (J_1 \cup J_1' \cup J_2 \cup J_2').$$

To prove it, one can see that two cases arise when an odd index j is not in J_0. The first case is when $\gamma(s_j)$ and $\gamma(s_j+2)$ are not in the same box, which leads to the fact that either $j \in J_1$ or $j \in J_1'$. The second case happens when $\gamma(s_j)$ or $\gamma(s_j+2)$ are not in W, which leads to either $I_j \subset W^c$ or $I_{j+2} \subset W^c$. Therefore, we proved that an odd index is either in J_0 or in $J_1 \cup J_1' \cup J_2 \cup J_2'$.

We have that $P(\gamma) \leq \sqrt{\theta} L$, hence $\#J_2 \leq \frac{\sqrt{\theta}L}{\sqrt{\theta}R} = \frac{L}{R}$ and similarly $\#J_2' \leq \frac{L}{R}$. Furthermore there are less than $\frac{L}{R}$ boxes intersecting γ, therefore $\#J_1 \leq \frac{t_2-t_1}{R} \leq \frac{L}{R}$ and $\#J_1' \leq \frac{L}{R}$, hence

$$\#(J_1 \cup J_1' \cup J_2 \cup J_2') \leq 4\frac{L}{R}; \quad \#J_0 = \#J - \#(J_1 \cup J_1' \cup J_2 \cup J_2') \geq \frac{t_2-t_1}{2\sqrt{\theta}R} - 4\frac{L}{R}.$$

We see that the "good" indexes are in majority compared to the "bad" indexes. We now use that fact to prove that $|t_2 - t_1|$ is smaller than $\sqrt{\theta}L$. Let us define $q(t) := h \circ \Phi \circ \gamma(t)$ for all $t \in [0, L]$. We assume that N is odd, the case where N is even is treated identically. By assumption $q(t_1) = q(t_2)$ therefore

$$\begin{aligned} 0 &= q(t_2) - q(t_1) = q(t_2) - q(s_N) + \sum_{i \in J} (q(s_{i+2}) - q(s_i)) + q(s_1) - q(t_1) \\ &= q(t_2) - q(s_N) + \sum_{i \in J_0} (q(s_{i+2}) - q(s_i)) + \sum_{i \in J \setminus J_0} (q(s_{i+2}) - q(s_i)) + q(s_1) - q(t_1). \end{aligned} \tag{99}$$

However we proved that $\#J_0$ is much bigger than $\#(J \setminus J_0)$, and for any $i \in J_0$, $q(s_{i+2}) - q(s_i)$ is a positive number by the upward orientation of the quasi-isometry on W. Therefore we will show that $|t_1 - t_2|$ must be small for this equality to hold. First, we have to consider that, $\forall i \in \{0, \dots, N\}$,

$$\begin{aligned} l(I_{i+1}) \leq |s_i - s_{i+2}| \leq l(I_i) + l(I_{i+1}) + l(I_{i+2}) &\implies \sqrt{\theta}R \leq |s_i - s_{i+2}| \leq 3\sqrt{\theta}R \\ &\implies |q(s_i) - q(s_{i+2})| \preceq_{k,c,\bowtie} \sqrt{\theta}R. \end{aligned}$$

Hence for all $i \in J \setminus J_0$ we have $q(s_{i+2}) - q(s_i) \succeq_{k,c,\bowtie} -\sqrt{\theta}R$. Furthermore for all $i \in J_0$, s_i and s_{i+2} are in the same box and in W, therefore by Corollary 2.7, there exists $M(k, c, \bowtie)$ such that

$$q(s_{i+2}) - q(s_i) \geq \frac{1}{k}|s_i - s_{i+2}| - M\varepsilon R \succeq_{k,c,\bowtie} \sqrt{\theta}R \quad \big(\text{since } \sqrt{\theta} \geq 2M\varepsilon\big).$$

Combined with equality (99)

$$0 \succeq_{k,c,\bowtie} \sqrt{\theta}R\#J_0 - \sqrt{\theta}R\#(J_1 \cup J_1' \cup J_2 \cup J_2') \succeq |t_2 - t_1| - \sqrt{\theta}L.$$

Hence $|t_2 - t_1| \preceq_{k,c,\bowtie} \sqrt{\theta}L$, which proves that there exists $M(k, c, \bowtie)$ such that γ is $M\sqrt{\theta}$-monotone. □

Let M be the constant involved in Lemma 4.11, let $\theta' = \theta^{\frac{1}{16}}$ and let $\varepsilon' := 2M\sqrt{\theta'}$. We now show that almost all vertical geodesic segments of boxes at scale L have ε'-monotone images under Φ.

Let us denote by $V^g\mathcal{B} \subset V\mathcal{B}$ the set of vertical geodesic segments of $V\mathcal{B}$ whose image by Φ are ε'-monotone.

Lemma 4.12 *For $L \succeq_{k,c,\bowtie} \frac{1}{\theta}$ and for any box $\mathcal{B}$ at scale L we have that*

$$\eta(V^g\mathcal{B}) \geq (1-\theta^{\frac{1}{32}})\eta(V\mathcal{B}). \tag{100}$$

Proof Lemma 4.11 tells us that $P(\gamma) \geq \sqrt{\theta}L$ for all $\gamma \in V^b\mathcal{B}$. Computing the measure λ of W^c,

$$\lambda(W^c) = \int_0^L \lambda_z(W_z^c)\,\mathrm{d}z \asymp_\bowtie \int_0^L \eta(V_{\mathcal{B}}(W_z^c))\,\mathrm{d}z \quad \text{(by Property 3.23)}$$

$$\asymp_\bowtie \int_0^L \int_{V\mathcal{B}} \mathbb{1}_{V_{\mathcal{B}}(W_z^c)}(\gamma)\,\mathrm{d}\eta(\gamma)\,\mathrm{d}z \asymp_\bowtie \int_{V\mathcal{B}} \int_0^L \mathbb{1}_{V_{\mathcal{B}}(W_z^c)}(\gamma)\,\mathrm{d}z\,\mathrm{d}\eta(\gamma) \quad \text{(by Fubini's theorem).} \tag{101}$$

However we have

$$\mathbb{1}_{V_{\mathcal{B}}(W_z^c)}(\gamma) = \begin{cases} 0 & \text{if } z \in \gamma^{-1}(W), \\ 1 & \text{if } z \in \gamma^{-1}(W^c). \end{cases} \tag{102}$$

Therefore $\mathbb{1}_{V_{\mathcal{B}}(W_z^c)}(\gamma) = \mathbb{1}_{\gamma^{-1}(W^c)}(z)$. With inequality (101) it gives us

$$\lambda(W^c) \asymp_\bowtie \int_{V\mathcal{B}} \int_0^L \mathbb{1}_{\gamma^{-1}(W^c)}(z)\,\mathrm{d}z\,\mathrm{d}\eta(\gamma) \geq \int_{V^b\mathcal{B}} \int_0^L \mathbb{1}_{\gamma^{-1}(W^c)}(z)\,\mathrm{d}z\,\mathrm{d}\eta(\gamma) \quad \text{(since } V^b\mathcal{B} \subset V\mathcal{B})$$

$$\geq \int_{V^b\mathcal{B}} \mathrm{Leb}(\gamma^{-1}(W^c))\,\mathrm{d}\eta(\gamma) = \int_{V^b\mathcal{B}} P(\gamma)\,\mathrm{d}\eta(\gamma). \tag{103}$$

Let us assume by contradiction that $\eta(V^g\mathcal{B}) < \left(1-\sqrt{\theta'}\right)\eta(V\mathcal{B})$, hence we have $\eta(V^b\mathcal{B}) > \sqrt{\theta'}\eta(V\mathcal{B})$. Therefore by inequality (103)

$$\begin{aligned} \lambda(W^c) &\succeq_\bowtie \eta(V^b\mathcal{B})\sqrt{\theta'}L \geq \sqrt{\theta'}\eta(V\mathcal{B})\sqrt{\theta'}L \\ &\asymp_\bowtie \theta'\lambda(\mathcal{B}), \end{aligned}$$

which contradicts the first conclusion of Theorem 4.9 for $\theta \preceq_{k,c,\bowtie} 1$. □

As in Section 4.2, we deduce that, in boxes which have almost only vertical geodesic segment with $2M\sqrt{\theta'}$-monotone image, Φ is close to a product map. Let us define $\varepsilon' := 2M\theta^{\frac{1}{16}}$ and $\theta' := 2M\theta^{\frac{1}{16}}$. Then for $0 < \theta' \preceq_{k,c,\bowtie} 1$ we have that $\theta' \leq \varepsilon' \leq \sqrt{\theta'}$.

Proof of Theorem 4.9 The proof is similar to Theorem 4.5. Lemma 4.12 plays the role of the second conclusion of Proposition 4.3, with ε' instead of ε. In a box at scale L, almost all vertical geodesic segment have ε'-monotone image by Φ.

Then, because $\varepsilon' \preceq_{k,c,\bowtie} \sqrt{\theta'}$, Lemma 4.4 provides us with a dominant orientation. In combination with Lemma 3.26, we get Lemma 4.6, which provides us with the vertical quadrilateral.

Afterwards, we make use of them, as in the proof of Theorem 4.5, to construct the quasi-isometry product $\widehat{\Phi}$. In a box at scale R, the upper-bound εR on the distance between Φ and $\widehat{\Phi}$ is achieved since $\theta' \leq \varepsilon$, and in our box at scale L, it is achieved since $\theta' \leq \varepsilon'$.

Finally, the exponents on θ of Theorem 4.9 can be removed since we can fix θ, then do the proof with a parameter $\tilde{\theta} = \theta^8$, then replace $\tilde{\theta}$ by θ^8. □

This is a step forward since now, Theorem 4.9 holds for all boxes at scale L, and not only a significant proportion of boxes at scale R.

4.5 A quasi-isometry quasi-respects the height

Let $p, q \in X \bowtie Y$ be such that $h(p) = h(q)$. In this section we aim to prove the following theorem, which estimates the difference of height between the images of p and q under Φ.

Theorem 4.13 *For $0 < \theta \preceq_{k,c,\bowtie} 1$, there exists $M(k, c, \bowtie, \theta)$ (here M depends also on θ) such that for all p and q in $X \bowtie Y$ with $h(p) = h(q)$ we have*

$$\Delta h(\Phi(p), \Phi(q)) \leq \theta d_{\bowtie}(p, q) + M. \tag{104}$$

By the previous section, we know that in a box of a sufficiently large scale, the quasi-isometry Φ is (on a set of almost full measure) close to a product map. We first show that this product map is coarsely an homothety along the height function.

Let L_0 be the constant of Theorem 4.9, let $L \geq L_0$ and let $\mathcal{B}$ be a box at scale L. Let us define $h^+ := \sup\{h(p) \mid p \in \mathcal{B}\}$ and $h^- := \inf\{h(p) \mid p \in \mathcal{B}\}$. Let $\widehat{\Phi} := (\widehat{\Phi}^X, \widehat{\Phi}^Y) : X \bowtie Y \to X' \bowtie Y'$ be the corresponding product map of Theorem 4.9.

Lemma 4.14 *Let $a \in \mathcal{B}_{h^+}$ and $b \in \mathcal{B}_{h^-}$ be two points of $\mathcal{B}$, one on its top part and one on its bottom part. Then we have both*

$$\left|\Delta h(\widehat{\Phi}(a), \widehat{\Phi}(b)) - \frac{m}{m'}L\right| \preceq_{k,c,\bowtie} \theta^{\frac{1}{2}} L;$$

$$\left|\Delta h(\widehat{\Phi}(a), \widehat{\Phi}(b)) - \frac{n}{n'}L\right| \preceq_{k,c,\bowtie} \theta^{\frac{1}{2}} L.$$

Proof Let $U \subset \mathcal{B}$ be the set involved in Theorem 4.9. We recall that $\lambda(U) \geq (1-\theta)\lambda(\mathcal{B})$ and that for all $p \in U$, we have $d_{\bowtie'}(\Phi(p), \widehat{\Phi}(p)) \preceq_{\bowtie,k,c} \theta L$. Since the measure λ identically weights the level sets of $\mathcal{B}$, by a Markov inequality there exists $z^+ \in [h^+ - \theta^{\frac{1}{2}} L, h^+]$ and $z^- \in [h^-, h^- + \theta^{\frac{1}{2}} L]$ such that

$$\lambda_{z^+}(U_{z^+}) \geq (1 - \theta^{\frac{1}{2}})\lambda_{z^+}(\mathcal{B}_{z^+});$$

$$\lambda_{z^-}(U_{z^-}) \geq (1 - \theta^{\frac{1}{2}})\lambda_{z^-}(\mathcal{B}_{z^-}).$$

By the definition of λ_{z^+} we have that

$$\tfrac{1}{2}\mu_{z^+}(\mathcal{B}_{z^+}) \leq \mu_{z^+}(U_{z^+}) \leq \mu_{z^+}(\mathcal{B}_{z^+}).$$

Furthermore $\mu_{z^+}(\mathcal{B}_{z^+}) \asymp_{\bowtie} e^{nL} e^{(m-n)|h^+ - z^+|}$ since we went down by a height $|h^+ - z^+|$ in the box. Therefore

$$e^{nL} e^{(m-n)\theta^{\frac{1}{2}} L} \preceq_{k,c,\bowtie} \mu_{z^+}(\mathcal{N}_{kc+c}(U_{z^+})).$$

Furthermore, B_{z^+} resembles a rectangle of width $2|h^+ - z^+|$ in X and $2(L - |h^+ - z^+|)$ in Y, hence

$$\mu^Y_{z^+}(\mathcal{N}_{kc+c}(U^Y_{z^+})) \succeq_{k,c,\bowtie} e^{nL} e^{(m-n)\theta^{\frac{1}{2}} L} \frac{1}{\mu^X_{z^+}(\mathcal{B}^X_{z^+})} \succeq_{k,c,\bowtie} e^{nL} e^{2(m-n)\theta^{\frac{1}{2}} L}.$$

By Lemma 3.18, and since $\widehat{\Phi}$ is close to Φ on U, we deduce

$$e^{nL}e^{2(m-n)\theta^{\frac{1}{2}}L} \preceq_{k,c,\bowtie} \mu^Y_{z+}\big(\mathcal{N}_1(\widehat{\Phi}(U^Y_{z+}))\big). \tag{105}$$

Let $\Delta \geq 0$ be $|h(\widehat{\Phi}(U_{z+}))-h(\widehat{\Phi}(U_{z-}))|$. For all $p \in U_{z+}$ there exists a vertical geodesic V_p of θ-monotone image under Φ passing close to p. Furthermore, $d_Y(V^Y_p(z^-), U^Y_{z-}) \leq 2\theta^{\frac{1}{2}}L$ since $\mathcal{B}^Y_{z-}$ has a relatively small diameter. Therefore, all vertical geodesics starting at $\mathcal{N}_1(\widehat{\Phi}^Y(U^Y_{z+}))$ intersect $\mathcal{N}_{M\theta^{\frac{1}{2}}L}(\widehat{\Phi}^Y(U^Y_{z-}))$. Hence we have that

$$\mathcal{N}_1(\widehat{\Phi}^Y(U^Y_{z+})) \subset \pi_{h(\widehat{\Phi}(U_{z+}))}\big(\mathcal{N}_{M\theta^{\frac{1}{2}}L}(\widehat{\Phi}^Y(U^Y_{z-}))\big).$$

Therefore

$$\begin{aligned}
\mu^{Y'}\big(\mathcal{N}_1(\widehat{\Phi}^Y(U^Y_{z+}))\big) &\leq \mu^{Y'}\big(\pi_{h(\widehat{\Phi}(U_{z+}))}\big(\mathcal{N}_{M\theta^{\frac{1}{2}}L}(\widehat{\Phi}^Y(U^Y_{z-}))\big)\big)\\
&\preceq_{k,c,\bowtie} e^{n'\Delta}\mu^{Y'}\big(\mathcal{N}_{M\theta^{\frac{1}{2}}L}(\widehat{\Phi}^Y(U^Y_{z-}))\big)\\
&\preceq_{k,c,\bowtie} e^{n'\Delta}e^{-n'M\theta^{\frac{1}{2}}L}\mu^{Y'}\big(\mathcal{N}_1(\widehat{\Phi}^Y(U^Y_{z-}))\big) \quad \text{(by Corollary 3.16)}\\
&\preceq_{k,c,\bowtie} e^{n'\Delta}e^{n'M\theta^{\frac{1}{2}}L}e^{\theta^{\frac{1}{2}}L} \quad \text{(because } \mathcal{B}^Y_{z-}\text{has small } \mu \text{ measure)}\\
&= e^{n'\Delta}e^{n'(M+1)\theta^{\frac{1}{2}}L}.
\end{aligned}$$

Combined with inequality (105) we obtain

$$e^{nL}e^{2(m-n)\theta^{\frac{1}{2}}L} \preceq_{k,c,\bowtie} e^{n'\Delta}e^{n'(M+1)\theta^{\frac{1}{2}}L},$$

which provides us with $e^{nL} \preceq_{k,c,\bowtie} e^{n'\Delta}e^{M'\theta^{\frac{1}{2}}L}$, where M' is a constant depending on k, c, $\bowtie$ and $\bowtie'$. Then there exists $M''(k,c,\bowtie,\bowtie')$ such that by taking the logarithm we get

$$nL \leq n'\Delta + M''\theta^{\frac{1}{2}}L.$$

Similarly, we do the same proof on Φ^{-1}, on the box of height Δ containing $\widehat{\Phi}(U_{z+} \cup U_{z-})$ which provides us with

$$n'\Delta \leq nL + M''\theta^{\frac{1}{2}}L.$$

Therefore $|\Delta - \frac{n}{n'}L| \preceq_{k,c,\bowtie} \theta^{\frac{1}{2}}L$. To obtain the same results with the constants m and m', we focus on the sets U^X_{z+} and U^X_{z-} instead of U^Y_{z+} and U^Y_{z-}. □

As a corollary we obtain a first quasi-isometry invariant for horospherical products.

Proposition 4.15 *If $X \bowtie Y$ and $X' \bowtie Y'$ are quasi-isometric, then $\frac{m}{n} = \frac{m'}{n'}$.*

Proof By Lemma 4.14, and by the triangle inequality we have that $\left|\frac{m}{m'} - \frac{n}{n'}\right| \preceq_{k,c,\bowtie} \frac{1}{L}$ for all $L \geq L_0$. Therefore, $\frac{m}{m'} = \frac{n}{n'}$, hence $\frac{m}{n} = \frac{m'}{n'}$. □

Lemma 4.16 *Let $0 < \theta \preceq_{k,c,\bowtie} 1$. Let $p := (p^X, p^Y), q := (q^X; q^Y) \in X \bowtie Y$ such that $d_\bowtie(p,q) \geq L_0^2$ and such that $p^Y = q^Y$ (hence $h(p) = h(q)$). Then we have*

$$\Delta h(\Phi(p), \Phi(q)) \preceq_{k,c,\bowtie} \theta^{\frac{1}{2}} d_\bowtie(p,q).$$

Proof Let $\mathcal{B}$ be a box of scale $L = d_\bowtie(p,q)$, such that p and q are contained in its bottom part. Let $V_p^X \in V\mathcal{B}^X$ be the vertical geodesic segment of X of length L starting at p. We apply Proposition 4.3 on $V_p^X \bowtie \mathcal{B}^Y$ (as a box of an embedded copy of $\mathbb{R} \bowtie Y$ inside $X \bowtie Y$) with $r_0 = L_0$ and $L \geq L_0^2$. We obtain that there exists $R \geq L_0$, a box tiling $\mathcal{B} \cup \bigcup_{i\in I} \mathcal{B}_i$ of boxes at scale R and $I_g \subset I$ such that

(1) $\lambda\big(\bigcup_{i\in I_g} \mathcal{B}_i\big) \geq (1-\theta)\lambda(\mathcal{B})$ (boxes indexed by I_g cover almost all $\mathcal{B}$);

(2) $\forall i \in I_g$, $\frac{\eta_i(V^b\mathcal{B}_i)}{\eta_i(V\mathcal{B}_i)} \leq \theta$ (almost all vertical geodesic segments in $\mathcal{B}_i$ have ε-monotone image),

where $\eta_i := \eta_{V\mathcal{B}_i}$. In this setting, we have that $V\mathcal{B}_i := \{(V_p^X, V^Y) \mid V^Y \in V\mathcal{B}_i^Y\}$, hence most vertical geodesics in $\mathcal{B}_i^Y$ are a good vertical geodesic of $\mathcal{B}_i$ when coupled with a portion of V_p^X.

Let us define $J := \{0, \dots, \frac{L}{R} - 1\}$ and for all $j \in J$ let us define $p_j^X := V_p^X(jR)$. Then we have

$$V_p^X(jR) := \bigcup_{j\in J} [p_j^X; p_{j+1}^X].$$

Since the measure of the good boxes cover almost all $\mathcal{B}$, and because the measure λ equally weights the level sets, by a Markov inequality argument there exists $J_g \subset J$ such that for all $j \in J_g$, $\mathcal{B}_{[jR;(j+1)R]}$ is almost entirely covered by boxes of I_g. Therefore, again by Markov inequality argument, there exists $W_p \subset V^Y\mathcal{B}$ such that

(1) $\eta(W_p) \geq (1-\theta^{\frac{1}{2}})\eta(V^Y\mathcal{B})$;

(2) $\forall V^Y \in W_p$ and $\forall j \in J_g$ we have $V^Y([-(j+1)R; -jR]) \in \bigcup_{i\in I_g} V^g\mathcal{B}_i^Y$.

Let $V^Y \in W_p$, for all $j \in J$ let us define $p_i := (V_p^X(jR), V^Y(-jR))$. By Lemma 4.14, for all $j \in J$,

$$\left|\Delta h(\widehat{\Phi}(p_j), \widehat{\Phi}(p_{j+1})) - \frac{m}{m'}R\right| \preceq_{k,c,\bowtie} \theta^{\frac{1}{2}} R. \tag{106}$$

For all $j \in J_g$, let us denote by $\mathcal{B}_j$ the box at scale R containing $[p_j; p_{j+1}]$. By the choice of R, we have that most vertical geodesic segments of $\mathcal{B}_j^Y$ have θ-monotone image when coupled with $[p_j^X; p_{j+1}^X]$.

Furthermore $\mathcal{B}_j$ contains almost only good vertical geodesic segments, therefore, there exists $v \in V^g\mathcal{B}_j$ such that $([p_j^X; p_{j+1}^X], v^Y) \in V^g\mathcal{B}_j$ and such that $(v^X, [p_j^Y; p_{j+1}^Y]) \in V^g\mathcal{B}_j$. Therefore there exists a good coarse vertical quadrilateral containing p_j and p_{j+1}, hence $d(\Phi(p_j), \widehat{\Phi}(p_j)) \preceq_{k,c,\bowtie} \theta R$. Similarly we have $d(\Phi(p_{j+1}), \widehat{\Phi}(p_{j+1})) \preceq_{k,c,\bowtie} \theta R$. Hence combined with inequality (106) we get

$$\left|\Delta h(\Phi(p_j), \Phi(p_{j+1})) - \frac{m}{m'}R\right| \preceq_{k,c,\bowtie} \theta^{\frac{1}{2}} R.$$

Therefore by the triangle inequality, there exists $M(k,c,\bowtie)$ and $M'(k,c,\bowtie)$ such that

$$\begin{aligned}\Delta h(\Phi(p_0),\Phi(p_{\frac{L}{R}-1})) &\le \sum_{j=0}^{\frac{L}{R}-1}\Delta h(\Phi(p_j),\Phi(p_{j+1}))\\ &\le \sum_{j\in J_g}\Delta h(\Phi(p_j),\Phi(p_{j+1}))+\sum_{j\in J\setminus J_g}\Delta h(\Phi(p_j),\Phi(p_{j+1}))\\ &\le \#J_g\left(\frac{m}{m'}R+M\theta^{\frac12}R\right)+\#(J\setminus J_g)(kR+c)\\ &\le \frac{L}{R}\left(\frac{m}{m'}R+M\theta^{\frac12}R\right)+\theta\frac{L}{R}(kR+c)\\ &\le \frac{m}{m'}L+M'\theta^{\frac12}L.\end{aligned}$$

Similarly we have $\Delta h(\Phi(p_0),\Phi(p_{\frac{L}{R}-1}))\ge\frac{m}{m'}L-M'\theta^{\frac12}L$. By doing the same reasoning on q we have that for all $V^Y\in W_q$, $\left|\Delta h(\Phi(q_0),\Phi(q_{\frac{L}{R}-1}))-\frac{m}{m'}L\right|\preceq_{k,c,\bowtie}\theta^{\frac12}L$, where $q_j:=(V_q^X(jR),V^Y(-jR))$. Furthermore $W_p\cap W_q$ is nonempty for $\theta^{\frac12}\preceq_{k,c,\bowtie}1$, then let $V^Y\in W_p\cap W_q$. Without loss of generality we can assume that $\Phi(p)\ge\Phi(q)$. We have

$$\begin{aligned}\Delta h(\Phi(p),\Phi(q)) &= h(\Phi(p))-h(\Phi(p_0))+h(\Phi(p_0))-h(\Phi(p_{\frac{L}{R}-1}))+h(\Phi(p_{\frac{L}{R}-1}))\\ &\qquad -h(\Phi(q_{\frac{L}{R}-1}))+h(\Phi(q_{\frac{L}{R}-1}))-h(\Phi(q_0))+h(\Phi(q_0))-h(\Phi(q))\\ &\preceq_{k,c,\bowtie} d_\bowtie(p,p_0)-\frac{m}{m'}L+M\theta^{\frac12}L+d_\bowtie(p_{\frac{L}{R}-1},q_{\frac{L}{R}-1})+\frac{m}{m'}L+M\theta^{\frac12}L+d_\bowtie(q,q_0)\\ &\preceq_{k,c,\bowtie} d_\bowtie(p,p_0)+d_\bowtie(p_{\frac{L}{R}-1},q_{\frac{L}{R}-1})+d_\bowtie(q,q_0)+2\theta^{\frac12}L.\end{aligned}$$

However $d_\bowtie(p,p_0)\le M_0$ since they share the same X coordinate and because the top part of $\mathcal{B}^Y$ as a diameter of at most M_0, similarly $d_\bowtie(q,q_0)\le M_0$. By construction $p^Y_{\frac{L}{R}-1}=q^Y_{\frac{L}{R}-1}$, furthermore the top part of $\mathcal{B}^X$ has a diameter of at most M_0, hence

$$d_\bowtie(p_{\frac{L}{R}-1},q_{\frac{L}{R}-1})\le M_0.$$

Finally we obtain

$$\Delta h(\Phi(p),\Phi(q))\preceq_{k,c,\bowtie}\theta^{\frac12}L=\theta^{\frac12}d_\bowtie(p,q).$$ □

Corollary 4.17 *Any vertical geodesic ray V of $X\bowtie Y$ satisfies, for all $t_1,t_2\in\mathbb{R}$,*

$$h(\Phi\circ V(t_1))=h(\Phi\circ V(t_2))\implies|t_1-t_2|\preceq_{k,c,\bowtie}1.$$

Proof Suppose V is a vertical geodesic segment parametrised by arclength. Suppose $0<t_1<t_2$ are such that $h\big(\Phi(V(t_1))\big)=h\big(\Phi(V(t_2))\big)$. We apply Theorem 4.13 on Φ^{-1} with $p=\Phi(V(t_1))$, $q=\Phi(V(t_2))$, where θ is here fixed and depends only on k,c and the metric measured space $(X\bowtie Y,d_\bowtie)$. Hence there exists $M(k,c,\bowtie)>0$ such that

$$\Delta h(V(t_1),V(t_2))\preceq_{k,c,\bowtie}\theta^{\frac12}|t_1-t_2|+M. \tag{107}$$

However $\Delta h(V(t_1), V(t_2)) = |t_1 - t_2|$, hence

$$(1 - \theta^{\frac{1}{2}})|t_1 - t_2| \preceq_{k,c,\bowtie} 1.$$

Therefore $|t_1 - t_2| \preceq_{k,c,\bowtie} 1$ since $\theta^{\frac{1}{2}} \leq \frac{1}{2}$. □

This is stronger than being ε-monotone since it holds for all $t_1, t_2 \in \mathbb{R}$.

4.6 Factorisation of a quasi-isometry on the whole space

Finally, we provide the proof of Theorem 4.1, which states that Φ is close to a product map $\widehat{\Phi}$ on the whole space $X \bowtie Y$.

Proof of Theorem 4.1 We first pick an arbitrary vertical geodesic V_0^X of X and an arbitrary vertical geodesic V_0^Y of Y. Then we work with the two embedded copies $X_0 := X \bowtie V_0^Y$ and $Y_0 := V_0^X \bowtie Y$ of X and Y in $X \bowtie Y$. Let $p \in X \bowtie Y$, there exist a unique $a \in X_0$ and a unique $b \in Y_0$ such that $p^X = a^X$ and $p^Y = b^Y$. We can construct a coarse vertical quadrilateral Q containing p and a as in Lemma 4.6. Thanks to Corollary 4.17, we know that $\Phi(Q)$ is in the $M(k, c, \bowtie)$-neighbourhood of a coarse vertical quadrilateral Q' on which we use Proposition 2.11. This gives us

$$d_{X'}(\Phi(p)^{X'}, \Phi(a)^{X'}) \preceq_{k,c,\bowtie} 1; \tag{108}$$

$$\Delta h(\Phi(p)^{X'}, \Phi(a)^{X'}) \preceq_{k,c,\bowtie} 1. \tag{109}$$

Similarly we have $d_{Y'}(\Phi(p)^{Y'}, \Phi(b)^{Y'}) \preceq_{k,c,\bowtie} 1$. Let us define

$$\widehat{\Phi}^X : X \to X', \quad x \mapsto \Phi\big(x, V_0^Y(-h(x))\big)^{X'}.$$

By rewriting inequality (108) we have

$$\begin{aligned} d_{X'}(\Phi(p)^{X'}, \widehat{\Phi}^X(p^X)) &= d_{X'}(\Phi(p)^{X'}, \widehat{\Phi}^X(a^X)) = d_{X'}\big(\Phi(p)^{X'}, \Phi\big(a^X, V_0^Y(-h(a^X))\big)^{X'}\big) \\ &= d_{X'}(\Phi(p)^{X'}, \Phi(a)^{X'}) \preceq_{k,c,\bowtie} 1. \end{aligned}$$

Similarly by defining $\widehat{\Phi}^Y := \Phi\big(V_0^X(-h(y)), y\big)^{Y'}$ for all $y \in Y$, we have

$$d_Y(\Phi(p)^Y, \widehat{\Phi}^Y(p^Y)) \preceq_{k,c,\bowtie} 1. \tag{110}$$

The last problem is that given a point p, the heights of $\widehat{\Phi}^X(p^X)$ and $\widehat{\Phi}^Y(p^Y)$ may differ. As in the proof of Theorem 4.5, inequality (109) guarantees that they are sufficiently close, which allows us to chose $\widehat{\Phi}^X$ and $\widehat{\Phi}^Y$ such that $\widehat{\Phi} := (\widehat{\Phi}^X, \widehat{\Phi}^Y)$ is a well-defined product map on $X \bowtie Y$. Then we have

$$\begin{aligned} d_{\bowtie'}(\Phi(p), \widehat{\Phi}(p)) &\preceq_{k,c,\bowtie} 1; \\ \Delta h(\Phi(p), \widehat{\Phi}(p)) &\preceq_{k,c,\bowtie} 1. \end{aligned}$$

We now prove that $\widehat{\Phi}^X$ and $\widehat{\Phi}^X$ are quasi-isometries. Let $x, x' \in X$, then

$$\begin{aligned}
&d_{X'}(\widehat{\Phi}^X(x), \widehat{\Phi}^X(x')) \\
&\qquad \preceq_{k,c,\bowtie} d_{X'}\big(\Phi\big(x, V_0^Y(-h(x))\big)^{X'}, \Phi\big(x', V_0^Y(-h(x'))\big)^{X'}\big) \\
&\qquad \le d_{\bowtie'}\big(\Phi\big(x, V_0^Y(-h(x))\big), \Phi\big(x', V_0^Y(-h(x'))\big)\big) \\
&\qquad \le k\, d_{\bowtie}\big(\big(x, V_0^Y(-h(x))\big), \big(x', V_0^Y(-h(x'))\big)\big) + c \\
&\qquad \le k\, d_X(x, x') + k\, d_Y\big(V_0^Y(-h(x)), V_0^Y(-h(x'))\big) + c + M(k, c, \bowtie) \quad \text{(by Theorem 1.11)} \\
&\qquad \le k\, d_X(x, x') + \Delta h(x, x') + c + M \le (k+1) d_X(x, x') + c + M.
\end{aligned}$$

Similarly, and because $d_{\bowtie'} \ge \frac{d_{X'} + d_{Y'}}{2}$,

$$\begin{aligned}
&d_{X'}(\widehat{\Phi}^X(x), \widehat{\Phi}^X(x')) \\
&\quad = d_{X'}\big(\Phi\big(x, V_0^Y(-h(x))\big)^{X'}, \Phi\big(x', V_0^Y(-h(x'))\big)^{X'}\big) \\
&\quad \ge 2 d_{\bowtie'}\big(\Phi\big(x, V_0^Y(-h(x))\big), \Phi\big(x', V_0^Y(-h(x'))\big)\big) - d_{Y'}\big(\Phi\big(x, V_0^Y(-h(x))\big)^{Y'}, \Phi\big(x', V_0^Y(-h(x'))\big)^{Y'}\big) \\
&\quad \ge \frac{1}{k} d_X(x, x') - c - d_{Y'}\big(\widehat{\Phi}^Y\big(V_0^Y(-h(x))\big), \widehat{\Phi}^Y\big(V_0^Y(-h(x))\big)\big) - 2M \quad \text{(by the triangle inequality)} \\
&\quad \ge \frac{1}{k} d_X(x, x') - c - 2M.
\end{aligned}$$

The proof that $\widehat{\Phi}^Y$ is a quasi-isometry is similar. □

5 Some solvable Lie groups as horospherical products

In this chapter, we provide a characterisation of the quasi-isometry group of the horospherical product of two Heintze groups. See Theorem 5.13 for the precise description.

5.1 Admissibility of Heintze groups

In this section we show that a Heintze group satisfies the conditions required to apply our main rigidity result Theorem 4.1.

Definition 5.1 (Heintze group) A Heintze group is a solvable Lie group $S = N \rtimes_A \mathbb{R}$ where N is a connected, simply connected, nilpotent Lie group, and A is a derivation of $\mathrm{Lie}(N)$ whose eigenvalues all have positive real parts.

Heintze [16] obtained that any negatively curved homogeneous manifold is isometric to a Heintze group.

Remark 5.2 A Heintze group admits a left-invariant metric with strictly negative sectional curvature; see [16] for further details. From now on we fix g a left-invariant metric on $N \rtimes_A \mathbb{R}$ with maximal sectional curvature -1. Since $N \rtimes_A \mathbb{R}$ is simply connected, it is a CAT(-1)-space.

From now on we fix the metric g such that $S = N \rtimes_A \mathbb{R}$ is a CAT(-1) space. Therefore S is a δ-hyperbolic, Busemann, proper, geodesically complete metric space. Moreover, we show that S satisfies all three assumptions of Definition 3.1. The assumption (E1) holds thanks to the decomposition $S = N \rtimes_A \mathbb{R}$. We have for all $(n, z) \in N \rtimes_A \mathbb{R}$, $g_{(n,z)} = \exp(-zA)(g_N)_n \exp(-zA)^t \oplus \mathrm{d}z^2$, where g_N is the restriction of g to the Lie algebra of N. Let us denote by $g_z := \exp(-zA) g_N \exp(-zA)^t$ a left invariant metric on N, then let us denote by $\mu := \mu_g$ the measure on S induced by g and by $\mu_z := \mu_{g_z}$ the measure on N induced by g_z. Then for all measurable subset $U \subset S$ we have

$$\begin{aligned}\mu(U) &:= \int_S \mathbb{1}_U(n, z)\, \mathrm{d}\mu_g(n, z) = \int_{\mathbb{R}} \int_N \mathbb{1}_U(n, z)\, \mathrm{d}\mu_{g_z}(n)\, \mathrm{d}z \\ &= \int_{\mathbb{R}} \mu_z(U_z)\, \mathrm{d}z,\end{aligned}$$

where $U_z := \{n \in N \mid (n, z) \in U\}$. Assumption (E2) holds with constant $M_0 = 1$ since $g_{n,z}$ is left-invariant, and assumption (E3) arises from the fact that $\det(g_z) = \exp(-2z \cdot \mathrm{tr}(A)) \det(g)$. Therefore, any Heintze group is an admissible horopointed space. Let us define $S_1 := N_1 \rtimes_{A_1} \mathbb{R}$ and $S_2 := N_2 \rtimes_{A_2} \mathbb{R}$, then

$$S_1 \bowtie S_2 = (N_1 \times N_2) \rtimes_A \mathbb{R},$$

with A the matrix $\mathrm{diag}(A_1, -A_2)$. Similarly let us denote by $S_1' := N_1' \rtimes_{A_1'} \mathbb{R}$ and $S_2' := N_2' \rtimes_{A_2'} \mathbb{R}$ two Heintze groups, with N_1', N_2' being two simply connected nilpotent Lie groups and A_1', A_2' being two derivations.

5.2 Precision on the components of the product map

We first refine Theorem 4.1 for Heintze groups.

Remark 5.3 For any vertical geodesics V of $(N_1 \times N_2) \rtimes_A \mathbb{R}$ there exist $n_1 \in N_1$, $n_2 \in N_2$ and an arclength parametrisation of V such that $V(t) = (n_1, n_2, t)$.

Let $\Phi : (N_1 \times N_2) \rtimes_A \mathbb{R} \to (N_1' \times N_2') \rtimes_{A'} \mathbb{R}$ be a (k, c)-quasi-isometry. Let us assume that $\mathrm{tr}(A_1) > \mathrm{tr}(A_2)$ and that $\mathrm{tr}(A_1') > \mathrm{tr}(A_2')$. By Theorem 4.1 there exist $\widehat{\Phi}_1 : S_1 \to S_1'$ and $\widehat{\Phi}_2 : S_2 \to S_2'$ such that

$$d_{\bowtie}(\Phi, (\widehat{\Phi}_1, \widehat{\Phi}_2)) \preceq_{k,c,\bowtie} 1.$$

Lemma 5.4 *Let $i \in \{1, 2\}$. Then for any vertical geodesic $V \in S_i$, there exists a vertical geodesic $V' \in S_j'$ such that*

$$d_{\mathrm{Hff}}(\widehat{\Phi}_i(V), V') \preceq_{k,c,\bowtie} 1.$$

This lemma also holds for any horospherical product where our main result, the geometric rigidity, applies.

Proof Since $S_i = N_i \rtimes_{A_i} \mathbb{R}$ is a Gromov hyperbolic space, there exists $M(k, c, \bowtie)$ such that the image of a vertical geodesic by $\widehat{\Phi}_i$ is in a M-neighbourhood of a geodesic γ of S_i'. By Corollary 4.17 γ is a vertical geodesic, hence for $V' := \gamma$ we have $d_{\mathrm{Hff}}(\widehat{\Phi}_i(V), V') \preceq_{k,c,\bowtie} 1$. □

Let $n \in N_i$ and let us denote by V_n the vertical geodesic $V_n : \mathbb{R} \to S_i; t \mapsto (n,t)$. By Lemma 5.4 there exists a vertical geodesic V_n' such that

$$d_{\mathrm{Hff}}(\widehat{\Phi}_i(V_n), V_n') \preceq_{k,c,\bowtie} 1. \tag{111}$$

Furthermore V_n' is unique since it is an infinite geodesic of the Heintze group S_i. We define a map $\Psi_i : N_i \to N_i'$ as

$$\Psi_i(n) = P(V_n'(0)) \quad \text{for all } n \in N_i, \tag{112}$$

where $P : N_i' \rtimes_{A_i} \mathbb{R} \to N_i'$ is the natural projection on N_i.

The goal of this subsection is to prove the following theorem.

Theorem 5.5 *There exists $t_0 \in \mathbb{R}$ such that for the aforementioned Ψ_i we have*

$$d_{\bowtie}\left(\Phi, \left(\Psi_1, \Psi_2, \frac{\mathrm{tr}(A_1)}{\mathrm{tr}(A_1')}\mathrm{id}_{\mathbb{R}} + t_0\right)\right) \preceq_{k,c,\bowtie} 1.$$

We can replace $\frac{\mathrm{tr}(A_1)}{\mathrm{tr}(A_1')}$ by $\frac{\mathrm{tr}(A_2)}{\mathrm{tr}(A_2')}$ thanks to Proposition 4.15. We first show $\widehat{\Phi}_i$ and Ψ_i are related.

Lemma 5.6 *Let $i \in \{1,2\}$. There exists $f_i : \mathbb{R} \to \mathbb{R}$ such that, for all $(n,t) \in S_i$,*

$$d_{S_i}\big(\widehat{\Phi}_i(n,t), (\Psi_i(n), f_i(t))\big) \preceq_{k,c,\bowtie} 1.$$

Proof Let $f_i : \mathbb{R} \to \mathbb{R}; t \mapsto h(\widehat{\Phi}_i(e_{N_i}, t))$. Then by Theorem 4.1 we have that $h(\widehat{\Phi}_i(n,t)) = f_i(t)$ for all $n \in N_i$. Therefore by the definition of Ψ_i we have $(\Psi_i(n), f_i(t)) = V_n'(f_i(t))$. Hence

$$d_{S_i'}\big(\widehat{\Phi}_i(n,t), (\Psi_i(n), f_i(t))\big) = d_{S_i'}\big(\widehat{\Phi}_i(n,t), V_n'(f_i(t))\big). \tag{113}$$

However by inequality (111), there exists $s_t \in \mathbb{R}$ such that

$$d_{S_i'}(\widehat{\Phi}_i(n,t), V_n'(s_t)) \preceq_{k,c,\delta} 1. \tag{114}$$

Furthermore we know that

$$1 \succeq_{k,c,\bowtie} d_{S_i'}(\widehat{\Phi}_i(n,t), V_n'(s_t)) \geq \Delta h(\widehat{\Phi}_i(n,t), V_n'(s_t)) = |f_i(t) - s_t|. \tag{115}$$

Therefore

$$\begin{aligned} &d_{S_i'}\big(\widehat{\Phi}_i(n,t), V_n'(f_i(t))\big) \\ &\qquad \leq d_{S_i'}(\widehat{\Phi}_i(n,t), V_n'(s_t)) + d_{S_i'}\big(V_n'(s_t), V_n'(f_i(t))\big) \quad \text{(by the triangle inequality)} \\ &\qquad = d_{S_i'}(\widehat{\Phi}_i(n,t), V_n'(s_t)) + |f_i(t) - s_t| \preceq_{k,c,\bowtie} 1 \quad \text{(by inequalities (114) and (115)).} \end{aligned}$$

Combined with equality (113) it provides us with $d_{S_i'}\big(\widehat{\Phi}_i(n,t), (\Psi_i(n), f_i(t))\big) \preceq_{k,c,\bowtie} 1$. □

Corollary 5.7 (quasi-isometries quasi-preserve the horosphere volume) *Let $t \in \mathbb{R}$, $r > 0$ and $n \in N_i$. Then the map $\widetilde{\Phi}_i := (\Psi_i, f_i)$ quasi-preserves the volume of any disk $D := D_r(n,t)$:*

$$\mu_t^{S_i}(D) \asymp_{k,c,\bowtie} \mu_{f_i(t)}^{S_i'}\big(\mathcal{N}_1(\widetilde{\Phi}_i(D))\big).$$

Proof By Lemma 5.6, there exists $M(k,c,\bowtie)$ such that $\widetilde{\Phi}_i$ is M-close to $\widehat{\Phi}_i$. Therefore, there exist k', c' depending only on k, c and $S_1 \bowtie S_2$ such that $\widetilde{\Phi}_i$ is a (k',c')-quasi-isometry.

We first pick a $2k'(c'+1)$-maximal separating set Z of D. Then $\widetilde{\Phi}_i(Z)$ is such that

(1) the disks $D_1(p)$ with $p \in \widetilde{\Phi}_i(Z)$ are pairwise disjoint;

(2) $\bigcup_{p\in\widetilde{\Phi}_i(Z)} D_1(p) \subset \mathcal{N}_1(\widetilde{\Phi}_i(D)) \subset \bigcup_{p\in\widetilde{\Phi}_i(Z)} D_{2k'\cdot k'(c'+1)+c'+1}(p)$.

Furthermore by Lemma 3.2, we have, $\forall (n,t) \in Z$,

$$\mu_t^{S_i}(D_{k'(c'+1)}(n,t)) \asymp_{k,c,\bowtie} 1 \asymp_{k,c,\bowtie} \mu_t^{S_i}(D_{2k'(c'+1)}(n,t)).$$

Hence $\mu_t^{S_i}(D) \asymp_{k,c,\bowtie} \#Z$. Furthermore, by Lemma 3.2 we also have, $\forall (n,t) \in Z$,

$$\mu_{f_i(t)}^{S_i'}\big(D_1(\Phi_i(n,t))\big) \asymp_{k,c,\bowtie} 1 \asymp_{k,c,\bowtie} \mu_{f_i(t)}^{S_i'}\big(D_{2k'\cdot k'(c'+1)+c'+1}(\Phi_i(n,t))\big).$$

Therefore

$$\mu_t^{S_i}(D) \asymp_{k,c,\bowtie} \#Z \asymp_{k,c,\bowtie} \mu_{f_i(t)}^{S_i'}\big(\mathcal{N}_1(\widetilde{\widetilde{\Phi}}_i(D))\big). \qquad \square$$

Lemma 5.8 (quasi-isometries quasi-translate the height) *Let $f_i : \mathbb{R} \to \mathbb{R}$ be the function involved in Lemma 5.6. Then, for all $t \in \mathbb{R}$,*

$$\left| \frac{\operatorname{tr}(A_1)}{\operatorname{tr}(A_1')} t - (f_i(t) - f_i(0)) \right| \preceq_{k,c,\bowtie} 1.$$

Proof We recall that for all $t \in \mathbb{R}$, we have $f_i(t) := h(\widehat{\Phi}_i(e_{N_i}, t))$. Let $n \in N_i$, $r > 0$, $t \in \mathbb{R}$, and let us define $U \subset N_i$ such that $D_r(n,0) = (U,0)$. Then we have

$$\mu_0^{S_i}(U,0) = e^{2\operatorname{tr}(A_i)t} \mu_t^{S_i}(U,t). \tag{116}$$

However $\widetilde{\Phi}_i(U,0) = (\Psi_i(U),\, f_i(0))$ and $\widetilde{\Phi}_i(U,t) = (\Psi_i(U),\, f_i(t))$. Therefore

$$\begin{aligned} \mu_{f_i(0)}^{S_i'}\big(\mathcal{N}_1(\widetilde{\Phi}_i(U,0))\big) &= \mu_{f_i(0)}^{S_i'}\big(\mathcal{N}_1(\Psi_i(U),\, f_i(0))\big) \\ &= e^{2\operatorname{tr}(A_i')(f_i(t)-f_i(0))} \mu_{f_i(t)}^{S_i'}\big(\mathcal{N}_1(\Psi_i(U),\, f_i(t))\big) \\ &= e^{2\operatorname{tr}(A_i')(f_i(t)-f_i(0))} \mu_{f_i(t)}^{S_i'}\big(\mathcal{N}_1(\widetilde{\Phi}_i(U,t))\big). \end{aligned} \tag{117}$$

Furthermore by Corollary 5.7 we have

$$\mu_0^{S_i}(U,0) \asymp_{k,c,\bowtie} \mu_{f_i(0)}^{S_i'}\big(\mathcal{N}_1(\widetilde{\Phi}_i(U,0))\big);$$
$$\mu_t^{S_i}(U,t) \asymp_{k,c,\bowtie} \mu_{f_i(t)}^{S_i'}\big(\mathcal{N}_1(\widetilde{\Phi}_i(U,t))\big).$$

In combination with equalities (116) and (117), it provides us with

$$\begin{aligned}\mu_0^{S_i}(U,0) &= e^{2\operatorname{tr}(A_i)t}\mu_t(U,t) \asymp_{k,c,\bowtie} e^{2\operatorname{tr}(A_i)t}\mu_{f_i(t)}^{S_i'}\big(\mathcal{N}_1(\widetilde{\Phi}_i(U,t))\big)\\ &= e^{2\operatorname{tr}(A_i)t}e^{2\operatorname{tr}(A_i')(f_i(0)-f_i(t))}\mu_{f_i(0)}^{S_i'}\big(\mathcal{N}_1(\widetilde{\Phi}_i(U,0))\big)\\ &\asymp_{k,c,\bowtie} e^{2\operatorname{tr}(A_i)t}e^{2\operatorname{tr}(A_i)(f_i(0)-f_i(t))}\mu_0^{S_i}(U,0).\end{aligned}$$

Hence we have $e^{2\operatorname{tr}(A_i)t} \asymp_{k,c,\bowtie} e^{2\operatorname{tr}(A_i')(f_i(t)-f_i(0))}$, which, composed with the logarithm, gives us

$$\left|\frac{\operatorname{tr}(A_1)}{\operatorname{tr}(A_1')}t - (f_i(t)-f_i(0))\right| \preceq_{k,c,\bowtie} 1. \tag{118}$$

□

Corollary 5.9 *There exists $t_0 \in \mathbb{R}$ such that, for $i \in \{1,2\}$ and for all $(n,t) \in N_i \times \mathbb{R}$,*

$$d_{S_i}\left(\widehat{\Phi}_i(n,t), \left(\Psi_i(n), \frac{\operatorname{tr}(A_1)}{\operatorname{tr}(A_1')}t + t_0\right)\right) \preceq_{k,c,\bowtie} 1.$$

Proof The proof is a direct application of Lemmas 5.6 and 5.8 by taking $t_0 := f_i(0)$. □

In this corollary t_0 depends on Φ.

Proof of Theorem 5.5 Using Corollary 5.9 on N_1 and N_2 provides us with Theorem 5.5. □

5.3 Hamenstädt distance and product maps of bilipschitz maps

As presented in [6, Section 5.3], the parabolic visual boundary of $N_i \rtimes \mathbb{R}$ may be identified with the Lie group N_i endowed with the following A_i-homogeneous Hamenstädt distance.

Definition 5.10 (Hamenstädt distance) For any $n, m \in N_i$, we define their Hamenstädt distance as

$$d_{N_i,A_i,H}(n,m) := \exp\big(-\tfrac{1}{2}\lim_{s\to+\infty}\big(2s - d_{N_i\rtimes_{A_i}\mathbb{R}}((n,-s),(m,-s))\big)\big).$$

We might omit A_i and N_i in the notation. We denote by $\mathrm{Bilip}(N)$ the group of bilipschitz maps of N for the Hamenstädt distance. It is defined as

$$\mathrm{Bilip}(N_i) := \{\Psi : (N_i, d_H) \to (N_i, d_H) \mid \exists k \geq 1, \Psi \text{ is a } (k,0)\text{-quasi-isometry}\}.$$

This is indeed a distance when the left invariant metric g is normalised so that $\mathbb{R} \ltimes_{A_i} N_i$ is a $\mathrm{CAT}(-1)$ space.

Two quasi-isometries Φ and Φ' are said to be equivalent when they are at finite distance from each other:

$$\Phi \sim \Phi' \iff \sup_x d_{\bowtie}(\Phi(x), \Phi'(x)) < +\infty.$$

In this section we prove the following characterisation of the quasi-isometry group of the product $S_1 \bowtie S_2 = (N_1 \times N_2) \rtimes_A \mathbb{R}$.

Theorem 5.11 *Let $N_1 \rtimes_{A_1} \mathbb{R}$ and $N_2 \rtimes_{A_2} \mathbb{R}$ be two Heintze group such that $\mathrm{tr}(A_1) \neq \mathrm{tr}(A_1)$, let $\Phi \in \mathrm{QI}((N_1 \times N_2) \rtimes_A \mathbb{R})$ and let Ψ_1, Ψ_2 be as in Theorem 5.5. Then we have the isomorphism*

$$f : \mathrm{QI}((N_1 \times N_2) \rtimes_A \mathbb{R})/\sim \to \mathrm{Bilip}(N_1) \times \mathrm{Bilip}(N_2), \quad \Phi \mapsto (\Psi_1, \Psi_2).$$

This distance is related to the height divergence of vertical geodesics in the following way.

Lemma 5.12 (extended backward lemma) *Let $n, m \in N_i$, let $V : t \mapsto (n, t)$ and let $W : t \mapsto (m, t)$. Then*

$$d_H(n, m) \preceq_{k,c,\bowtie} \exp(h_{\mathrm{Div}}(V, W)).$$

See Corollary 1.4 for the definition of $h_{\mathrm{Div}}(V, W)$.

Proof By Corollary 1.4 there exists a height $h_{\mathrm{Div}}(V, W) \in \mathbb{R}$ such that V and W diverge from each other at the height $h_{\mathrm{Div}}(V, W)$. Hence there exists $M(k, c, \bowtie)$ such that, for all $s_1 \leq s_2 \leq h_{\mathrm{Div}}(V, W)$,

$$d(V(s_2), W(s_2)) - M \leq d_{S_i}(V(s_1), W(s_1)) + 2|s_2 - s_1| \leq d_{S_i}(V(s_2), W(s_2)) + M.$$

Therefore

$$\exp\big(d_{S_i}(V(s_1), W(s_1)) + 2|s_2 - s_1|\big) \preceq_{k,c,\bowtie} \exp\big(d_{S_i}(V(s_2), W(s_2))\big). \tag{119}$$

Let us define $h_0 := h_{\mathrm{Div}}(V, W)$. Then we can compute the Hamenstädt distance $d_H(n, m)$:

$$\begin{aligned}
d_H(n, m) &= \exp\big(-\tfrac{1}{2} \lim_{s \to +\infty} \big(2s - d_{S_i}(V(-s), W(-s))\big)\big) \\
&\preceq_{k,c,\bowtie} \exp\big(-\tfrac{1}{2} \lim_{s \to +\infty} \big(2s - d_{S_i}(V(h_0), W(h_0)) - (2h_0 + 2s)\big)\big) \quad \text{(by inequality (119))} \\
&\preceq_{k,c,\bowtie} \exp\big(-\tfrac{1}{2} \lim_{s \to +\infty} \big(-d_{S_i}(V(h_0), W(h_0)) - 2h_0\big)\big) \\
&= \exp\left(\frac{d_{S_i}(V(h_0), W(h_0))}{2} + h_0\right) = \exp\left(\frac{d_{S_i}(V(h_0), W(h_0))}{2}\right) \exp(h_0) \\
&\preceq_{k,c,\bowtie} \exp(h_0) \quad \text{(by definition of } h_{\mathrm{Div}}(V, W)).
\end{aligned}$$

□

We show that the aforementioned maps Ψ_i are bilipschitz.

Theorem 5.13 *Let Ψ_i be the map of Theorem 5.5. Then Ψ_i is a bilipschitz homeomorphism either from (N_i, d_H) to $(N_i', (d_H)^{\mathrm{tr}(A_1)/\mathrm{tr}(A_1')})$ or from $(N_i, (d_H)^{\mathrm{tr}(A_1')/\mathrm{tr}(A_1)})$ to (N_i', d_H).*

Proof Let $n, m \in N_i$ and let $V : t \mapsto (n, t)$ and $W : t \mapsto (m, t)$ be two vertical geodesics of $N_i \rtimes_{A_i} \mathbb{R}$. Let us define $\lambda_0 := \frac{\mathrm{tr}(A_1)}{\mathrm{tr}(A_1')}$. By Lemma 5.12 we have

$$d_H(n, m) \preceq_{k,c,\bowtie} \exp(h_{\mathrm{Div}}(V, W)).$$

Since $\Phi_i := (\Psi_i, \lambda_0 \mathrm{id}_{\mathbb{R}} + t_0)$ is a (k', c')-quasi-isometry, we have

(1) $d_{S_i}\big((\Psi_i(n), \lambda_0 h_{\mathrm{Div}}(V, W) + t_0), (\Psi_i(m), \lambda_0 h_{\mathrm{Div}}(V, W) + t_0)\big) \preceq_{k,c,\bowtie} 1$;

(2) $\forall s \geq h_{\mathrm{Div}}(V, W)$, $d_{S_i}\big((\Psi_i(n), \lambda_0 s + t_0), (\Psi_i(m), \lambda_0 s + t_0)\big) \preceq_{k,c,\bowtie} 1$.

Furthermore, for all $n \in N_i$, $\widetilde{\Phi}_i(V_n) = V_{\Psi_i(n)}$ hence $\widetilde{\Phi}_i(V_n)$ is a vertical geodesics of S_i'. Then there exists $M(k, c, \bowtie)$ such that

$$(\lambda_0 h_{\mathrm{Div}}(V, W) + t_0) - M \leq h_{\mathrm{Div}}(\widetilde{\Phi}_i(V), \widetilde{\Phi}_i(W)) \leq (\lambda_0 h_{\mathrm{Div}}(V, W) + t_0) + M.$$

Consequently Lemma 5.12 provides us with

$$\begin{aligned} d_H(\Psi_i(n), \Psi_i(m)) &\asymp_{k,c,\bowtie} \exp(h_{\mathrm{Div}}(V_{\Psi_i(n)}, W_{\Psi_i(m)})) = \exp\big(h_{\mathrm{Div}}(\widetilde{\Phi}_i(V), \widetilde{\Phi}_i(W))\big) \\ &\asymp_{k,c,\bowtie} \exp(t_0) \exp(\lambda_0 h_{\mathrm{Div}}(V, W)) \\ &\asymp_{k,c,\bowtie} \exp(t_0)(d_H(n, m))^{\lambda_0} \quad \text{(by Lemma 5.12).} \end{aligned}$$

Here t_0 depends only on Φ. Furthermore, if $\lambda_0 \leq 1$, $(d_H)^{\lambda_0}$ is still a distance by concavity. Hence, depending on the value of λ_0, either $\Psi_i : (N_i, d_H) \to (N_i', (d_H)^{\lambda_0})$ or $\Psi_i : (N_i, (d_H)^{\lambda_0}) \to (N_i', d_H)$ is a bilipschitz map. □

We now focuses on self quasi-isometries of $(N_1 \times N_2) \rtimes_A \mathbb{R}$.

Proof of Theorem 5.11: Let Ψ_1, Ψ_2 be as in Theorem 5.5, and let f be the map

$$f : \mathrm{QI}((N_1 \times N_2) \rtimes_A \mathbb{R})/\sim \to \mathrm{Bilip}(N_1) \times \mathrm{Bilip}(N_2), \quad \Phi \mapsto (\Psi_1, \Psi_2).$$

We first show that this application is well defined. Let $\Phi, \Phi' \in \mathrm{QI}((N_1 \times N_2) \rtimes_A \mathbb{R})$ be such that $\Phi \sim \Phi'$, which means that $d_{\bowtie}(\Phi, \Phi') \preceq_{k,c,\bowtie} 1$.

By Theorems 5.5 and 5.13, there exist $\Psi_i, \Psi_i' \in \mathrm{Bilip}(N_i)$ such that

(1) $d(\Phi, (\Psi_1, \Psi_2, \mathrm{id}_{\mathbb{R}})) \preceq_{k,c,\bowtie} 1$;

(2) $f(\Phi) = (\Psi_1, \Psi_2)$;

(3) $d(\Phi', (\Psi_1', \Psi_2', \mathrm{id}_{\mathbb{R}})) \preceq_{k,c,\bowtie} 1$;

(4) $f(\Phi') = (\Psi_1', \Psi_2')$.

By the definition of Ψ_i and Ψ_i', for all $n \in N$ we have

$$\begin{aligned} \Psi_i(n) &= P(V_n'(0)); \\ \Psi_i'(n) &= P(V_n''(0)), \end{aligned}$$

where V_n' is the unique vertical geodesic close to $\widehat{\Phi}_i(V_n)$ and V_n'' the unique vertical geodesic close to $\widehat{\Phi}_i'(V_n)$. However $\Phi \sim \Phi'$, then $\widehat{\Phi}_i(V_n)$ and $\widehat{\Phi}_i'(V_n)$ are M-close to each other for some $M(k, c, \bowtie)$, therefore $d_{\mathrm{Hff}}(V_n', V_n'') \preceq_{k,c,\bowtie} 1$. However these vertical geodesics are unique, then $V_n' = V_n''$. Consequently, $\Psi_i(n) = \Psi_i'(n)$, hence $\Psi_i = \Psi_i'$, therefore f is well defined.

Let us now prove that f is injective. Let Φ and Φ' be two quasi-isometries of $(N_1 \times N_2) \rtimes_A \mathbb{R}$ such that $f(\Phi) = f(\Phi')$. Then by Theorem 5.5 and by the triangle inequality

$$d_{\bowtie}(\Phi, \Phi') \leq d_{\bowtie}(\Phi, (\Psi_1, \Psi_2, \mathrm{id}_{\mathbb{R}})) + d_{\bowtie}((\Psi_1, \Psi_2, \mathrm{id}_{\mathbb{R}}), \Phi') \preceq_{k,c,\bowtie,\Phi,\Phi'} 1.$$

Hence $\Phi \sim \Phi'$, which proves that f is injective.

Let $\Psi_i \in \mathrm{Bilip}(N_i, d_H)$, our goal is to show that $(\Psi_i, \mathrm{id}_{\mathbb{R}})$ is a quasi-isometry of $(N_i \rtimes_A \mathbb{R}, d_{S_i})$. Let $(n, t_n), (m, t_m) \in S_i$. By Lemma 5.12 applied on n and m, there exists a constant $M(k, c, \bowtie)$ such that

$$\ln(d_H(n,m)) - M \leq h_{\mathrm{Div}}(V_n, V_m) \leq \ln(d_H(n,m)) + M. \tag{120}$$

Similarly, by Lemma 5.12 applied on $\Psi_i(n)$ and $\Psi_i(m)$,

$$\ln\big(d_H(\Psi_i(n), \Psi_i(m))\big) - M \leq h_{\mathrm{Div}}(V_{\Psi_i(n)}, V_{\Psi_i(m)}) \leq \ln\big(d_H(\Psi_i(n), \Psi_i(m))\big) + M. \tag{121}$$

We know that $\Psi_i \in \mathrm{Bilip}(N_i, d_H)$ hence $d_H(n,m) \asymp d_H(\Psi_i(n), \Psi_i(m))$. Therefore by inequalities (120) and (121) we have

$$|h_{\mathrm{Div}}(V_n, V_m) - h_{\mathrm{Div}}(V_{\Psi_i(n)}, V_{\Psi_i(m)})| \preceq 1. \tag{122}$$

Moreover by Lemma 1.3 we can characterise the distance between two points thanks to the height of divergence of their associated vertical geodesics. Let us denote $h_0 = h_{\mathrm{Div}}(V_n, V_m)$. By inequality (122) and by Lemma 1.3, if $h_0 \geq \max(t_n, t_m)$ we have both

$$\begin{aligned} \big|d_{S_i}((n,t_n),(m,t_m)) - (|t_m - h_0| + |t_n - h_0|)\big| &\preceq_\delta 1; \\ \big|d_{S_i}\big((\Psi_i(n),t_n),(\Psi_i(m),t_m)\big) - (|t_m - h_0| + |t_n - h_0|)\big| &\preceq_\delta 1. \end{aligned}$$

Consequently by the triangle inequality there exists $M(\delta)$ such that

$$d_{S_i}((n,t_n),(m,t_m)) - M \leq d_{S_i}\big((\Psi_i(n),t_n),(\Psi_i(m),t_m)\big) \leq d_{S_i}((n,t_n),(m,t_m)) + M.$$

Similarly, if $h_0 \leq \max(t_n, t_m)$ we have both

$$\begin{aligned} \big|d_{S_i}((n,t_n),(m,t_m)) - (|t_m - t_n|)\big| &\preceq_\delta 1; \\ \big|d_{S_i}\big((\Psi_i(n),t_n),(\Psi_i(m),t_m)\big) - (|t_m - t_n|)\big| &\preceq_\delta 1. \end{aligned}$$

Hence again

$$d_{S_i}((n,t_n),(m,t_m)) - M \leq d_{S_i}\big((\Psi_i(n),t_n),(\Psi_i(m),t_m)\big) \leq d_{S_i}((n,t_n),(m,t_m)) + M.$$

Therefore $(\Psi_i, \mathrm{id}_{\mathbb{R}})$ is a $(1, M)$-quasi-isometry of $N_i \rtimes \mathbb{R}$, hence $(\Psi_1, \Psi_2, \mathrm{id}_{\mathbb{R}})$ is also a $(1, M)$-quasi-isometry, which provides us with $f(\Psi_1, \Psi_2, \mathrm{id}_{\mathbb{R}}) = (\Psi_1, \Psi_2)$. Hence f is surjective, and finally bijective.

Let us now prove that f is a morphism. Let $\Phi, \Phi' \in \mathrm{QI}((N_1 \times N_2) \rtimes_A \mathbb{R})$. Furthermore, we have $d_{\bowtie}(\Phi', (\Psi_1', \Psi_2', \mathrm{id}_{\mathbb{R}})) \preceq 1$, hence $d_{\bowtie}(\Phi \circ \Phi', \Phi \circ (\Psi_1', \Psi_2', \mathrm{id}_{\mathbb{R}})) \preceq 1$ since Φ is a quasi-isometry. Moreover, $d_{\bowtie}(\Phi, (\Psi_1, \Psi_2, \mathrm{id}_{\mathbb{R}})) \preceq 1$, therefore by the triangle inequality

$$d_{\bowtie}(\Phi \circ \Phi', (\Psi_1, \Psi_2, \mathrm{id}_{\mathbb{R}}) \circ (\Psi_1', \Psi_2', \mathrm{id}_{\mathbb{R}})) \preceq 1.$$

However

$$(\Psi_1, \Psi_2, \mathrm{id}_{\mathbb{R}}) \circ (\Psi_1', \Psi_2', \mathrm{id}_{\mathbb{R}}) = (\Psi_1 \circ \Psi_1', \Psi_2 \circ \Psi_2', \mathrm{id}_{\mathbb{R}}),$$

which provides us with

$$d_{\bowtie}(\Phi \circ \Phi', (\Psi_1 \circ \Psi_1', \Psi_2 \circ \Psi_2', \mathrm{id}_{\mathbb{R}})) \preceq 1.$$

Consequently $f(\Phi \circ \Phi') = (\Psi_1 \circ \Psi_1', \Psi_2 \circ \Psi_2')$. □

In this proof we showed that $\Phi \sim (\Psi_1, \Psi_2, \mathrm{id}_{\mathbb{R}})$, therefore any quasi-isometry is in the equivalence class of an $(1, M)$-quasi-isometry.

5.4 Quasi-isometric classification and necessary conditions to being quasi-isometric

Thanks to Proposition 4.15 and Theorem 5.13 we are able to provide necessary conditions and quasi-isometric classifications for families of solvable Lie groups of the form $\mathbb{R} \ltimes_{\mathrm{Diag}(A_1,A_2)} (N_1 \times N_2)$.

Let us recall two consequences implied by being quasi-isometric in the Lie group setting. For $i \in \{1, 2\}$, let N_i, N_i' be two simply connected, nilpotent Lie groups and let A_i, A_i' be two matrices whom eigenvalues have positive real parts, acting by derivation on the corresponding Lie algebra. Let us assume that $\mathrm{tr}(A_1) > \mathrm{tr}(A_2)$ and $\mathrm{tr}(A_1') > \mathrm{tr}(A_2')$. If $\mathbb{R} \ltimes_{\mathrm{Diag}(A_1,A_2)} (N_1 \times N_2)$ and $\mathbb{R} \ltimes_{\mathrm{Diag}(A_1',A_2')} (N_1' \times N_2')$ are quasi-isometric then

(1) $\frac{\mathrm{tr}(A_1)}{\mathrm{tr}(A_2)} = \frac{\mathrm{tr}(A_1')}{\mathrm{tr}(A_2')}$ (Proposition 4.15);

(2) for $i \in \{1, 2\}$, N_i and N_i' are bilipschitz. (Theorem 5.13).

Let us define

$$S_{N_1,N_2} := \mathbb{R} \ltimes_{\mathrm{Diag}(A_1,-A_2)} (N_1 \times N_2).$$

Combining [2, Lemma 4.1] and Theorem 5.13 we obtain the following statement.

Proposition 5.14 *Let us assume that* $\mathrm{tr}(A_1) > \mathrm{tr}(A_2)$ *and* $\mathrm{tr}(A_1') > \mathrm{tr}(A_2')$. *If* S_{N_1,N_2} *and* $S_{N_1',N_2'}$ *are quasi-isometric, then we have that for* $i \in \{1, 2\}$, A_i *and* $\frac{\mathrm{tr}(A_1)}{\mathrm{tr}(A_1')} A_i'$ *share the same characteristic polynomial.*

A *Carnot* group N is a simply connected, nilpotent Lie group with a Lie algebra $\mathrm{Lie}(N)$ which admits a grading: there exists a family of subspaces V_i with $i \in \{1, \dots, r\}$ for some $r \geq 1$ such that $V_{i+1} = [V_1, V_i]$ for $i < r$ and such that

$$\mathrm{Lie}(N) = \bigoplus_{i=1}^{r} V_i.$$

A Carnot group is equipped with a 1-parameter family of automorphisms called *dilations* on N and defined for $t \in \mathbb{R}$ by $\delta_t := \exp(tD)$, with D a Lie derivation on $\mathrm{Lie}(N)$ satisfying that $Dv = iv$ for $v \in V_i$ and $i \in \{1, \dots, r\}$. Such a derivation is called a *Carnot derivation*. A Lie group $S(N_1, N_2)$ is *Carnot-Sol type* if N_1 and N_2 are Carnot groups and if their respective derivations A_1 and A_2 are Carnot derivations. Combining Theorem 5.13 and [19, Theorem 2], we get the following necessary condition.

Proposition 5.15 *Let* $S(N_1, N_2)$ *and* $S(N_1', N_2')$ *be two Carnot-Sol type Lie groups and assume that* $\mathrm{tr}(A_1) > \mathrm{tr}(A_2)$ *and that* $\mathrm{tr}(A_1') > \mathrm{tr}(A_2')$. *Then*

$$S_{N_1,N_2} \text{ and } S_{N_1',N_2'} \text{ are quasi-isometric} \implies \text{for } i \in \{1, 2\},\ N_i \text{ and } N_i' \text{ are isomorphic.}$$

Furthermore, for a given Carnot derivation A on a Carnot group N, there exists a positive real $\alpha > 0$ such that $\mathbb{R} \ltimes_A N = S_\alpha$ where $S_\alpha := \mathbb{R} \ltimes_\alpha N$ is the group defined by the action of $\mathbb{R}$ via the dilation

$(\delta_{\alpha t})_{t\in\mathbb{R}}$ on N. Let N_1 and N_2 be two Carnot groups and for any two positive reals $\alpha, \beta > 0$, let $G_{\alpha,\beta} := \mathbb{R} \ltimes_{\alpha,-\beta} (N_1 \times N_2)$ be the group defined by the action of $\mathbb{R}$ on $N_1 \times N_2$,

$$\mathbb{R} \to \mathrm{Aut}(N \times N), \quad t \mapsto (\delta_{\alpha t}, \delta_{-\beta t}).$$

Note that $G_{\alpha,\beta} = S_\alpha \bowtie S_\beta$. Thanks to the quasi-isometry invariant of Proposition 4.15, we obtain the quasi-isometry classification for Carnot-Sol type Lie groups.

Proposition 5.16 *Let (α, β) and (σ, τ) be two pairs of positive reals with $\alpha > \beta$ and $\sigma > \tau$. Then*

$$G_{\alpha,\beta} \text{ quasi-isometric to } G_{\sigma,\tau} \iff \frac{\alpha}{\beta} = \frac{\sigma}{\tau} \iff G_{\alpha,\beta} \text{ isomorphic to } G_{\sigma,\tau}.$$

Proof If $\frac{\alpha}{\beta} = \frac{\sigma}{\tau}$, then $G_{\alpha,\beta}$ and $G_{\sigma,\tau}$ are isomorphic and thus in particular quasi-isometric (or even bilipschitz) with respect to any left-invariant Riemannian metrics on the groups. Indeed, the map

$$G_{\alpha,\beta} \to G_{\sigma,\tau}, \quad (x, y, t) \mapsto (x, y, \lambda t),$$

is an isomorphism. For $(x_i, y_i, t_i) \in G_{\alpha,\beta}$ for $i \in \{1, 2\}$, we have, in $G_{\sigma,\tau}$,

$$\begin{aligned}(x_1, y_1, \lambda t_1) \cdot (x_2, y_2, \lambda t_2) &= (x_1 \cdot \delta_{\sigma\lambda t_1} x_2, y_1 \cdot \delta_{-\tau\lambda t_1} y_2, \lambda(t_1 + t_2)) \\ &= (x_1 \cdot \delta_{\alpha t_1} x_2, y_1 \cdot \delta_{-\beta t_1} y_2, \lambda(t_1 + t_2)),\end{aligned}$$

which is the image of $(x_1, y_1, t_1) \cdot (x_2, y_2, t_2)$. Proposition 4.15 conclude the proof since the ratios of traces of the respective derivations are $\frac{\alpha}{\beta}$ and $\frac{\sigma}{\tau}$. □

Otherwise stated, two nonunimodular Carnot-Sol type solvable Lie groups are quasi-isometric if and only if they are isomorphic.

Acknowledgements

This work was supported by the University of Montpellier, as well as by the Japan Society for the Promotion of Science (JSPS) and the ERC of Enrico Le Donne.

I am deeply grateful to my advisors, Jérémie Brieussel and Constantin Vernicos, for their guidance, and for their many reviews and comments. I also wish to thank Xiangdong Xie for the early and stimulating discussions regarding the paper of Eskin, Fisher and Whyte. My thanks extend to Ryokichi Tanaka and Gabriel Pallier for their detailed explanations of the implications of geometric rigidity in the context of Lie groups. Lastly, I am particularly grateful to the reviewer for a meticulous reading, detailed feedback, and thoughtful suggestions, which greatly improved the quality of this paper.

References

[1] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR

[2] **M Carrasco Piaggio**, **E Sequeira**, *On quasi-isometry invariants associated to a Heintze group*, Geom. Dedicata 189 (2017) 1–16 MR

[3] **M Coornaert**, **T Delzant**, **A Papadopoulos**, *Géométrie et théorie des groupes: les groupes hyperboliques de Gromov*, Lecture Notes in Mathematics 1441, Springer (1990) MR

[4] **Y de Cornulier**, *Asymptotic cones of Lie groups and cone equivalences*, Illinois J. Math. 55:1 (2011) 237–259 MR

[5] **Y de Cornulier**, *On the quasi-isometric classification of locally compact groups*, from "New directions in locally compact groups" (P-E Caprace, N Monod, editors), London Math. Soc. Lecture Note Ser. 447, Cambridge Univ. Press (2018) 275–342 MR

[6] **M G Cowling**, **V Kivioja**, **E Le Donne**, **S Nicolussi Golo**, **A Ottazzi**, *From homogeneous metric spaces to Lie groups*, C. R. Math. Acad. Sci. Paris 362 (2024) 943–1014 MR

[7] **T Dymarz**, *Large scale geometry of certain solvable groups*, Geom. Funct. Anal. 19:6 (2010) 1650–1687 MR

[8] **A Eskin**, **D Fisher**, **K Whyte**, *Coarse differentiation of quasi-isometries, I*: *Spaces not quasi-isometric to Cayley graphs*, Ann. of Math. (2) 176:1 (2012) 221–260 MR

[9] **A Eskin**, **D Fisher**, **K Whyte**, *Coarse differentiation of quasi-isometries, II*: *Rigidity for Sol and lamplighter groups*, Ann. of Math. (2) 177:3 (2013) 869–910 MR

[10] **B Farb**, **L Mosher**, *A rigidity theorem for the solvable Baumslag–Solitar groups*, Invent. Math. 131:2 (1998) 419–451 MR

[11] **T Ferragut**, *Geodesics and visual boundary of horospherical products* (2020) arXiv 2009.04698

[12] **E Ghys**, **P de la Harpe**, *Espaces métriques hyperboliques*, from "Sur les groupes hyperboliques d'après Mikhael Gromov" (Bern, 1988) (E Ghys, P de la Harpe, editors), Progr. Math. 83, Birkhäuser, Boston, MA (1990) 27–45 MR

[13] **S Gouëzel**, **V Shchur**, *A corrected quantitative version of the Morse lemma*, J. Funct. Anal. 277:4 (2019) 1258–1268 MR

[14] **M Gromov**, *Asymptotic invariants of infinite groups*, from "Geometric group theory, II" (Sussex, 1991) (G A Niblo, M A Roller, editors), London Math. Soc. Lecture Note Ser. 182, Cambridge Univ. Press (1993) 1–295 MR

[15] **J Heinonen**, *Lectures on analysis on metric spaces*, Springer (2001) MR

[16] **E Heintze**, *On homogeneous manifolds of negative curvature*, Math. Ann. 211 (1974) 23–34 MR

[17] **M Kapovich**, *Lectures on quasi-isometric rigidity*, from "Geometric group theory" (M Bestvina, M Sageev, K Vogtmann, editors), IAS/Park City Math. Ser. 21, Amer. Math. Soc., Providence, RI (2014) 127–172 MR

[18] **E Le Donne**, **G Pallier**, **X Xie**, *Rough similarity of left-invariant Riemannian metrics on some Lie groups*, Groups Geom. Dyn. 19:1 (2025) 227–263 MR

[19] **P Pansu**, *Métriques de Carnot–Carathéodory et quasiisométries des espaces symétriques de rang un*, Ann. of Math. (2) 129:1 (1989) 1–60 MR

[20] **A Papadopoulos**, *Metric spaces, convexity and nonpositive curvature*, IRMA Lectures in Mathematics and Theoretical Physics 6, European Mathematical Society, Zürich (2005) MR

[21] **I Peng**, *Coarse differentiation and quasi-isometries of a class of solvable Lie groups, I*, Geom. Topol. 15:4 (2011) 1883–1925 MR

[22] **I Peng**, *Coarse differentiation and quasi-isometries of a class of solvable Lie groups, II*, Geom. Topol. 15:4 (2011) 1927–1981 MR

[23] **P M Soardi**, **W Woess**, *Amenability, unimodularity, and the spectral radius of random walks on infinite graphs*, Math. Z. 205:3 (1990) 471–486 MR

[24] **W Woess**, *What is a horocyclic product*, *and how is it related to lamplighters*?, Internationale Mathematische Nachrichten 224 (2013) 1–27

[25] **X Xie**, *Large scale geometry of negatively curved* $\mathbb{R}^n \rtimes \mathbb{R}$, Geom. Topol. 18:2 (2014) 831–872 MR

TOM FERRAGUT tom.ferragut@univ-lyon1.fr
University Lyon 1, Villeurbanne, France

Received: December 16, 2022 Revised: February 4, 2025

msp

Algebraic & Geometric Topology **26**:3 (2026) 955–971
DOI: 10.2140/agt.2026.26.955
Published: April 1, 2026

Large volume fibered knots in 3-manifolds

J ROBERT OAKLEY

We prove that for hyperbolic fibered knots in any closed, connected, oriented 3-manifold the volume and genus are unrelated. As an application we answer a question of Hirose, Kalfagianni, and Kin about volumes of mapping tori that are double branched covers.

1 Introduction

Let M denote a closed, connected, oriented 3-manifold. Alexander in [1] proved that every such M contains a fibered link. This result has been strengthened in many ways such as the link being a knot [21] and the monodromy being right-veering and pseudo-Anosov [8]. More recently, this construction has been used in [2] to show that for hyperbolic fibered knots in the 3-sphere, volume and genus are unrelated. In this paper we generalize the approach in [2] to show that for hyperbolic, fibered knots in closed, connected, oriented 3-manifolds the volume and genus are unrelated. In particular we prove the following theorem.

Theorem 1.1 *Let M be a closed, connected, oriented 3-manifold. There exists some $g_0 > 1$ such that for all $g \geq g_0$ and $V > 0$ there exists a knot $K \subseteq M$ such that $M - K$ is fibered over the circle with genus g and $M - K$ is hyperbolic with* $\mathrm{vol}(M - K) > V$.

Let $\mathfrak{D}_g(M) \subseteq \mathrm{Mod}(S_g)$ be the subset of the mapping class group of a closed surface of genus g consisting of elements whose mapping tori are 2-fold branched covers of M branched along a link. As an application of Theorem 1.1 we answer the following question asked by Hirose, Kalfagianni, and Kin in [15].

Question 1 *For g sufficiently large, does $\mathfrak{D}_g(M)$ contain an infinite family of pseudo-Anosov mapping classes whose mapping tori have arbitrarily large volume*?

Combining Theorem 1.1 and Theorem 10 of [15] yields the following affirmative answer to Question 1.

Corollary 1.2 *For any closed, connected, oriented 3-manifold M there exists some $g_0 > 1$ such that for any $g > g_0$ the set $\mathfrak{D}_{2g}(M)$ contains an infinite family of pseudo-Anosov elements whose mapping tori have arbitrarily large volume.*

1.1 Outline of the proof

The strategy of the proof is to construct for every closed, connected, oriented 3-manifold M and for every sufficiently large genus g of the fiber, a family of fibered hyperbolic knots of genus g with linearly growing volume. We use three main tools to construct these families. The first is the theory of branched covers in dimensions 2 and 3. Using branched covers of the 3-sphere branched over a knot or link in braid

MSC2020: 57K10, 57K20, 57K32, 57M12.

position to obtain a fibered link in a 3-manifold goes back to Alexander. Here we use an improvement on Alexander's theorem due to Hilden and Montesinos [14; 20]. This refinement allows us to consider simple 3-fold branched covers. That the branched covers are degree 3 is crucial for control of the number of preimages of the braid axis which will become the fibered link. This control over the degree of the cover comes at the expense of the covers considered by Hilden and Montesinos being irregular. Fortunately these irregular covers have the property of being *simple* (see Definition 3.1). Simple branched covers are well studied in dimension 2 (see, for example, [11; 12; 23]). This allows us to control the monodromy of the fibered link.

In Section 4 we use this control of the monodromy and the second tool, open book decompositions and stabilization, to ensure that our fibered link becomes a fibered knot with pseudo-Anosov monodromy (see Section 4 for a brief discussion of open book decompositions). To that end we use the equivalence between open book decompositions and fibered links as well as the technique introduced by Colin and Honda in [8] to transform our fibered 2-component link with possibly reducible monodromy, obtained from the branched cover construction described above, into a fibered knot with pseudo-Anosov monodromy.

Crucially, even after this transformation we maintain the control over the monodromy. This allows us to use our third tool, subsurface projections and Brock's work in [5] relating translation distance of a pseudo-Anosov in the pants graph and the volume of the associated mapping torus. The control over the monodromy that we have maintained throughout the construction of these fibered hyperbolic knots ensures that in addition to the monodromies being pseudo-Anosov, they factor as $(A)(S^{-1}F^nS)(F^{-n})$ where A is constant as n varies and $F^{\pm n}$ are pseudo-Anosov on essential subsurfaces. Then we apply the work of Clay, Leininger, and Mangahas [7] to see that the monodromy has linearly growing subsurface projections to the supports of F^{-n} and $S^{-1}F^nS$. The Masur–Minsky distance formula for the pants graph [19] then implies that the monodromies have linearly growing translation distance in the pants graph. Finally, the work of Brock [5] implies that this linearly growing translation distance corresponds to linearly growing volume of the associated mapping tori.

2 Background

Throughout the paper $S = S_{g,n}$ denotes an orientable surface of genus g and n boundary components. We will now introduce some conventions and aspects of the study of surfaces and their symmetries that will be useful. The starting point of our construction will use the theory of braids in the 3-sphere. We leverage the group-theoretic structure of braids.

Definition 2.1 The *braid group on n strands*, $\mathfrak{B}_n$, is the group of isotopy classes of n braids. If $\eta, \psi \in \mathfrak{B}_n$, then we multiply η and ψ by scaling each of their heights by $\frac{1}{2}$ and stack the braid corresponding to η on top of the braid corresponding to ψ. The resulting braid is $\eta \cdot \psi \in \mathfrak{B}_n$.

The braid group $\mathfrak{B}_n$ is generated by the braids σ_i for $1 \leq i \leq n-1$ where the only crossing is the $(i+1)$-st strand passing over the i-th strand (see Figure 1).

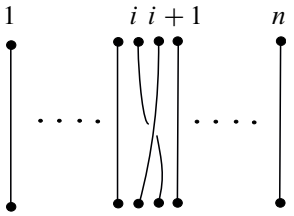


Figure 1: The generator σ_i of the braid group on n strands.

2.1 Mapping class group basics

Definition 2.2 The *mapping class group* of a surface, $\mathrm{Mod}(S_{g,n})$, is $\mathrm{Homeo}^+(S, \partial S)/\text{isotopy}$. This is the group of orientation preserving homeomorphisms of the surface $S_{g,n}$ that restrict to the identity on the boundary, up to isotopy.

For composing elements of $\mathrm{Mod}(S_{g,n})$, called mapping classes, we use functional notation. That is, for $\varphi, \psi \in \mathrm{Mod}(S_{g,n})$ the composition $\varphi \circ \psi$ refers to applying ψ and then applying φ.

Remark 2.3 We will occasionally take the perspective that $\mathfrak{B}_n$ is $\mathrm{Mod}(D_n)$ the mapping class group of the disk with n punctures. See Chapter 9 of [9].

A simple example of a nontrivial element of the mapping class group is a *Dehn twist.* Let $\alpha \subseteq S$ be an essential simple closed curve on a surface S. The Dehn twist about α, denoted by T_α, is the mapping class obtained by cutting S along α, twisting a neighborhood of one boundary component by an angle of 2π, and then regluing the surface. See Figure 2 for an illustration. Dehn twists are an important aspect of the theory of mapping class groups that we will use extensively. In fact, $\mathrm{Mod}(S)$ is generated by a finite collection of Dehn twists about simple closed curves. See Chapter 3 of [9] for an extensive discussion of Dehn twists.

In studying the mapping class group we consider different classes of mapping classes depending on their action on the surface. We say that a mapping class $f \in \mathrm{Mod}(S_{g,n})$ is *reducible* if there is a nonempty set of isotopy classes of mutually disjoint, simple closed curves, $\{a_1, \ldots, a_n\}$ such that

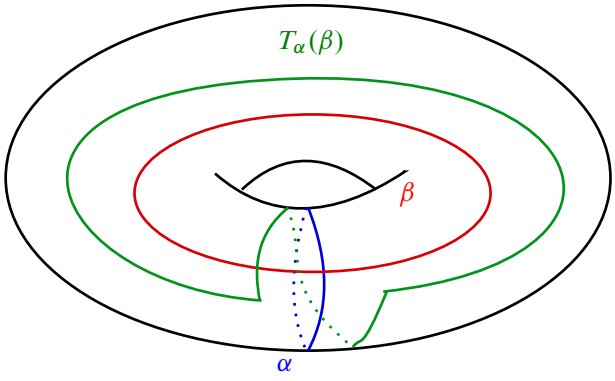


Figure 2: The action of the Dehn twist T_α on the curve β.

$\{a_1, \dots, a_n\} = \{f(a_1), \dots, f(a_n)\}$. We say that a mapping class is *periodic* if it is finite order. We say that a mapping class $f \in \mathrm{Mod}(S_{g,n})$ is *pseudo-Anosov* if there exists a pair of transverse measured foliations $(\mathcal{F}^u, \mu_u)$ and $(\mathcal{F}^s, \mu_s)$ on $S_{g,n}$, a number $\lambda > 1$, and a representative homeomorphism φ such that

$$\varphi(\mathcal{F}^u, \mu_u) = (\mathcal{F}^u, \lambda\mu_u) \quad \text{and} \quad \varphi(\mathcal{F}^s, \mu_s) = (\mathcal{F}^s, \lambda^{-1}\mu_s).$$

See Chapter 13 of [9] for more discussion of the three types of mapping classes. The following theorem says that these three classes characterize all elements of the mapping class group.

Theorem 2.4 (Nielsen–Thurston classification) *Every mapping class in* $\mathrm{Mod}(S_{g,n})$ *is either periodic, reducible, or pseudo-Anosov. Moreover, pseudo-Anosov mapping classes are neither periodic nor reducible.*

An interesting subclass of reducible mapping classes are *partial pseudo-Anosovs.* A mapping class $f \in \mathrm{Mod}(S)$ is called a partial pseudo-Anosov if it satisfies the following conditions.

(1) There exists a subsurface $S' \subseteq S$ such that $f|_{S'}$ is a pseudo-Anosov.

(2) $f|_{S-S'}$ is isotopic to the identity.

In this case, we call S' the support of f.

2.2 The curve graph

Given two isotopy classes of arcs or curves $a, b \subseteq S$, the *geometric intersection number* of a and b is defined to be the minimal number of intersections between any pair of curves or arcs representing each isotopy class. That is,

$$i(a, b) = \min_{\substack{\alpha \in a \\ \beta \in b}} |\alpha \cap \beta|.$$

Definition 2.5 Let $3g + n \geq 5$. Then the *curve graph of a surface* $S = S_{g,n}$, denoted by $\mathcal{C}(S)$, is the simplicial graph with vertices corresponding to isotopy classes of essential simple closed curves and whose edges represent pairs of curves that can be realized disjointly on S. Note that there is an adaption of the definition for surfaces of lower complexity by modifying the edge condition, but all of the surfaces we discuss here satisfy the complexity condition (see [9, Chapter 4]).

We will use $\mathcal{C}^{(0)}(S)$ to denote the 0-skeleton of $\mathcal{C}(S)$. We endow $\mathcal{C}(S)$ with a metric by assigning each edge length 1. With this metric, d_S, the curve graph is a geodesic metric space. There is a natural action of $\mathrm{Mod}(S)$, by isometries, on $\mathcal{C}(S)$. Given $f \in \mathrm{Mod}(S)$ we define the *translation distance of* f, *denoted by* $\tau_{\mathcal{C}}(f)$, as

$$\tau_{\mathcal{C}}(f) = \inf_{\eta \in \mathcal{C}^{(0)}(S)} \{d_S(\eta, f(\eta)\}.$$

We now state the following lemma originally due to Hempel in [13].

Lemma 2.6 *Let $\alpha, \beta \subseteq S$ be simple closed curves representing vertices a and b of $\mathcal{C}(S)$, respectively. Then*

$$d_S(a,b) \leq 2\log_2(i(\alpha,\beta)) + 2.$$

The theory of subsurface projections plays a central role in the theory of mapping class groups, and we introduce the basics of this here. For more discussion of subsurface projections, see Section 2 of [19]. An *essential subsurface* $Y \subseteq S$ is a subsurface such that the interior of Y embeds in S so that each component of ∂Y maps to either an essential simple closed curve in S or a component of ∂S. A *nonannular* subsurface is a subsurface that is not homeomorphic to an annulus. An essential simple closed arc or curve, η, is in *minimal position* with $Y \subseteq S$ when η is in minimal position with all of the components of ∂Y. That is, $i(\eta, C)$ is minimal for all components C of ∂Y.

Given $Y \subseteq S$ a nonannular essential subsurface, the *subsurface projection*

$$\pi_Y : \mathcal{C}(S) \to \mathcal{C}(Y)$$

takes a vertex $a \in \mathcal{C}^{(0)}(S)$ to a subset of $\mathcal{C}(Y)$ as follows. First, put a and Y in minimal position. If $a \subseteq Y$ then $\pi_Y(a) = a \in \mathcal{C}(Y)$. If a intersects Y in a collection of arcs we associate to the collection of arcs a collection of essential simple closed curves as described below. Suppose α is an arc in Y. We associate to α a collection of pairwise disjoint simple closed curves in Y, denoted by $\overline{\alpha}$. Concretely, let $N(\alpha)$ denote a thickening of the union of α with the (at most two) components of ∂Y that it meets. We then define $\overline{\alpha}$ to be the essential components of $\partial N(\alpha)$. See Figure 6 on page 967 for an example where the arc meets two boundary components. Thus, if a intersects Y in a collection of arcs, we define $\pi_Y(a)$ to be the collection of essential simple closed curves $\bar{a}$ obtained as described above. Now, given $a, b \in \mathcal{C}^{(0)}(S)$ we define $d_Y(a,b)$ by

$$d_Y(a,b) = \operatorname{diam}_{\mathcal{C}(Y)}(\pi_Y(a), \pi_Y(b)).$$

Finally, we note that π_Y is coarsely Lipschitz (Lemma 2.3 of [19]):

$$d_Y(a,b) \leq 2 \cdot d_S(a,b) + 2.$$

The following is known as the bounded geodesic image theorem. It is due to Masur and Minsky in [19] and appears as Theorem 3.1 there.

Theorem 2.7 (bounded geodesic image theorem) *There exists a constant M depending only on $\chi(S)$ such that the following holds. Let $Y \subseteq S$ be an essential subsurface. If ξ is a geodesic in $\mathcal{C}(S)$ all of whose vertices represent curves which intersect Y nontrivially, then $\operatorname{diam}(\pi_Y(\xi)) \leq M$.*

Remark 2.8 We will use the contrapositive of the above: If the projected image of ξ to $\mathcal{C}(Y)$ is larger than M, then there is a vertex on the geodesic, ξ, that represents a curve disjoint from Y. In particular, said vertex has distance 1 to ∂Y.

2.3 The pants graph

We will use another graph associated to the curves on a surface called the pants graph. Before we define the graph we need the following definitions.

A *pair of pants* is a compact surface of genus 0 with three boundary components. Let S be a surface with $\chi(S) < 0$. A *pants decomposition* of S is a collection of disjoint simple closed curves on S such that the result of cutting S along these curves results in a disjoint union of pairs of pants. Two pants decompositions $P, P' \subseteq S$ are related by an *elementary move* if P' can be obtained from P by replacing a curve $\alpha \in P$ with a curve $\beta \neq \alpha$ such that $i(\alpha, \beta)$ is minimal over all choices of β (i.e., over all β such that replacing α with β results in a pants decomposition). We are now equipped to define the pants graph.

Definition 2.9 The *pants graph of* S, denoted by $\mathcal{P}(S)$, is the graph with vertices corresponding to isotopy classes of pants decompositions of S and whose edges connect pants decompositions which differ by an elementary move.

The graph, $\mathcal{P}(S)$, also carries a natural metric, $d_{\mathbf{P}}$, by assigning each edge length 1. Again, the mapping class groups naturally acts by isometries on the pants graph. Using this metric structure we can define the translation distance of a mapping class on the pants graph, $\tau_{\mathbf{P}}(f)$, in the same way as in the curve graph.

The following is an important theorem which relates the distance between pants decompositions in the curve graph of a subsurface to distance in the pants graph. It is commonly referred to as the Masur–Minsky distance formula for the pants graph, and appears originally as Theorem 6.12 of [19]. It appears, as written, as Theorem 4.4 of [4].

Theorem 2.10 *There is a constant $M_0 = M_0(S) > 0$ such that, given $M \geq M_0$, there are constants e_0, e_1, for which, if $P, P' \subseteq S$ are any two pants decompositions of S, then*

$$e_0^{-1} d_{\mathbf{P}}(P, P') - e_1 \leq \sum d_Y(P, P') \leq e_0 d_{\mathbf{P}}(P, P') + e_1,$$

where the sum is taken over all nonannular subsurfaces, $Y \subseteq S$, such that $d_Y(P, P') \geq M$.

The primary use we have for the pants graph is the close relation between the geometry of a mapping torus and the action of the monodromy on the pants graph. In particular, we make use of the following theorem of Brock (see [5]).

Theorem 2.11 *Given S there is a constant $K > 1$ so that if M_f is the mapping torus of a pseudo-Anosov, $f \in \mathrm{Mod}(S)$, then*

$$K^{-1} \tau_{\mathbf{P}}(f) \leq \mathrm{vol}(M_f) \leq K \tau_{\mathbf{P}}(f).$$

2.4 Geodesic laminations

We now introduce some of the theory of geodesic laminations that we will use later. For a thorough exposition of the theory, the reader should see [3; 6; 10].

Definition 2.12 Fix a surface, S, with $\chi(S) < 0$ and fix a hyperbolic metric on S. A *geodesic measured lamination*, L, on S is a nonempty collection of disjoint simple geodesics whose union is closed, along with a transverse measure that is invariant as you flow along the geodesics.

We denote the space of geodesic measured laminations on S by $\mathcal{ML}(S)$. We now also define the space of projective measured laminations, $\mathcal{PML}(S)$, as the space $\mathcal{ML}(S)$ modulo the scaling of measures. We now require the following definitions.

Definition 2.13 A geodesic lamination L is *minimal* if L does not contain any proper sublaminations. A lamination, L, is *filling* if every component of $S - L$ is either an ideal polygon or a once punctured ideal polygon.

We now give a second, equivalent, definition of what it means for a homeomorphism to be pseudo-Anosov.

Definition 2.14 A homeomorphism, f, of S is pseudo-Anosov if it preserves a pair of transverse measured laminations on S, one expanding and one contracting. We call these laminations the *unstable* and *stable laminations* of f respectively.

Recall that, at the beginning of this section, a pseudo-Anosov homeomorphism was defined in terms of stable and unstable foliations rather than the stable and unstable laminations as above. In fact, there is an exact correspondence between measured laminations and measured foliations. Further, the stable lamination of a pseudo-Anosov, f, corresponds to the stable foliation of f under this correspondence. For a precise treatment of the correspondence, see [17].

Masur and Minsky showed in [18] that the curve graph is Gromov hyperbolic. Hence, we may consider the boundary at infinity, $\partial_\infty \mathcal{C}(S)$. Klarreich showed in [16] that $\partial_\infty \mathcal{C}(S)$ is the space of minimal filling laminations, $\mathcal{EL}(S) \subseteq \mathcal{ML}(S)$. Moreover, a sequence of curves $\alpha_i \in \mathcal{C}(S)$ converges to a point $\alpha \in \partial_\infty \mathcal{C}(S)$ if and only if the sequence converges to $\alpha \in \mathcal{EL}(S)$ equipped with the subspace topology inherited from $\mathcal{PML}(S)$.

The geometric intersection number as defined above for arcs and curves extends uniquely to a continuous, homogeneous function $i : \mathcal{ML}(S) \times \mathcal{ML}(S) \to \mathbb{R}$ that we also call the geometric intersection number (see [22, Chapter 9]).

3 Branched cover construction

Let M be a closed, connected, oriented 3-manifold. In this section we begin our construction of the desired knots in M. We start by constructing a fibered 2-component link in M using branched covers. We need the following definition.

Definition 3.1 A branched covering $p : M \to N$ of degree d is called *simple* if each point in N has at least $d - 1$ preimages in M. In particular, points away from the branch locus in N will have d preimages while points on the branch locus in N will have $d - 1$ preimages in M.

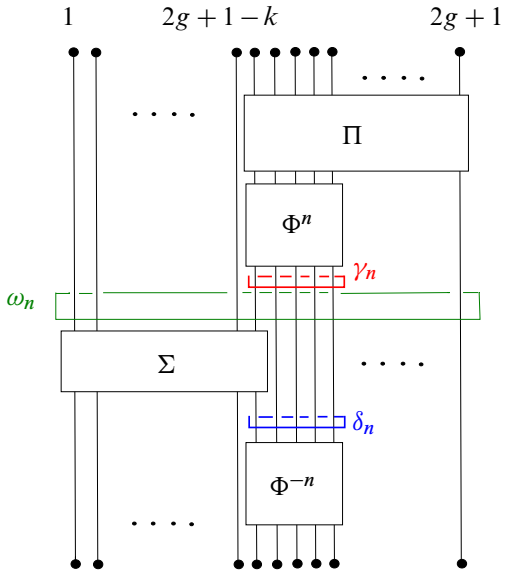


Figure 3: The braid whose closure will be the branching locus of our branched cover. The braid Π is the braid word representing the knot K which we branch over. The braid Φ is a pseudo-Anosov on its support. The word Σ totally consists of stabilizations. Note that γ_n and δ_n and the disks they bound do not change as n increases.

Due to a theorem of Hilden and Montesinos [14; 20] there is a 3-fold simple branched covering $p:M\to S^3$ branched along a knot K. Since the branched covers constructed are simple, points on the branch locus $K\subseteq S^3$ have two preimages in M. One preimage has branching index 1 while the other has branching index 2. By a theorem of Alexander [1] we can represent K as the closure of a braid $\Pi\in\mathfrak{B}_k$ for some k.

Fix $g>1$ such that $2g+1>k$. Let σ_i denote the positive half-twist between the i-th and $(i+1)$-st strands. We will consider $\Pi\in\mathfrak{B}_k$ under the natural inclusion into $\mathfrak{B}_{2g+1}$ where Π is on the last k strands (see Figure 3). In a variation on the construction of [2] we now define the following braids:

$$\begin{aligned}\Phi &= (\sigma_{2g+2-k})^{-1}(\sigma_{2g+3-k})(\sigma_{2g+4-k})^{-1}(\sigma_{2g+5-k}),\\ \Sigma &= (\sigma_{2g+1-k})(\sigma_{2g-k})^{-1}\cdots(\sigma_2)^{\pm}(\sigma_1)^{\mp},\\ \beta_n &= \Pi\Phi^n\Sigma\Phi^{-n}.\end{aligned}$$

Let $\widehat{\beta_n}$ be the braid closure of β_n. Let ω_n be its braid axis encircling the braid after the Φ^n factor and before the Σ factor of β_n. Let $\Lambda_n=\widehat{\beta_n}\cup\omega_n$. Note that ω_n bounds a disk Ω_n in S^3 meeting $\widehat{\beta_n}$ in $2g+1$ points. Hence, $S^3-\Lambda_n$ is a punctured disk bundle over the circle with monodromy β_n.

Lemma 3.2 *The braid closure $\widehat{\beta_n}$ is the knot K.*

Proof Note that $\widehat{\Pi}$ is K, and that $\widehat{\beta_n}$ is stabilized along the first $2g+1-k$ strands. Destabilizing smooths the crossings coming from Σ and deletes the first $2g+1-k$ strands. Now the Φ^n and Φ^{-n} factors cancel. We are left with the k-strand braid closure $\widehat{\Pi}$ which by definition is K. □

We will now introduce some notation. Let γ_n and δ_n be the simple closed curves encircling the $2g+2-k$, $2g+3-k$, $2g+4-k$, $2g+5-k$, and $2g+6-k$ strands of β_n taken immediately before and after the Σ factor of β_n (see Figure 3). Let Γ_n and Δ_n denote the disks bounded by γ_n and δ_n respectively. Note that Γ_n and Δ_n each intersect β_n in 5 points. Let w_n and W_n denote $p^{-1}(\omega_n)$ and $p^{-1}(\Omega_n)$ respectively.

Lemma 3.3 *The manifold $N_n = M - w_n$ is an $S_{g-1,2}$-bundle over the circle.*

Proof That N_n is fibered over the circle follows from the fact that $S^3 - \Lambda_n$ is a punctured disk bundle over the circle. Since p is a simple, 3-fold cover and Ω_n intersects β_n in $2g+1$ points, an Euler characteristic calculation shows that W_n is homeomorphic to $S_{g-1,2}$. Indeed, p restricts to a simple 3-fold branched cover of the disk, Ω_n, branched over the $2g+1$ points of intersection with β_n. By the definition of a 3-fold simple branched cover, each of the $2g+1$ branch points in Ω_n have two preimages. One of the preimages has branching index one while the other has branching index two. This is sufficient to compute the Euler characteristic of the fiber surface of N_n. First, we apply the Riemann–Hurwitz formula

$$\chi(W_n) = 3 \cdot \chi(\Omega_n) - \sum_{p \in W_n} (e_p - 1),$$

where e_p denotes the branching index at p. As discussed above, we know the branching indices of all points in W_n, so

$$\chi(W_n) = 3 \cdot 1 - (2g+1).$$

Hence, $\chi(W_n) = 2-2g$. The boundary of Ω_n is triply covered by the boundary of W_n. Therefore, ∂W_n has 1, 2, or 3 components. By the above calculation $\chi(W_n)$ is even. Thus, ∂W_n has 2 components. We also note here that by the definition a simple branched cover, W_n is connected. Indeed, if W_n were not connected, then it must be the disjoint union of two or three homeomorphic surfaces. By the above, W_n has two boundary components, so W_n must be the disjoint union of two surfaces, W_1 and W_2, each with one boundary component. However, the branched covering map, $p : W_n \to \Omega_n$, restricts to a branched cover of Ω_n on each of W_1 and W_2. In particular, we may assume without loss of generality that $p|_{W_1}$ is a one-sheeted cover and $p|_{W_2}$ is a two-sheeted cover. In this case, $p|_{W_1}$ is a homeomorphism, and thus $\Omega_n \cong W_1 \cong W_2$. This contradicts the above Euler characteristic calculation for W_n since $g > 1$. Combining the above, we see that W_n is homeomorphic to, $S_{g-1,2}$, a surface with genus $g-1$ and 2 boundary components. □

4 Hyperbolicity and volume

To prove the main theorem we need to construct from N_n the desired knot complements, verify their hyperbolicity, and show that their volumes grow linearly with n. In service of this we will change our perspective to a 2-dimensional perspective. We will focus on the fiber surface $W_n \subseteq N_n$ which is fixed as the monodromy of the fibration varies with n.

In particular we specify a marking of the disk $\Omega = \Omega_n \subseteq S^3$. Here, by a *marking* we mean an isotopy class of a diffeomorphism from the disk $\Omega_n \subseteq S^3$ to a fixed disk with $2g+1$ marked points. We let

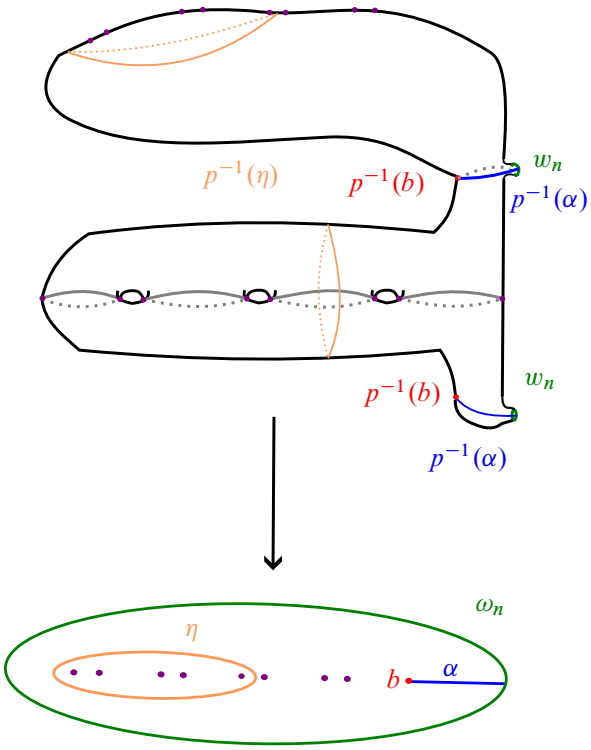


Figure 4: The unique simple cover from $S_{3,2}$ to a disk. The gray curves divide the covering surface into three "fundamental domains" for the cover. See the papers by Winarski and Fuller ([23, Section 3.2] and [11]) for more on a similar simple branched cover.

$S^3 = \mathbb{R}^3 \cup \infty$. Let $\Omega \subseteq \mathbb{R}^2 \times 0$ be the unit disk in the plane with marked points $\{1, \dots, 2g+1\}$ where the marked point i is at coordinate $\left(\frac{-1}{2} + \frac{i-1}{2g}, 0, 0\right)$. Using this marking, we identify Γ_n and Δ_n with subsurfaces of Ω_n by flowing them along the braid until they meet Ω_n. Moreover, by fixing a simple 3-fold branched covering of the disk, we fix a marking of the fiber surface, $W = W_n$, as shown in Figure 4, by composing the marking of the disk with the branched covering. Again, by working in the marked surface we view the subsurfaces $p^{-1}(\Gamma_n)$ and $p^{-1}(\Delta_n)$ as subsurfaces of $W = p^{-1}(\Omega)$.

From our 2-dimensional perspective we observe that γ_n, δ_n, Γ_n, and Δ_n do not depend on n, so from now on we will set $\gamma = \gamma_n$, $\delta = \delta_n$, $\Gamma = \Gamma_n$, and $\Delta = \Delta_n$. Let $b_n = P \circ F^n \circ S \circ F^{-n}$ be the monodromy of N_n, where P, F, and S are the lifts of Π, Φ, and Σ respectively.

Let c, d, and w_n denote $p^{-1}(\gamma)$, $p^{-1}(\delta)$, and $p^{-1}(\omega_n)$ respectively. Also, let C^+ and D^+ denote $p^{-1}(\Gamma)$ and $p^{-1}(\Delta)$ respectively.

Lemma 4.1 *C^+ and D^+ are each homeomorphic to the disjoint union of a disk and $S_{2,1}$. Furthermore, $C^+ \cap D^+$ is the disjoint union of a $S_{1,2}$ and a disk.*

Proof If we restrict p to a single fiber of N_n we get a simple 3-fold branched cover $p' : S_{g-1,2} \to D^2$ branched over $2g+1$ points. It follows from the work of Gabai and Kazez in [12] that 3-fold simple

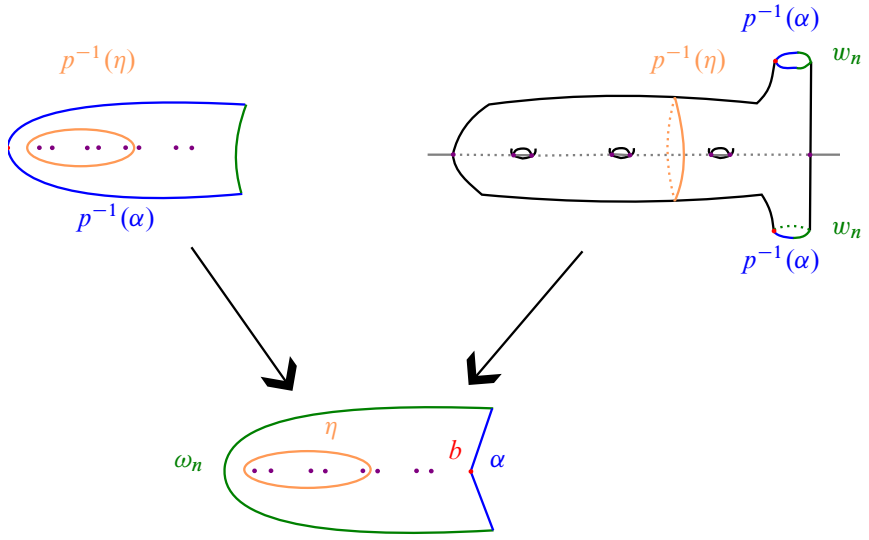


Figure 5: After cutting the disk $\Omega = \Omega_n$ along α and the covering surface along $p^{-1}(\alpha)$ we obtain two components of the covering surface. One component covers the cut disk trivially. The other component covers the cut disk by an order 2 involution that identifies the two boundary components.

branched covers of the disk are unique up to equivalence. By uniqueness of the cover we may assume it takes on the configuration shown in Figure 4 in the disk fiber of S^3 immediately following Π. That is, the distinguished branch point labeled b may be taken to lie on the $(2g+1)$-st strand of the braid in the fiber Ω_n. In particular, every simple closed curve that bounds a disk containing any 5 of the branch points that also does not intersect the arc α can be taken by a homeomorphism of the disk to the simple closed curve η as shown in Figure 4. Observe that c and d will consist of two simple closed curves since γ and δ each bound a disk with an odd number of branch points. By construction, γ and δ do not intersect the arc α, and hence one of the two preimages of γ and of δ is completely contained in the upper fundamental domain of the cover and bounds a disk as in the left panel of Figure 5. For example, every simple closed curve that bounds a disk containing five branch points other than the distinguished branch point, b, in Figure 4 is homeomorphic to the curve η by the change of coordinates principle. Again by an Euler characteristic calculation we conclude that the other component is homeomorphic to $S_{2,1}$, a surface of genus two with one boundary component. Note that $\Gamma \cap \Delta$ may be taken to be a disk that intersects β_n in 4 points. Thus, $C^+ \cap D^+$ consists of 2 disconnected components. One is the intersection of two disks in the upper fundamental domain that must be connected by the properties of the cover and hence is a disk. The other component is the double branched cover of a disk, branched over 4 points all of which have branching index 2. By another Euler characteristic calculation this is a surface of genus 1 with 2 boundary components. □

Observe that b_n factors as $(P \circ S) \circ (S^{-1} \circ F^n \circ S) \circ F^{-n}$. Let C and D denote the $S_{2,1}$ component of C^+ and D^+ respectively. To improve readability, we will frequently omit the $\circ$ symbol. We however

maintain the same convention that mapping classes are multiplied according to functional notation. That is, PS refers to applying first S then P.

Lemma 4.2 *F is a partial pseudo-Anosov with $\operatorname{supp}(F) = D$ and $\operatorname{supp}(S^{-1}FS) = C$.*

Proof Recall that $p^{-1}(\operatorname{supp}(\Phi)) = p^{-1}(\Gamma) = C^+$. By Lemma 4.1 C^+ is the disjoint union of C and a disk. Moreover, the component of C^+ that contains branching index one points is the disk component again by Lemma 4.1 (see also Figure 5). We observe that $C^+ - C$ and $D^+ - D$ are both disks and hence any mapping class restricted to either one is trivial. Note that p restricted to the subsurface C that contains only index two points is a double branched cover of Γ. Since Φ is a pseudo-Anosov on Γ it preserves two transverse singular foliations on Γ. The foliations lift through the branched cover to give singular transverse foliations on the surface C. Any possible 1-prong singularities at punctures of Ω_n are double covered in C and D and hence become 2-prong singularities. Therefore, F is a partial pseudo-Anosov with $\operatorname{supp}(F) = D$ and $\operatorname{supp}(S^{-1}FS) = C$. See [9, Section 14.1] for more on pseudo-Anosov mapping classes arising from branched covers. □

Lemma 4.3 *b_n has linearly growing translation distance in the curve complexes of C and D. In particular there exists some $n_0 > 0$ such that b_n is not periodic for all $n \geq n_0$.*

Proof By Lemma 4.2, $S^{-1}F^nS$ and F^{-n} are each pseudo-Anosov on their support, namely C and D. By Lemma 4.1 $C \cap D$ is homeomorphic to a surface with genus 1 and 2 boundary components. Hence, Theorem 5.2 of [7] implies that $(S^{-1}F^nS) \cdot F^{-n}$ has linearly growing translation distance in the curve complexes of each of C and D. Therefore, there exists some n_0 such that for all $n \geq n_0$, b_n has large translation distance in each of C and D, and hence is infinite order. □

While we have shown that our monodromies are eventually not periodic, we need monodromies that are eventually pseudo-Anosov. Since, this may not be true for b_n, we need to modify our manifolds N_n slightly. To that end we introduce a more 2-dimensional way of discussing fibered links in closed oriented 3-manifolds. We will use the equivalent notion of an *open book decomposition* of our closed, oriented 3-manifold M. An open book decomposition of M is a pair (S, φ) where S is the fiber surface ($\partial S \neq \varnothing$) and φ is the monodromy. The mapping torus M_φ of φ is homeomorphic to a link complement in M. Then M is homeomorphic to the quotient of M_φ under the identification $(x, t) \sim (x, t')$ where $x \in \partial S$ and $t, t' \in [0, 1]$. The quotient of ∂M_φ under the above identification is the fibered link in M which is called the *binding* of (S, φ). We now describe an operation on an open book decomposition (S, φ) of M called a stabilization.

Definition 4.4 Let (S, φ) be an open book decomposition of M, and let $(\gamma, \partial\gamma) \subseteq (S, \partial S)$ be a properly embedded arc. A *stabilization* along γ of the open book decomposition (S, φ) is the open book decomposition (S', φ') of M. where S' is obtained by attaching a 1-handle to S that connects the endpoints of γ on ∂S and $\varphi' = T_{\gamma'} \circ \varphi$ where γ' is the curve that agrees with γ in S and intersects the cocore of the 1-handle once (see Figure 6) and $T_{\gamma'}$ denotes the Dehn twist along γ'.

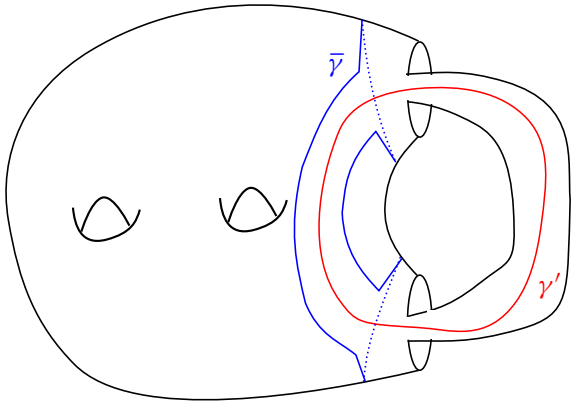


Figure 6: The surface obtained by stabilizing $S_{2,2}$. The curve γ' is the curve we twist along while stabilizing. The curve $\overline{\gamma}$ is the result of the "closing up" of the arc γ which agrees with γ' on $S_{2,2}$.

The following proposition is essentially Theorem 1.1 of [8]. In the same way as described in Section 2.2, we associate to any arc, γ, a collection of mutually disjoint simple closed curves, $\overline{\gamma}$.

Proposition 4.5 *Let $(S_{g-1,2}, \varphi)$ be an open book decomposition of a closed, oriented 3-manifold M such that φ is not periodic. Let γ be an arc connecting the two boundary components of $S_{g-1,2}$. If $d_{S_{g-1,2}}(\overline{\gamma}, \varphi(\overline{\gamma})) = N > 16$ then stabilization along the arc γ yields an open book decomposition $(S_{g,1}, T_{\gamma'} \circ \varphi)$ such that $T_{\gamma'} \circ \varphi$ is pseudo-Anosov. Recall that γ' is the extension of γ to the stabilized surface.*

Proof While Theorem 1.1 of [8] contains the hypothesis that the monodromy φ is right-veering this is not needed to show that $T_{\gamma'} \circ \varphi$ is pseudo-Anosov. All that is required is that φ is not periodic. While the lemma is essentially Theorem 1.1 of [8], the required distance between $\overline{\gamma}$ and its image under φ is not made explicit there. For the sake of completeness, we sketch their proof and make the distance explicit. Throughout this proof, $d(\cdot, \cdot)$ will refer to $d_{S_{g-1,2}}(\cdot, \cdot)$.

Let $\varphi' = T_{\gamma'} \circ \varphi$. We argue that φ' is pseudo-Anosov by showing that it is neither reducible nor periodic. We will first argue that $\varphi'(\delta) \neq \delta$ for any multicurve $\delta \subseteq S_{g,1}$. First suppose that $\delta \subseteq S_{g-1,2} \subseteq S_{g,1}$. If $i(\varphi(\delta), \gamma') = n$ then there is a representative, g of $\varphi'(\delta)$ that intersects the cocore of the 1-handle, a, n times. If $i(\varphi'(\delta), a) < n$ then there is a bigon consisting of a subarc of a and an arc of g. Then one checks that this implies there is a bigon consisting of an arc of γ and an arc of $\varphi(\delta)$. See the third paragraph of the proof of case 1 of Theorem 1.1 of [8] for this argument. This is a contradiction if $n > 0$. If $n = 0$ then $i(\varphi(\beta), \overline{\gamma}) = 0$ for any component β of δ. Hence, $d(\varphi(\beta), \overline{\gamma}) = 1$ for all such β. However, $d(\overline{\gamma}, \varphi(\overline{\gamma})) = N$ and therefore, $d(\varphi(\overline{\gamma}), \varphi(\beta)) \geq N - 1$ for every β. Since $n = 0$, we see that $i(\varphi'(\delta), \overline{\gamma}) = 0$. However, $d(\overline{\gamma}, \beta) = d(\varphi(\overline{\gamma}), \varphi(\beta)) > N - 1$, and hence $i(\overline{\gamma}, \beta) > 2^{((N-1)-2)/2}$. This yields contradiction so long as $N > 2$.

Now suppose that $\delta \not\subseteq S_{g-1,2}$. Let $i(\delta, a) = k > 0$. Let $B = S_{g,1} - S_{g-1,2}$ denote the stabilization band. Normalize δ so that it intersects γ' and a transversely and efficiently. We then subdivide δ into arcs

$\delta_1, \dots, \delta_k, \delta_1', \dots, \delta_k'$ so that $\delta_i \subseteq S_{g-1,2}$ and $\delta_i' \subseteq B$. The δ_i' are all linear arcs in B. If $i(\delta_i', \gamma') = 0$ we call such an arc *vertical*. If $i(\delta_i', \gamma') > 0$ we say it has *positive slope* or *negative slope*. See Figure 7. We normalize and subdivide $\varphi(\delta)$ in the same way. Up to isotopy we may assume that there is no triangle contained in $S_{g-1,2}$ with boundary an arc of γ', an arc of δ, and an arc of $\partial S_{g-1,2}$. Now let m denote the number of arcs of $\varphi(\delta) \cap B$ that are not vertical. Let n denote the number of intersections between $\varphi(\delta)$ and γ' that are outside of B so that $i(\varphi(\delta), \gamma') = m + n$. Note that this definition of n agrees with the definition of n as $i(\varphi(\delta), \gamma')$ in the above paragraph when $\varphi(\delta)$ does not intersect the band B. Colin and Honda show that $i(\varphi'(\delta), a) = k \pm m + n$ where the sign depends on whether the nonvertical arcs have positive or negative slope. If $m \neq n$ then $i(\varphi'(\delta), a) \neq i(\delta, a) = k$, so $\varphi'(\delta) \neq \delta$. It remains to deal with the case that $m = n$.

Suppose that $m = n$. Then $i(\varphi'(\delta), \gamma') = m + n = 2m \leq 2k$ by the definitions of m and k. Suppose for the sake of contradiction that $\varphi'(\delta) = \delta$. Then $i(\delta, \gamma') = i(\varphi'(\delta), \gamma')$. Therefore $i(\delta, \gamma') \leq 2k$, and hence there is some $1 \leq i \leq k$ such that $i(\delta_i, \gamma) \leq 2k/k = 2$. Now we note that

$$(*) \qquad i(\overline{\delta_i}, \overline{\gamma}) \leq 2 + 4 = 6.$$

By Lemma 2.6

$$d(\overline{\delta_i}, \overline{\gamma}) \leq \lfloor 2\log_2(6) + 2 \rfloor = 7.$$

Observe that $\overline{\delta_i}$ and $\overline{\delta_j}$ cannot jointly fill the surface, so

$$d(\overline{\delta_i}, \overline{\delta_j}) \leq 2 \quad \text{for all } i, j.$$

Therefore,

$$d(\overline{\delta_j}, \overline{\gamma}) \leq 2 + 7 = 9 \quad \text{for all } j.$$

Applying φ yields

$$d(\varphi(\overline{\delta_j}), \varphi(\overline{\gamma})) \leq 9 \quad \text{for all } j.$$

Thus,

$$d(\varphi(\overline{\delta_j}), \overline{\gamma}) \geq N - 9.$$

Applying Hempel's lemma again,

$$i(\varphi(\overline{\delta_j}), \overline{\gamma}) \geq 2^{(N-9)/2-1}.$$

Passing from arcs to the corresponding closed curves we may increase intersection number by at most 4 (as noted in $(*)$, for example). Hence, when passing from closed curves back to arcs we may decrease intersection number by at most 4. Therefore,

$$i(\varphi(\delta_j), \gamma) \geq 2^{(N-9)/2-1} - 4.$$

Multiplying by k, we get

$$i(\varphi(\delta), \gamma') \geq (2^{(N-9)/2-1} - 4)k > 2k;$$

this is a contradiction so long as $N > 16$ which proves that φ' is not reducible.

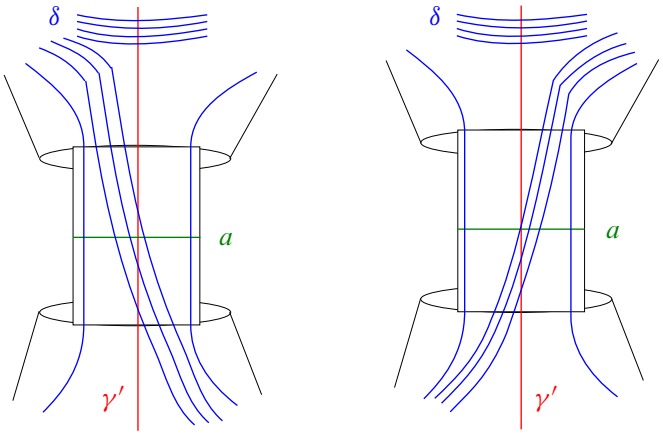


Figure 7: On the left is the case that δ has negative slope, and on the right is the case that δ has positive slope. In both cases $k = 5$, $m = 3$, and $n = 4$.

It just remains to show that φ' is not periodic. Colin and Honda show that φ' is not periodic by considering one of the connected components, η, of $\partial S_{g-1,2}$ as a curve in $S_{g,1}$. They argue that $i((\varphi')^n(\eta), a) \to \infty$ as $n \to \infty$. □

We now apply the above lemma to $M - w_n$ to obtain the following:

Proposition 4.6 *Let $(S_{g-1,2}, b_n)$ be the open book decomposition of M corresponding to the fibered link $w_n \subseteq M$, and let n_0 be the constant from Lemma 4.3. There exists an arc of stabilization γ and $n_1 \geq n_0 > 0$ such that $T_{\gamma'} \circ b_n$ is pseudo-Anosov for all $n \geq n_1$, and hence $M - \overline{w_n}$ is hyperbolic, where $\overline{w_n}$ is the image of w_n after stabilization.*

Proof First, note that by Lemma 4.3, for all $n \geq n_0$ the mapping class b_n is not periodic. We now explain how to choose an arc of stabilization γ such that $d_{S_{g-1,2}}(\overline{\gamma}, b_{n_0}(\overline{\gamma})) > 16$ following [8]. First let γ_0 be any properly embedded arc connecting the two boundary components of $S_{g-1,2}$. In the case that b_{n_0} is pseudo-Anosov, fix a pseudo-Anosov, $f \in \mathrm{Mod}(S_{g-1,2})$, that does not share a stable or unstable lamination with b_{n_0}. In the case that b_{n_0} is reducible, any pseudo-Anosov, f, will suffice. Let ν be the stable lamination of f. Note that this implies that $b_{n_0}(\nu) \neq \nu$. The sequence $f^i(\overline{\gamma_0}) \to \nu$ as $i \to \infty$. We claim that $d_{S_{g-1,2}}\big(f^i(\overline{\gamma_0}), b_{n_0}(f^i(\overline{\gamma_0}))\big) \to \infty$. If $d_{S_{g-1,2}}\big(f^i(\overline{\gamma_0}), b_{n_0}(f^i(\overline{\gamma_0}))\big)$ does not approach ∞, then up to passing to a subsequence $d_{S_{g-1,2}}\big(f^i(\overline{\gamma_0}), b_{n_0}(f^i(\overline{\gamma_0}))\big) = N$ for some constant N. Let $f^i(\overline{\gamma_0}) = v_{i,0}, v_{i,1}, \ldots, v_{i,N-1}, v_{i,N} = b_{n_0}(f^i(\overline{\gamma_0}))$ be a geodesic in the curve graph. Note that $v_{i,0}$ and $v_{i,1}$ are disjoint. Up to passing to a subsequence, $v_{i,1}$ converges to a lamination ν' such that $i(\nu, \nu') = 0$. Similarly, $v_{i,N-1}$ converges to a lamination λ such that $i(\lambda, b_{n_0}(\nu)) = 0$. Since ν is minimal and filling and $i(\nu, \nu') = 0$, we have that $\nu = \nu'$ and similarly $b_{n_0}(\nu) = \lambda$. Hence, we have $v_{i,1} \to \nu$ and $v_{i,N-1} \to b_{n_0}(\nu)$ as $i \to \infty$, but $d_{S_{g-1,2}}(v_{i,1}, v_{i,N-1}) < N$. After repeating this process finitely many times, we conclude that $b_{n_0}(\nu) = \nu$. This contradicts the choice of f. Therefore, there exists some $m > 0$ such that $d_{S_{g-1,2}}\big(f^m(\overline{\gamma_0}), b_{n_0}(f^m(\overline{\gamma_0}))\big) > 16$. Moreover, f is pseudo-Anosov and

hence the action of f on $\mathcal{C}(S_{g-1,2})$ has positive translation distance. In particular, the forward translates of a vertex in $\mathcal{C}(S_{g-1,2})$ are eventually arbitrarily far from any fixed vertex of $\mathcal{C}(S_{g-1,2})$. Therefore, there exists some $l_0 > 0$ so that for all $l \geq l_0$ $d_{S_{g-1,2}}(f^l(\overline{\gamma_0}), \partial C) > 18$. Let $k = \max\{m, l\}$. Let γ be the arc such that $f^k(\overline{\gamma_0}) = \overline{\gamma}$.

Note that although the above only shows how to choose an arc of stabilization for b_{n_0} we may choose such an arc of stabilization uniformly for all open book decompositions $(S_{g-1,2}, b_n)$ where n is large enough. To see why this is true consider the subsurface C. Observe that there exist constants $K, L > 0$ such that $K^{-1} d_C(\overline{\gamma}, S^{-1}F^nSF^{-n}(\overline{\gamma})) - L \leq d_C(\overline{\gamma}, b_n(\overline{\gamma})) \leq K d_C(\overline{\gamma}, S^{-1}F^nSF^{-n}(\overline{\gamma})) + L$ for all n. This is due to the factorization of b_n as $(PS)(S^{-1}F^nSF^{-n)}$ where the (PS) factor is fixed as $n \to \infty$. By Lemma 4.3 there exists $n_1 \geq n_0 > 0$ such that $d_C(\overline{\gamma}, S^{-1}F^nSF^{-n}(\overline{\gamma})) \gg 0$ for all $n \geq n_1$. Hence, by Theorem 2.7 (see also Remark 2.8) the geodesic in $\mathcal{C}(S_{g-1,2})$, between $\overline{\gamma}$ and $b_n(\overline{\gamma})$ must pass within distance 1 of ∂C for all $n \geq n_1$. Therefore

$$d_{S_{g-1,2}}(\overline{\gamma}, \partial C) - 1 \leq d_{S_{g-1,2}}(\overline{\gamma}, b_n(\overline{\gamma}))$$

Now note that, $d_{S_{g-1,2}}(\overline{\gamma}, \partial C)$ is unchanged as $n \to \infty$, and hence $d_{S_{g-1,2}}(\overline{\gamma}, b_n(\overline{\gamma})) > 16$ for all $n \geq n_1$. Thus, by Proposition 4.5 the family of open book decompositions $(S_{g,1}, T_{\gamma'} \circ b_n)$ have pseudo-Anosov monodromies for all $n \geq n_1$. □

Remark 4.7 Stabilization does not affect the conclusions of Lemmas 4.2 and 4.3. We consider the subsurfaces C and D included into the stabilized surface and the mapping class F as a mapping class on the stabilized surface. Then, $T_{\gamma'} \circ b_n$ has linearly growing translation distance in the curve graphs of C and D, and F is a partial pseudo-Anosov on $S_{g,1}$, the stabilized surface.

Proposition 4.8 *In fixed genus g, as n tends to infinity, the knot complements $M_n = M - \overline{w_n}$ are eventually hyperbolic, with volumes tending to infinity.*

Proof As is shown in Proposition 4.6 $b_n' = T_{\gamma'} \circ b_n$ are pseudo-Anosov for $n \geq n_1$. Hence, their mapping tori, M_n, are hyperbolic for $n \geq n_1$. As is noted in Remark 4.7 and Lemma 4.3, b_n' has linearly growing subsurface projections to C_n and D_n. Now, by applying Theorem 2.10 we see that b_n' has linearly growing translation distance in the pants graph of $S_{g,1}$. By Theorem 2.11, the mapping tori of b_n' have linearly growing volume. These mapping tori are precisely the fibered knot complements M_n. □

For fixed g the family $\overline{w_n}$ are the promised family of fibered knots in M satisfying the conclusion of Theorem 1.1.

Acknowledgements

The author thanks his advisor Dave Futer for pointing him towards this problem as well as his guidance. The author thanks the NSF for its support via grant DMS-1907708. The author also thanks the anonymous referee for comments which expanded and improved the exposition of this paper.

References

[1] **J W Alexander**, *A lemma on systems of knotted curves*, Proc. Natl. Acad. Sci. U.S.A. 9:3 (1923) 93–95

[2] **K L Baker**, **D Futer**, **J S Purcell**, **S Schleimer**, *Large volume fibred knots of fixed genus*, Math. Res. Lett. 31:5 (2024) 1305–1314 MR

[3] **F Bonahon**, *Geodesic laminations on surfaces*, from "Laminations and foliations in dynamics, geometry and topology" (Stony Brook, NY, 1998) (M Lyubich, J W Milnor, Y N Minsky, editors), Contemp. Math. 269, Amer. Math. Soc., Providence, RI (2001) 1–37 MR

[4] **J F Brock**, *The Weil–Petersson metric and volumes of 3-dimensional hyperbolic convex cores*, J. Amer. Math. Soc. 16:3 (2003) 495–535 MR

[5] **J F Brock**, *Weil–Petersson translation distance and volumes of mapping tori*, Comm. Anal. Geom. 11:5 (2003) 987–999 MR

[6] **A J Casson**, **S A Bleiler**, *Automorphisms of surfaces after Nielsen and Thurston*, London Mathematical Society Student Texts 9, Cambridge Univ. Press (1988) MR

[7] **M T Clay**, **C J Leininger**, **J Mangahas**, *The geometry of right-angled Artin subgroups of mapping class groups*, Groups Geom. Dyn. 6:2 (2012) 249–278 MR

[8] **V Colin**, **K Honda**, *Stabilizing the monodromy of an open book decomposition*, Geom. Dedicata 132 (2008) 95–103 MR

[9] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Mathematical Series 49, Princeton Univ. Press (2012) MR

[10] **A Fathi**, **F Laudenbach**, **V Poénaru**, *Thurston's work on surfaces*, Mathematical Notes 48, Princeton Univ. Press (2012) MR

[11] **T Fuller**, *On fiber-preserving isotopies of surface homeomorphisms*, Proc. Amer. Math. Soc. 129:4 (2001) 1247–1254 MR

[12] **D Gabai**, **W H Kazez**, *The classification of maps of surfaces*, Invent. Math. 90:2 (1987) 219–242 MR

[13] **J Hempel**, *3-manifolds as viewed from the curve complex*, Topology 40:3 (2001) 631–657 MR

[14] **H M Hilden**, *Three-fold branched coverings of S^3*, Amer. J. Math. 98:4 (1976) 989–997 MR

[15] **S Hirose**, **E Kalfagianni**, **E Kin**, *Volumes of fibered 2-fold branched covers of 3-manifolds*, J. Topol. Anal. 17:2 (2025) 409–425 MR

[16] **E Klarreich**, *The boundary at infinity of the curve complex and the relative Teichmüller space*, Groups Geom. Dyn. 16:2 (2022) 705–723 MR

[17] **G Levitt**, *Foliations and laminations on hyperbolic surfaces*, Topology 22:2 (1983) 119–135 MR

[18] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves, I: Hyperbolicity*, Invent. Math. 138:1 (1999) 103–149 MR

[19] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves, II: Hierarchical structure*, Geom. Funct. Anal. 10:4 (2000) 902–974 MR

[20] **J M Montesinos**, *Three-manifolds as 3-fold branched covers of S^3*, Quart. J. Math. Oxford Ser. (2) 27:105 (1976) 85–94 MR

[21] **R Myers**, *Open book decompositions of 3-manifolds*, Proc. Amer. Math. Soc. 72:2 (1978) 397–402 MR

[22] **W P Thurston**, *The geometry and topology of three-manifolds*, lecture notes, Princeton University (1979) Available at https://url.msp.org/gt3m

[23] **R R Winarski**, *Symmetry, isotopy, and irregular covers*, Geom. Dedicata 177 (2015) 213–227 MR

J ROBERT OAKLEY roboakley01@gmail.com
Wachman Hall Mathematics Department, Temple University, Philadelphia, PA, United States

Received: November 13, 2023 Revised: October 29, 2024

msp

Algebraic & Geometric Topology **26**:3 (2026) 973–988
DOI: 10.2140/agt.2026.26.973
Published: April 1, 2026

The primitive curve complex for a handlebody

SANGBUM CHO AND JUNG HOON LEE

A simple closed curve in the boundary surface of a handlebody is called primitive if there exists an essential disk in the handlebody whose boundary circle intersects the curve transversely in a single point. The primitive curve complex is then defined to be the full subcomplex of the curve complex for the boundary surface, spanned by the vertices of primitive curves. Given any two primitive curves, we construct a sequence of primitive curves from one to the other one satisfying a certain property. As a consequence, we prove that the primitive curve complex for the handlebody is connected.

1 Introduction

Let V be a genus-g handlebody for $g \geq 2$, and let Σ be the boundary of V, a closed orientable surface of the same genus. A simple closed curve C in $\Sigma = \partial V$ is called a *primitive curve* if there exists an essential disk D in V such that C intersects ∂D transversely in a single point. Such a disk D is called a *dual disk* of C. Of course any primitive curve in Σ admits infinitely many nonisotopic dual disks, and conversely any nonseparating essential disk in V can be a dual disk of infinitely many nonisotopic primitive curves.

Two primitive curves C and C' are said to be *separated* if there exist dual disks D and D' of C and C', respectively, such that $C \cup D$ and $C' \cup D'$ are disjoint. If two primitive curves C and C' are separated with their dual disks D and D', then one can find quickly their common dual disk by taking a "band sum" of D and D'. On the other hand, there are infinitely many disjoint primitive curves C and C' that are not separated although they have a common dual disk (even though $C \cup C'$ is nonseparating in Σ). The main result of this work is stated as follows.

Theorem 1.1 *Let C and C' be primitive curves in Σ, the boundary of a genus-g handlebody V for $g \geq 2$. Then there exists a sequence $C = C_1, C_2, \ldots, C_n = C'$ of primitive curves in Σ such that C_i and C_{i+1} are separated for each $i \in \{1, 2, \ldots, n-1\}$.*

We first prove a weaker version of Theorem 1.1, in which C_i and C_{i+1} are only required to be disjoint and to have a common dual disk for each $i \in \{1, 2, \ldots, n-1\}$ (Theorem 2.4), and then Theorem 1.1 will be proved for the case of $g = 2$ (Theorem 3.1) and for the case of $g \geq 3$ (Theorem 4.4). In any case, the dual disks of C and C' can be chosen arbitrarily in the first line of the proof.

The curve complex $\mathcal{C}(\Sigma)$ for a closed orientable surface Σ of genus g with $g \geq 2$ is a simplicial complex defined as follows. The vertices of $\mathcal{C}(\Sigma)$ are the isotopy classes of essential simple closed curves in Σ, and a collection of $k+1$ distinct vertices of $\mathcal{C}(\Sigma)$ spans a k-simplex if it admits a collection of representatives, all of which are pairwise disjoint. The combinatorial structure of $\mathcal{C}(\Sigma)$ has been

MSC2020: 57K30.

widely studied for the past decades. In particular, it is well known that $\mathcal{C}(\Sigma)$ is a $(3g-4)$-dimensional connected complex, and in fact it is homotopy equivalent to a wedge sum of spheres [1]. In this work, we are interested in a special subcomplex of $\mathcal{C}(\Sigma)$ when Σ bounds a handlebody V, called the primitive curve complex.

The *primitive curve complex* $\mathcal{PC}(V)$ for V is the full subcomplex of $\mathcal{C}(\Sigma)$ spanned by vertices whose representatives are primitive curves in Σ. The following is a direct consequence of Theorem 1.1 (it is enough to see that C_i and C_{i+1} are disjoint in the theorem).

Corollary 1.2 *The primitive curve complex $\mathcal{PC}(V)$ for a genus-g handlebody V is connected for every $g \geq 2$.*

Understanding the combinatorial structure of a simplicial complex like $\mathcal{PC}(V)$ is not only an interesting problem itself, but it also has many useful applications. For example, Wajnryb [2] showed that "the cut-system complex" for a handlebody is simply connected, and found a finite presentation of the handlebody group by investigating the action of the group on the complex. The handlebody group also acts on the primitive curve complex $\mathcal{PC}(V)$ in a natural way, and so $\mathcal{PC}(V)$ could be an alternate to find a (simpler) presentation of the group.

We are also interested in a subcomplex of $\mathcal{PC}(V)$, called the primitive disk complex, when the handlebody V is standardly embedded in the 3-sphere. That is, the closure W of the complement of V in the 3-sphere is also a handlebody, which forms a Heegaard splitting of the 3-sphere together with the handlebody V. The *primitive disk complex* $\mathcal{P}(W)$ is then defined to be the full subcomplex of $\mathcal{PC}(V)$ spanned by the vertices of $\mathcal{PC}(V)$ whose representatives bound disks in the handlebody W. The vertices of $\mathcal{P}(W)$ can be considered as the isotopy classes of the disks themselves, which we call primitive disks in W. If $\mathcal{P}(W)$ is a connected complex and moreover if Theorem 1.1 is true for primitive disks in W, i.e., if any two primitive disks in W can be joined by a sequence of primitive disks in W in which any two consecutive disks in the sequence are separated, then one can show quickly that the reducing sphere complex for the splitting is also connected, which implies that the Powell conjecture is true [3].

Throughout the paper, $\overline{X}$ will denote the closure of X and $N(X)$ a regular neighborhood of X for a subspace X of a space, where the ambient space will always be clear from the context.

2 Primitive curves having a common dual disk

The goal of this section is to prove Theorem 2.4, a weaker version of our main theorem. To avoid repeating the expression of "sequence of curves" in the arguments, we simply say that two primitive curves C and C' *are c-connected* if there exists a sequence $C = C_1, C_2, \dots, C_k = C'$ of primitive curves for some k such that C_i and C_{i+1} are disjoint and have a common dual disk for each $i \in \{1, 2, \dots, k-1\}$. We define the *distance* $d(C, C')$ between C and C', concerning the c-connectedness, to be the minimum number $k-1$ over all such sequences $C = C_1, C_2, \dots, C_k = C'$. If all of $C = C_1, C_2, \dots, C_k = C'$ admit a single common dual disk D, then we say that C and C' *are c-connected with the common dual disk* D.

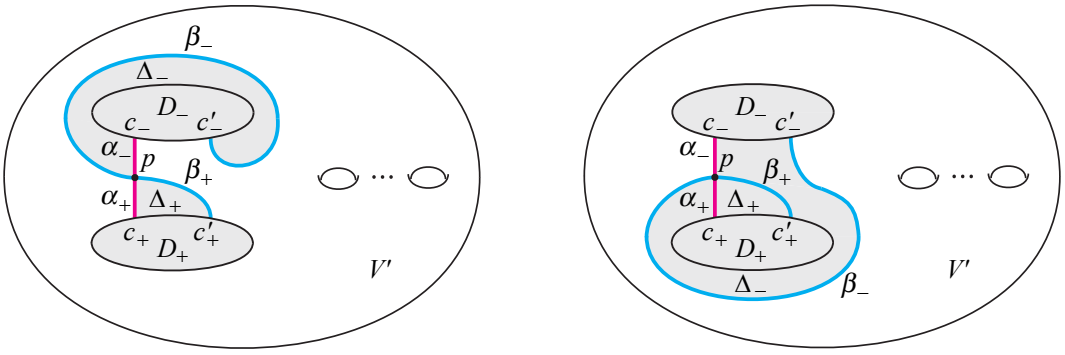


Figure 1: (a) $\Delta_+ \cap \Delta_- = p$ and (b) $\Delta_+ \subset \Delta_-$.

Lemma 2.1 (intersecting in a single point) *Let C and C' be primitive curves in Σ, the boundary of a genus-g handlebody V for $g \geq 2$. Suppose that C and C' have a common dual disk D and that $|C \cap C'| = 1$. Then C and C' are c-connected with the common dual disk D, and $d(C, C') \leq 5$.*

Proof Cutting V along D, we have a genus-$(g-1)$ handlebody V' with two copies D_+ and D_- of D on $\Sigma' = \partial V'$. The primitive curve C is then cut into an arc α with one of its endpoints c_+ in ∂D_+ and the other endpoint c_- in ∂D_-. The primitive curve C' is also cut into an arc β with its endpoints c'_+ and c'_- in ∂D_+ and ∂D_-, respectively. Let $p = \alpha \cap \beta$. Let α_+ and α_- be the subarcs of α with endpoints $\{c_+, p\}$ and $\{p, c_-\}$, respectively. Let β_+ and β_- be the subarcs of β with endpoints $\{c'_+, p\}$ and $\{p, c'_-\}$, respectively. First we consider the following special case.

Case 1 (both $\alpha_+ \cup \beta_+$ and $\alpha_- \cup \beta_-$ cut off disks Δ_+ and Δ_- from $\overline{\Sigma' - D_+}$ and $\overline{\Sigma' - D_-}$, respectively) There are two subcases (see Figure 1(a) and (b)).

(a) The two disks Δ_+ and Δ_- intersect only in the point p.

(b) One of Δ_+ and Δ_-, say Δ_-, contains Δ_+.

(a) Take a point d_+ in $\partial D_+ \cap \Delta_+$ between c_+ and c'_+. Let d_- be the point in ∂D_- that is identified with d_+. Then d_- is not in Δ_-. Take a point q in α_+ between c_+ and p. Let δ_+ be an arc in Δ_+ connecting d_+ and q. Since $D_+ \cup D_- \cup \Delta_+ \cup \Delta_-$ is homotopy equivalent to a point, we can take a nonseparating arc δ_- in $\overline{\Sigma' - (D_+ \cup D_- \cup \Delta_+ \cup \Delta_-)}$ connecting q and d_- (see Figure 2).

Let $\delta = \delta_+ \cup \delta_-$. If we identify D_+ and D_-, then δ becomes a primitive curve with a dual disk D and it is disjoint from C'. The arc δ intersects α in a single point q. Let Δ be the triangular disk determined by $\{q, c_+, d_+\}$. Since δ_- is nonseparating in $\overline{\Sigma' - (D_+ \cup \Delta \cup N(\alpha) \cup D_-)}$, we can take an arc ϵ in $\overline{\Sigma' - (D_+ \cup \Delta \cup N(\alpha) \cup D_-)}$ with endpoints e_+ and e_- in ∂D_+ and ∂D_-, respectively, with the following properties (see Figure 3).

- The arc ϵ is disjoint from α and δ.
- The two points e_+ and e_- become the same point if we identify D_+ and D_-.

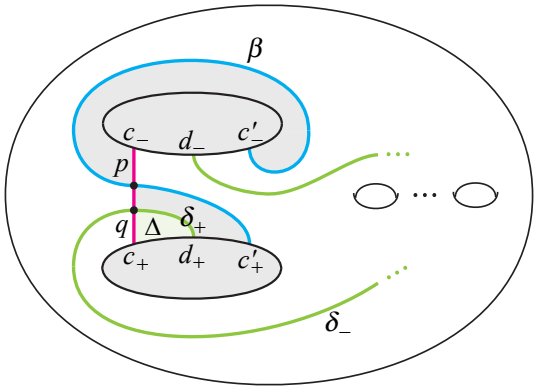


Figure 2: The arc $\delta = \delta_+ \cup \delta_-$.

If we identify D_+ and D_-, then the arc ϵ becomes a primitive curve with the dual disk D, and it is disjoint from C and δ. Hence we see that C and C' are c-connected with the common dual disk D via the curves ϵ and δ, and $d(C, C') \leq 3$ in this case.

(b) We push β_+ along Δ_+ to the opposite side of α_+. Then c'_+ is moved to the opposite side of c_+. Of course, c'_- is also moved to the opposite side of c_-. The point p and the disk Δ_+ disappear, but instead a new intersection point p' of $\alpha \cap \beta$ near c_- and a small triangular disk Δ'_- determined by $\{p', c_-, c'_-\}$ are created. The disk Δ_- disappears, but instead a disk Δ'_+ determined by $\{p', c_+, c'_+\}$ intersecting Δ'_- only in p' is created. Now the situation is like the case (a) in the above, and hence C and C' are c-connected with a common dual disk D, and we have $d(C, C') \leq 3$.

Case 2 (at least one of $\alpha_+ \cup \beta_+$ and $\alpha_- \cup \beta_-$, say $\alpha_+ \cup \beta_+$, does not cut off a disk from $\overline{\Sigma' - D_+}$) Without loss of generality, assume that β_+ is incident to p in the right side of α, and then take a point d_+ in ∂D_+ near c_+ and in the right side of c_+ as in Figure 4. Take a point q in α_+ near c_+. Let δ_+ be a short arc connecting d_+ and q such that the interior of δ_+ is disjoint from $D_+ \cup D_- \cup \alpha \cup \beta$. Let $\delta_- =$ (the subarc of α_+ from q to p) $\cup\, \beta_-$. Slide the endpoint of δ_- in ∂D_- to the point d_- that is

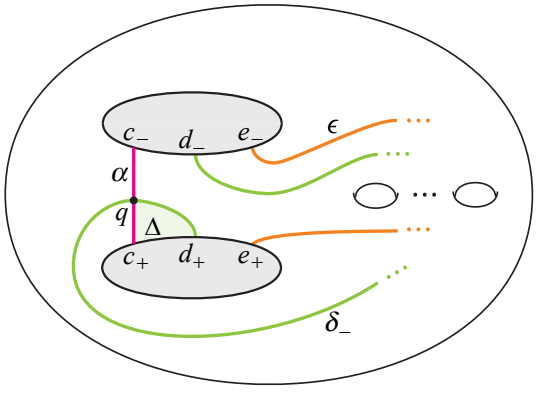


Figure 3: The arc ϵ.

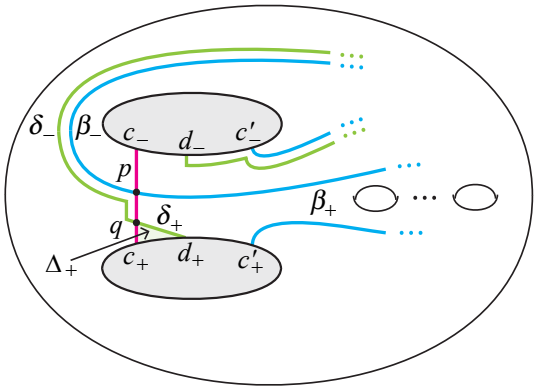


Figure 4: The arc $\delta = \delta_+ \cup \delta_-$.

identified with d_+ in Σ. Then isotope δ_- slightly so that $\delta = \delta_+ \cup \delta_-$ intersects α in a single point q, and δ is disjoint from β (see Figure 4).

Let α'_+ be the subarc of α_+ from c_+ to q, and α'_- be the subarc of α from q to c_-. The union $\alpha'_+ \cup \delta_+$ cuts off a small disk Δ_+ from $\overline{\Sigma' - D_+}$. If $\alpha'_- \cup \delta_-$ also cuts off a disk from $\overline{\Sigma' - D_-}$, then we apply case 1 to α and δ. Since the distance between curves α and δ in V is less than or equal to 3, and δ is disjoint from β, the curves C and C' are c-connected with a common dual disk D, and we have $d(C, C') \leq 4$ in this case. If $\alpha'_- \cup \delta_-$ does not cut off a disk from $\overline{\Sigma' - D_-}$, then we apply case 2 again to α and δ with respect to $\alpha'_- \cup \delta_-$. More precisely, we take an arc ϵ such that ϵ is disjoint from δ, and α and ϵ give rise to two disks intersecting in a single point as in case 1(a) (see Figure 5). Since the distance between curves α and ϵ in V is less than or equal to 3, and since δ is disjoint from ϵ and β, the curves C and C' are c-connected with a common dual disk D, and we have $d(C, C') \leq 5$ in this case. □

Lemma 2.2 (a stronger version of Lemma 2.1) *Let C and C' be primitive curves in Σ, the boundary of a genus-g handlebody V for $g \geq 2$. Suppose that C and C' have a common dual disk D. Then C and C' are c-connected with the common dual disk D, and $d(C, C') \leq 5\,|C \cap C'|$ if $C \cap C' \neq \varnothing$.*

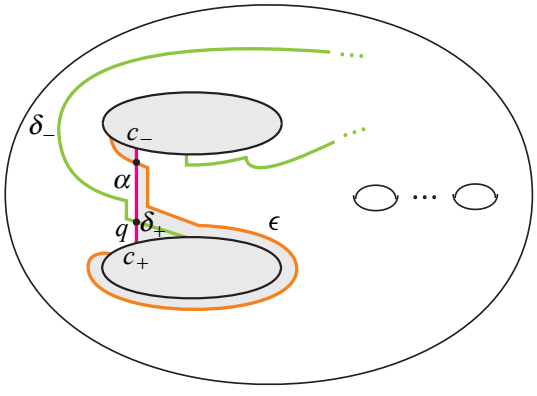


Figure 5: The arcs α and ϵ.

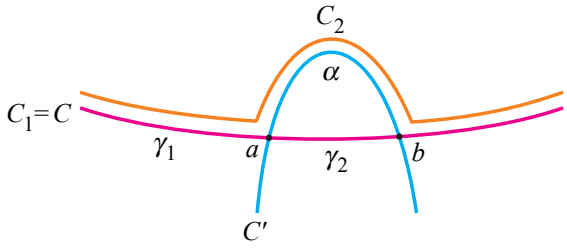


Figure 6: The primitive curve C_2 in case 1.

Proof If $C \cap C' = \varnothing$, then we are done. If $|C \cap C'| = 1$, then the result holds by Lemma 2.1. So suppose that $|C \cap C'| \geq 2$. Let $p = C \cap \partial D$ and $p' = C' \cap \partial D$.

Consider a subarc α of C' cut off by C such that the interior of α is disjoint from C. Let a and b be the two endpoints of α. The two points a and b cut C into two arcs γ_1 and γ_2. If α contains p', then let γ_1 be the arc that does not contain p. If α does not contain p', then let γ_1 be the arc that contains p. So in any case, $\gamma_1 \cup \alpha$ intersects ∂D in a single point. There are two cases according to the sides that α is incident to C at a and b.

Case 1 (α is incident to C in the same side of C at a and b) By a surgery of C along α, we obtain a new primitive curve $C_2 = \gamma_1 \cup \alpha$ with a dual disk D. After a slight isotopy, C_2 is disjoint from C_1 $(= C)$, and $|C_2 \cap C'| \leq |C_1 \cap C'| - 2$ (see Figure 6).

Case 2 (α is incident to C in the opposite sides of C at a and b) By a surgery of C along α, we obtain a new primitive curve $\Gamma = \gamma_1 \cup \alpha$ with a dual disk D. After a slight isotopy, $|\Gamma \cap C| = 1$, and $|\Gamma \cap C'| \leq |C \cap C'| - 1$ (see Figure 7). By Lemma 2.1, C and Γ are c-connected via primitive curves with a common dual disk D, and $d(C, \Gamma) \leq 5$.

By an inductive argument, we get a sequence of primitive curves with a common dual disk D, from C to C'. An upper bound of $d(C, C')$ in worst case occurs if every step of the surgery is case 2. Since the intersection number of curves decreases by only one in worst case, we have $d(C, C') \leq 5\,|C \cap C'|$. □

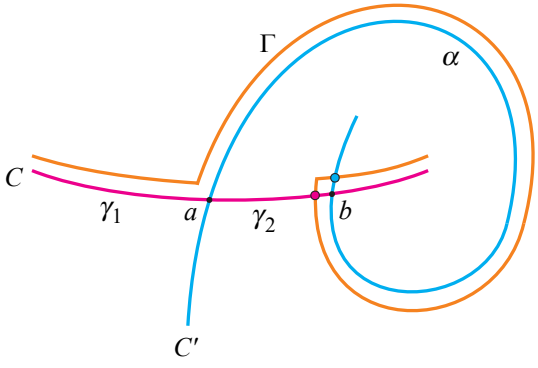


Figure 7: The primitive curve Γ in case 2.

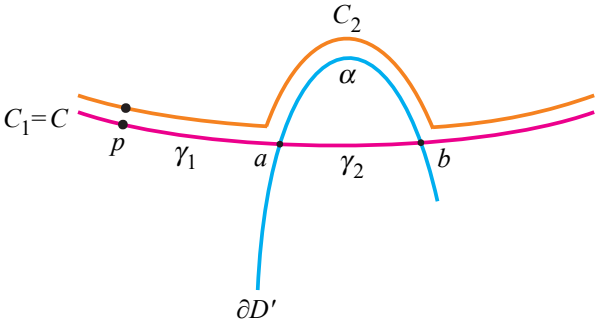


Figure 8: The primitive curve C_2 in case 1.

A key idea in the next lemma is that we do arc surgery with respect to a pair of a primitive curve and the boundary curve of an essential disk.

Lemma 2.3 (arc surgery) *Let C be a primitive curve in Σ, the boundary of a genus-g handlebody V for $g \geq 2$ with a dual disk D. Let D' be an essential disk in V disjoint from D. Then there exists a primitive curve C'' in Σ such that $|C'' \cap \partial D'| \leq 1$, and C and C'' are c-connected with the common dual disk D, and $d(C, C'') \leq 5(|C \cap \partial D'| - 1)$ if $|C \cap \partial D'| \geq 1$.*

Proof If $|C \cap \partial D'| \leq 1$, then we are done by taking $C'' = C$. So suppose that $|C \cap \partial D'| \geq 2$. Let $p = C \cap \partial D$. Consider a subarc α of $\partial D'$ cut off by C such that the interior of α is disjoint from C. Let a and b be the two endpoints of α. The two points a and b cut C into two arcs γ_1 and γ_2, and let γ_1 be the arc that contains p. There are two cases according to the sides that α is incident to C at a and b.

Case 1 (α is incident to C in the same side of C at a and b) By a surgery of C along α, we obtain a new primitive curve $C_2 = \gamma_1 \cup \alpha$ with a dual disk D. After a slight isotopy, C_2 is disjoint from C_1 $(= C)$, and $|C_2 \cap \partial D'| \leq |C_1 \cap \partial D'| - 2$ (see Figure 8).

Case 2 (α is incident to C in the opposite sides of C at a and b) By a surgery of C along α, we obtain a new primitive curve $\Gamma = \gamma_1 \cup \alpha$ with a dual disk D. After a slight isotopy, $|\Gamma \cap C| = 1$, and $|\Gamma \cap \partial D'| \leq |C \cap \partial D'| - 1$ (see Figure 9). By Lemma 2.1, C and Γ are c-connected via primitive curves with a common dual disk D.

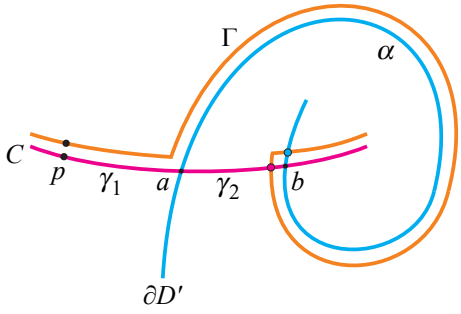


Figure 9: The primitive curve Γ in case 2.

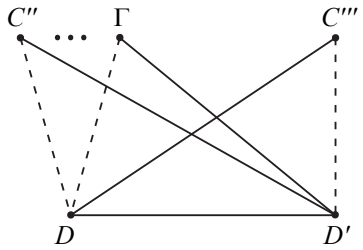


Figure 10: $|(\Gamma \cup D) \cap (C''' \cup D')| \leq 1$.

By an inductive argument, we get a primitive curve C'' with a dual disk D and $|C'' \cap \partial D'| \leq 1$. An upper bound of $d(C, C'')$ in worst case is $5(|C \cap \partial D'| - 1)$ because the surgery process finishes if $|C'' \cap \partial D'| = 1$. □

Now we are ready to prove the following theorem, a weaker version of Theorem 1.1.

Theorem 2.4 *Let C and C' be primitive curves in Σ, the boundary of a genus-g handlebody V for $g \geq 2$. Then C and C' are c-connected.*

Proof Take any dual disks D and D' of C and C', respectively. By the standard disk surgery argument, there exists a sequence $D = D_1, D_2, \ldots, D_n = D'$ of nonseparating disks in V such that D_j and D_{j+1} are disjoint for each $j \in \{1, 2, \ldots, n-1\}$. Let $C_1 = C$ and $C_n = C'$. Take a primitive curve C_j for each $j \in \{2, \ldots, n-1\}$ such that D_j is a dual disk of C_j.

To show the statement of Theorem 2.4 for C and C', it is enough to show the same statement for C_j and C_{j+1}, and each $j \in \{1, 2, \ldots, n-1\}$. So without loss of generality, we assume that C and C' admit dual disks D and D', respectively, such that D and D' are disjoint.

By Lemma 2.3, there exists a primitive curve C'' such that C and C'' are c-connected with a common dual disk D and $|C'' \cap \partial D'| \leq 1$. Similarly, there exists a primitive curve C''' such that C' and C''' are c-connected with a common dual disk D' and $|C''' \cap \partial D| \leq 1$. It remains to show that C'' and C''' are c-connected.

Suppose that one of $|C'' \cap \partial D'|$ or $|C''' \cap \partial D|$, say $|C'' \cap \partial D'|$, is 1. Then D' is a common dual disk of C'' and C'''. By Lemma 2.2, C'' and C''' are c-connected.

Now suppose that both $|C'' \cap \partial D'|$ and $|C''' \cap \partial D|$ are 0. We do an arc surgery for the pair C'' and C''' to reduce $|C'' \cap C'''|$. More precisely, we do a surgery of C'' along an arc component α of C''' cut off by C'' such that the interior of α is disjoint from C'' and that α does not contain the point $C''' \cap \partial D'$. Then we obtain a primitive curve Γ_1 satisfying $|\Gamma_1 \cap C'''| < |C'' \cap C'''|$ with a common dual disk D, and with an additional property that $|\Gamma_1 \cap \partial D'| = 0$ since $|\Gamma_1 \cap \partial D'| \leq |C'' \cap \partial D'| = 0$ by the choice of α. We iterate such an arc surgery to obtain a primitive curve Γ from C'' with a common dual disk D, and $|\Gamma \cap C'''| \leq 1$, and $|\Gamma \cap \partial D'| = 0$. The two primitive curves C'' and Γ are c-connected (using Lemma 2.1 if necessary). See Figure 10. In Figure 10, the dotted lines mean that $|C'' \cap \partial D| = |\Gamma \cap \partial D| = 1$ and $|C''' \cap \partial D'| = 1$, and the solid lines mean that $D \cap D' = \varnothing$ and $(C'' \cup \Gamma) \cap D' = \varnothing$ and $C''' \cap D = \varnothing$.

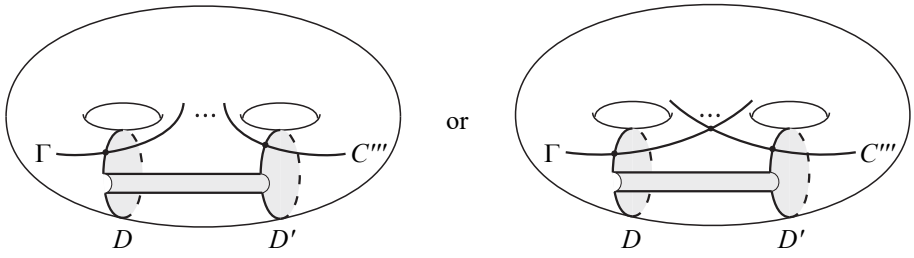


Figure 11: A band sum of D and D' is a common dual disk of Γ and C'''.

Since $|(\Gamma \cup D) \cap (C''' \cup D')| = |\Gamma \cap C'''| \leq 1$, we can take a band sum of D and D' that is a common dual disk of Γ and C'''. See Figure 11.

If $|\Gamma \cap C'''| = 0$, then Γ and C''' are c-connected obviously. If $|\Gamma \cap C'''| = 1$, then Γ and C''' are again c-connected by Lemma 2.1. □

3 Proof of the main theorem in the case of $g = 2$

For convenience, we simply say that two primitive curves C and C' *are s-connected* if there exists a sequence $C = C_1, C_2, \ldots, C_k = C'$ of primitive curves for some k such that C_i and C_{i+1} are separated for each $i \in \{1, 2, \ldots, k-1\}$. If C and C' are s-connected, that is, C_i and C_{i+1} are separated with their disjoint dual disks D_i and D_{i+1}, respectively, then we see that C and C' are also c-connected by taking a band sum of D_i and D_{i+1} that forms a common dual disk of C_i and C_{i+1}. We will prove the following, which is Theorem 1.1 in the case of $g = 2$.

Theorem 3.1 *Let C and C' be primitive curves in Σ, the boundary of a genus-2 handlebody V. Then C and C' are s-connected.*

Proof By Theorem 2.4, C and C' are c-connected (with an arbitrary choice of dual disks of C and C'), and hence it suffices to assume that C and C' are disjoint and have a common dual disk.

Let D be a common dual disk of C and C'. Cutting V along D, we have a solid torus V' with two copies D_+ and D_- of D on $\Sigma' = \partial V'$. The primitive curve C is cut into an arc α with one of its endpoints in ∂D_+ and the other in ∂D_-. The primitive curve C' is also cut into an arc β with its endpoints in ∂D_+ and ∂D_-, respectively. If we crush D_+ and D_- to points, then $\alpha \cup \beta \cup D_+ \cup D_-$ becomes a loop, denoted by ℓ. The loop ℓ is homotopic to an inessential loop in Σ' only if C and C' are isotopic in Σ. So we may assume that ℓ is homotopic to an essential loop in the torus Σ'.

Fix a meridian m_0 and a longitude l_0 of $\Sigma' = \partial V'$. Suppose that ℓ is homotopic to a (p, q)-torus knot $K_{p,q}$. Here $K_{p,q}$ winds V' $|p|$ times in longitudinal direction and $|q|$ times in meridional direction in such a way that $|K_{p,q} \cap m_0| = |p|$ and $|K_{p,q} \cap l_0| = |q|$.

Suppose that $(p, q) = (0, 1)$, i.e., ℓ is homotopic to a meridian. Take a meridian disk D_1 of V' such that $D_1 \cap (\alpha \cup \beta \cup D_+ \cup D_-) = \varnothing$. Take another meridian disk D_2 of V' such that $D_2 \cap (D_1 \cup D_+ \cup D_-) = \varnothing$ and $|\partial D_2 \cap \alpha| = |\partial D_2 \cap \beta| = 1$. See Figure 12.

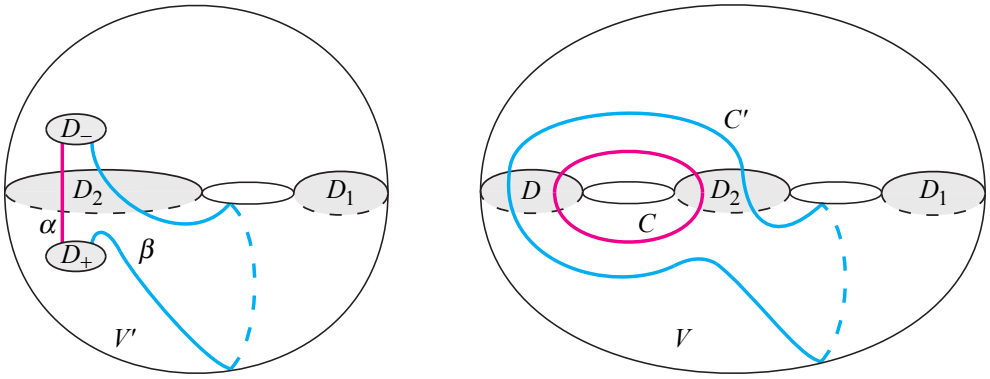


Figure 12: The loop ℓ is homotopic to a meridian.

We take a primitive curve C'' in Σ with the following properties (see Figure 13(a)).

- $C'' \cap (C \cup D) = \varnothing$.
- $|C'' \cap \partial D_1| = |C'' \cap \partial D_2| = 1$.

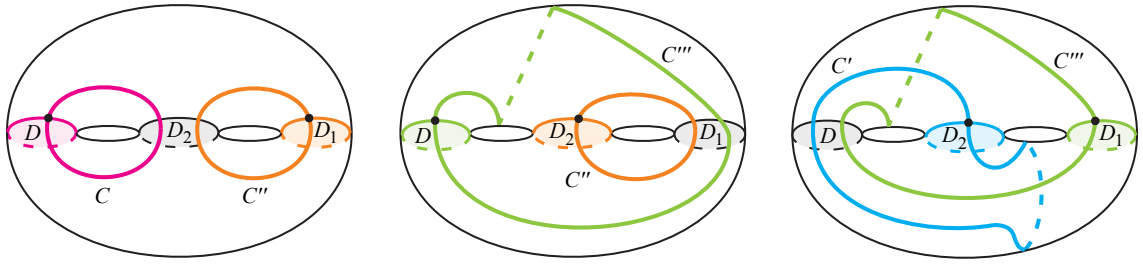
It is not difficult to take another primitive curve C''' in Σ with the following properties (see Figure 13(b) and (c)).

- $C''' \cap (C'' \cup D_2) = \varnothing$.
- $C''' \cap C' = \varnothing$.
- $|C''' \cap \partial D| = |C''' \cap \partial D_1| = 1$.

The curves C and C' are s-connected because $(C \cup D) \cap (C'' \cup D_1) = \varnothing$ and $(C'' \cup D_2) \cap (C''' \cup D) = \varnothing$ and $(C''' \cup D_1) \cap (C' \cup D_2) = \varnothing$. See Figure 13.

Suppose that $(p, q) = (1, 0)$, i.e., the loop ℓ is homotopic to a longitude. Figure 14 illustrates that $(C \cup D_2) \cap (C' \cup D_1) = \varnothing$. So C and C' are s-connected directly.

The signs of p and q just reflect left-handed or right-handed directions of the winding of $K_{p,q}$ with a choice of orientation. Thus without loss of generality, we only consider those p, q with $p > 0$ and $q > 0$.



(a) (b) (c)

Figure 13: C and C' are s-connected via C'' and C'''.

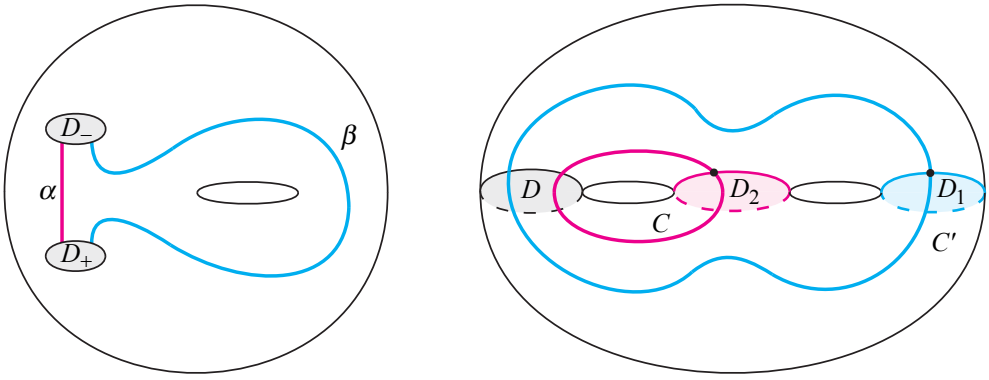


Figure 14: C and C' are separated.

Moreover, if $q \geq p > 0$, then by sufficient Dehn twists of $K_{p,q}$ along m_0, we make q is less than p. So we may assume that $p > q > 0$.

We use induction on p. For that purpose, we consider two types of arcs $\beta_{p,q}$ and $\beta'_{p,q}$ as in Figure 15(a) and Figure 15(b), respectively. Both $\beta_{p,q}$ and $\beta'_{p,q}$ have one of their endpoints in ∂D_+ and the other in ∂D_-. The arc $\beta_{p,q}$ is disjoint from α, but $\beta'_{p,q}$ intersects α in a single point, cutting off a small triangular disk together with D_-. Let $\ell_{p,q}$ and $\ell'_{p,q}$ be the loops obtained from $\beta_{p,q} \cup \alpha \cup D_+ \cup D_-$ and $\beta'_{p,q} \cup \alpha \cup D_+ \cup D_-$ by crushing D_+ and D_- to points, respectively. Both $\ell_{p,q}$ and $\ell'_{p,q}$ are homotopic to $K_{p,q}$. Figure 15 illustrates an example of the case $(p, q) = (7, 5)$.

Let r and s be positive integers satisfying $ps - qr = 1$ and $r < p$ and $s \leq q$. Since $\det\left(\begin{smallmatrix} p & r \\ q & s \end{smallmatrix}\right) = ps - qr = 1$, $|K_{p,q} \cap K_{r,s}| = 1$. For $\beta_{p,q}$ and $\beta'_{p,q}$, we take arcs $\beta_{r,s}$ and $\beta'_{r,s}$, respectively, winding r times in longitudinal direction and s times in meridional direction and $|\beta_{r,s} \cap \beta_{p,q}| = |\beta'_{r,s} \cap \beta'_{p,q}| = 1$ as in Figure 16(a) and Figure 16(b). Both $\beta_{r,s}$ and $\beta'_{r,s}$ have one of their endpoints in ∂D_+ and the other

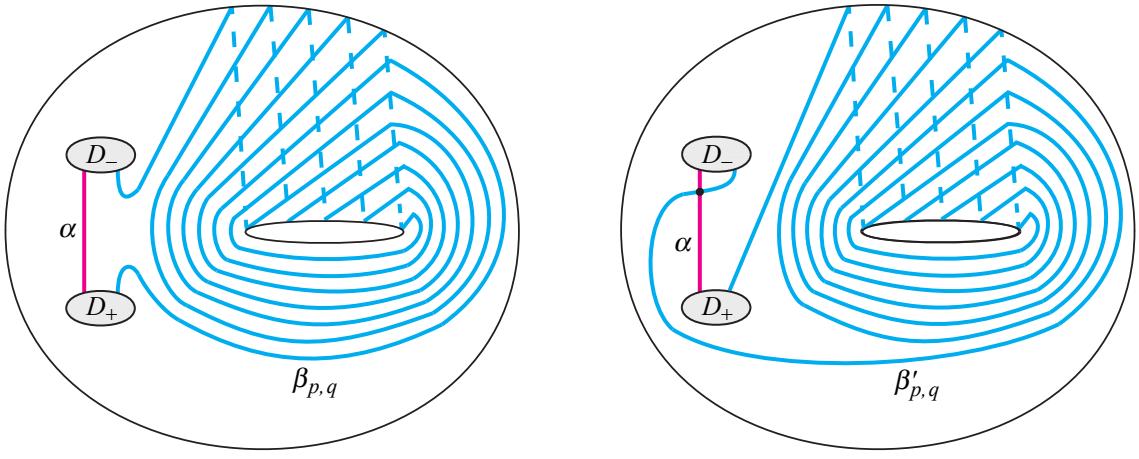


Figure 15: (a) $|\beta_{p,q} \cap \alpha| = 0$ and (b) $|\beta'_{p,q} \cap \alpha| = 1$.

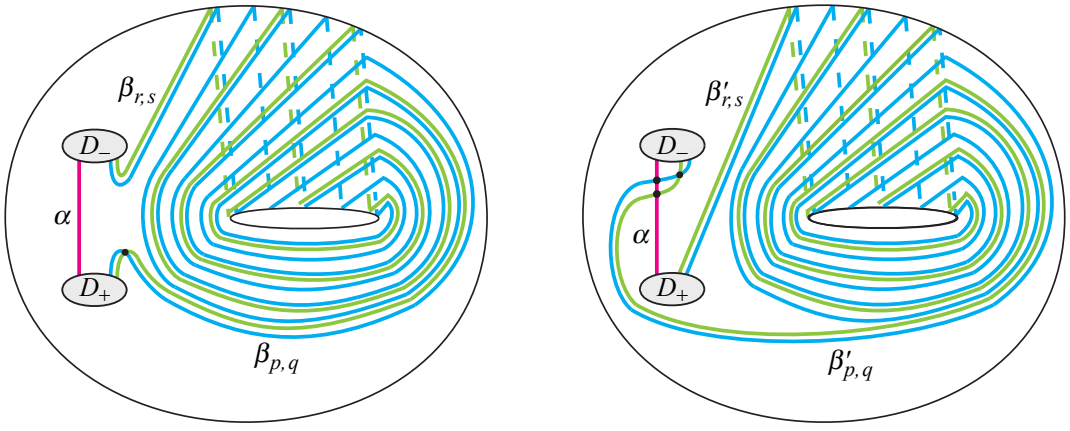


Figure 16: (a) $|\beta_{r,s} \cap \alpha| = 0$ and (b) $|\beta'_{r,s} \cap \alpha| = 1$, and $|\beta_{r,s} \cap \beta_{p,q}| = |\beta'_{r,s} \cap \beta'_{p,q}| = 1$.

in ∂D_-. Moreover, $|\beta_{r,s} \cap \alpha| = 0$ and $|\beta'_{r,s} \cap \alpha| = 1$. The arc $\beta_{r,s}$ is like a stepping stone between α and $\beta_{p,q}$, and the arc $\beta'_{r,s}$ is like a stepping stone between α and $\beta'_{p,q}$.

The disk D_- in Figure 16(a) is isotoped along two long parallel subarcs of $\beta_{r,s}$ and $\beta_{p,q}$ as in Figure 17(a) so that $\beta_{r,s}$ becomes a short arc and $\beta_{p,q}$ winds $p-r$ times in longitudinal direction. Denote the isotoped $\beta_{r,s}$ and $\beta_{p,q}$ by α_0 and $\beta'_{p-r,t}$ for some $t < q$, respectively. Similarly, D_+ in Figure 16(b) is isotoped along two long parallel subarcs of $\beta'_{r,s}$ and $\beta'_{p,q}$ as in Figure 17(b) so that $\beta'_{r,s}$ becomes a short arc and $\beta'_{p,q}$ winds $p-r$ times in longitudinal direction. Denote the isotoped $\beta'_{r,s}$ and $\beta'_{p,q}$ by α_0 and $\beta'_{p-r,t}$ for some $t < q$, respectively. In both cases, let $\ell'_{p-r,t}$ be the loop obtained from $\alpha_0 \cup \beta'_{p-r,t} \cup D_+ \cup D_-$ by crushing D_+ and D_- to points. Then $\ell'_{p-r,t}$ is homotopic to $K_{p-r,t}$. Since $K_{p-r,t}$ intersects the original $K_{p,q}$ in a single point and $\det\left(\begin{smallmatrix} p & p-r \\ q & q-s \end{smallmatrix}\right) = p(q-s) - q(p-r) = -1$, we have that $t = q-s$.

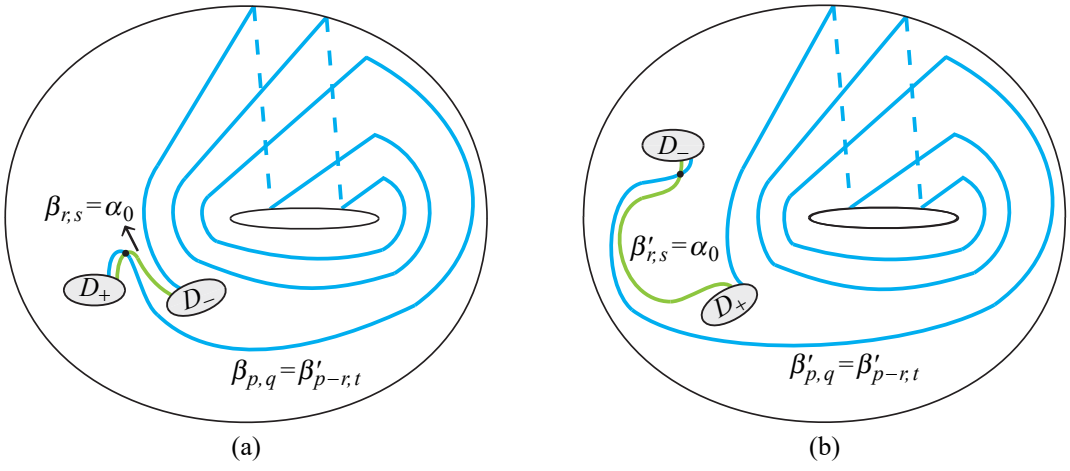


Figure 17: $\beta'_{p-r,t} = \beta'_{p-r,q-s}$.

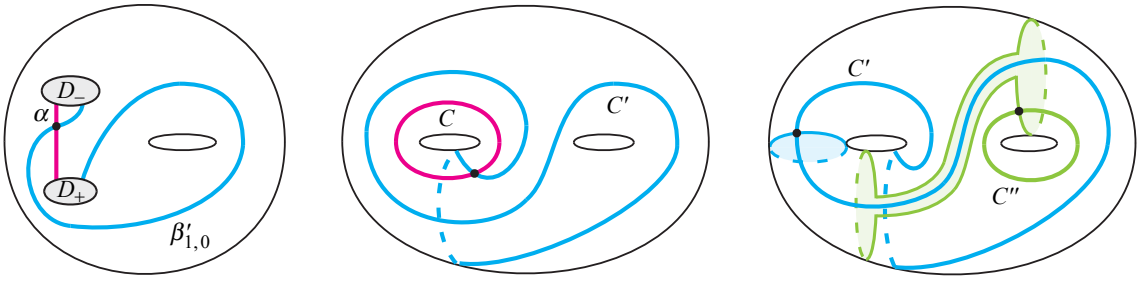


Figure 18: C and C' are s-connected via C''.

Hence the longitudinal and meridional parameters of $\beta_{p,q}$ with respect to $\beta_{r,s}$ and those of $\beta'_{p,q}$ with respect to $\beta'_{r,s}$ are $(p-r, q-s)$. Figures 15, 16, and 17 illustrate an example of $(p,q) = (7,5)$ and $(r,s) = (4,3)$ and $(p-r, q-s) = (3,2)$.

Now it remains only the case of $(p,q) = (1,0)$, the initial condition of the induction. We only need to consider $|\beta'_{1,0} \cap \alpha| = 1$ because we already dealt with the case of $|\beta_{1,0} \cap \alpha| = 0$ (see Figure 14). Figure 18 shows that C and C' are s-connected via C'', where C and C' are obtained from α and $\beta'_{1,0}$, respectively. (It is easy to see that C and C'' are separated.)

Let C, C'', C' be the primitive curves in ∂V obtained by identifying endpoints of α, $\beta_{r,s}$ (or $\beta'_{r,s}$), $\beta_{p,q}$ (or $\beta'_{p,q}$), respectively. Since $r < p$, C and C'' are s-connected by our induction hypothesis. Since $p - r < p$, C'' and C' are s-connected by our induction hypothesis. Hence C and C' are s-connected. □

4 Proof of the main theorem in the case of $g \geq 3$

Throughout the section, V will be assumed to be a genus-g handlebody for $g \geq 3$. For a primitive curve C in $\Sigma = \partial V$ with a dual disk D, we call the pair (C, D) simply a *dual pair* for V. For convenience, we simply say that two dual pairs (C, D) and (C', D') *are p-connected* if there exists a sequence $(C, D) = (C_1, D_1), (C_2, D_2), \ldots, (C_k, D_k) = (C', D')$ of dual pairs for some k such that $C_i \cup D_i$ and $C_{i+1} \cup D_{i+1}$ are disjoint for each $i \in \{1, 2, \ldots, k-1\}$. It is obvious that if two dual pairs (C, D) and (C', D') are p-connected then C and C' are s-connected.

Lemma 4.1 (common primitive curve) *Let (C, D) and (C, D') be dual pairs for a genus-g handlebody V with $g \geq 3$. Suppose that D and D' are disjoint. Then (C, D) and (C, D') are p-connected.*

Proof Cut V along $D \cup D'$. The resulting manifold V' is a genus-$(g-2)$ handlebody, or two handlebodies of genus g_1 and g_2, respectively, where $g_1 + g_2 = g - 1$. The primitive curve C is cut into two arcs α_1 and α_2. Whether V' is connected or not, there are two copies D_+ and D_- of D and two copies D'_+ and D'_- of D' on $\Sigma' = \partial V'$. One of α_1 and α_2, say α_1, connects D_+ and D'_- and the other α_2 connects D'_+ and D_-. Since each of $D_+ \cup \alpha_1 \cup D'_-$ and $D'_+ \cup \alpha_2 \cup D_-$ is homotopy equivalent to a point, we can take a dual pair (C'', D'') in V' disjoint from them. See Figure 19. Then (C'', D'') can be regarded as a dual pair also in V, and (C, D) and (C, D') are p-connected via (C'', D''). □

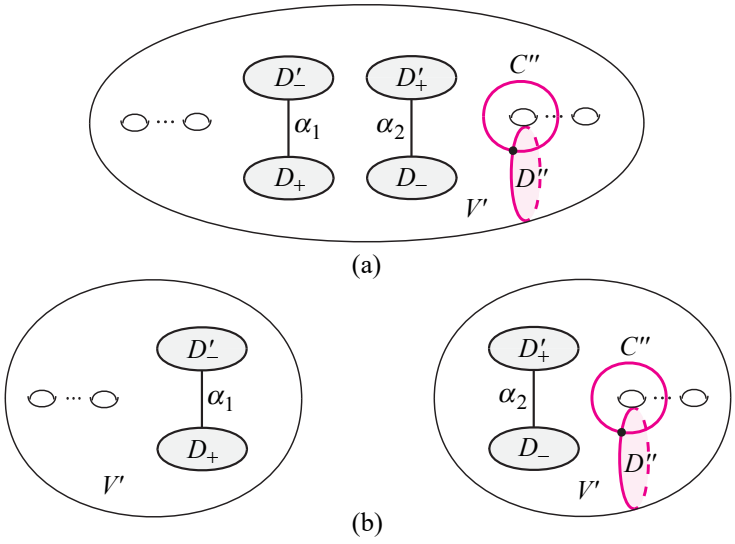


Figure 19: (a) V' is a handlebody and (b) V' is two handlebodies.

Lemma 4.2 *Let (C, D) and (C', D') be dual pairs for a genus-g handlebody V with $g \geq 3$. Suppose that $(C' \cup D') \cap D = \varnothing$. Then (C, D) and (C', D') are p-connected.*

Proof Cut V along D. Then we have a genus-$(g-1)$ handlebody V' with two copies D_+ and D_- of D on $\Sigma' = \partial V'$. The primitive curve C is cut into an arc α in Σ' such that one endpoint is in ∂D_+ and the other endpoint is in ∂D_-. Since $(C' \cup D') \cap D = \varnothing$, (C', D') is a dual pair also in V'. If $(C' \cup D') \cap \alpha = \varnothing$, then (C, D) and (C', D') are p-connected and we are done. So we may assume that $(C' \cup D') \cap \alpha \neq \varnothing$.

We use induction on $|(C' \cup D') \cap \alpha|$. Let p be a point of $(C' \cup D') \cap \alpha$ which is closest to one of the two endpoints of α, say c_+, in ∂D_+. There are two cases.

Case 1 $(p \in C')$ Let $C'_1 = C'$, and C'_2 be a parallel copy of C'_1 such that a point $q \in C'_2 \cap \alpha$ is closer to c_+ than p. Let α_+ be the subarc of α from q to c_+. Slide a small neighborhood of q in C'_2 along $\alpha_+ \cup D_+$. Then $C'_1 \cup C'_2$ bounds an annulus A containing D_+. See Figure 20.

Since (C'_1, D') is a dual pair in a genus-$(g-1)$ handlebody V', and $C'_1 \cup C'_2$ bounds an annulus, and D_- is disjoint from $D' \cup A$, we can take a new dual pair (C_3, D_3) in V' disjoint from $D' \cup A \cup D_-$. All of (C'_1, D') and (C'_2, D') and (C_3, D_3) can be regarded as dual pairs in V because they are disjoint from $D_+ \cup D_-$. The dual pairs (C'_1, D') and (C'_2, D') are p-connected via (C_3, D_3). Since $|(C'_2 \cup D') \cap \alpha| = |(C'_1 \cup D') \cap \alpha| - 1$, by an inductive argument (C'_2, D') and (C, D) are p-connected.

Case 2 $(p \in \partial D')$ Let $D'_1 = D'$, and D'_2 be a parallel copy of D'_1 such that a point $q \in \partial D'_2 \cap \alpha$ is closer to c_+ than p. Let α_+ be the subarc of α from q to c_+. Slide a small neighborhood of q in D'_2 along $\alpha_+ \cup D_+$. Then $D'_1 \cup D'_2$ bounds a region B in V' that is homeomorphic to $D' \times I$ containing D_+.

Since (C', D'_1) is a dual pair in a genus-$(g-1)$ handlebody V', and $D'_1 \cup D'_2$ bounds a 3-ball, and D_- is disjoint from $B \cup C'$, we can take a new dual pair (C_3, D_3) in V' disjoint from $B \cup C' \cup D_-$. All of (C', D'_1)

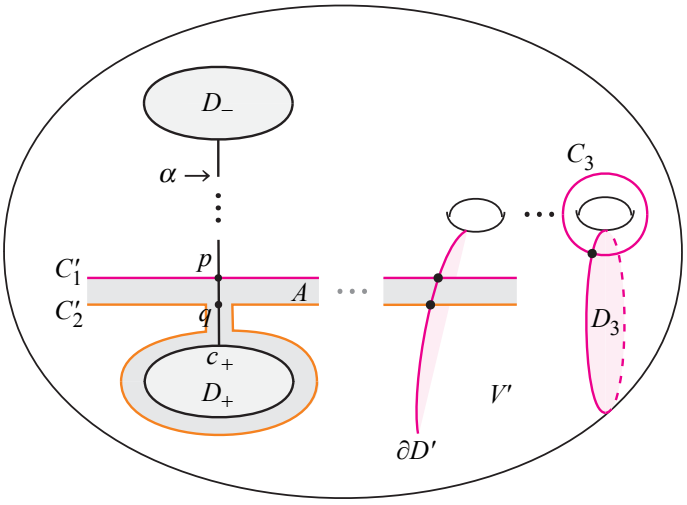


Figure 20: The dual pairs (C_1', D') and (C_2', D').

and (C', D_2') and (C_3, D_3) can be regarded as dual pairs in V because they are disjoint from $D_+ \cup D_-$. The dual pairs (C', D_1') and (C', D_2') are p-connected via (C_3, D_3). Since $|(C' \cup D_2') \cap \alpha| = |(C' \cup D_1') \cap \alpha| - 1$, by an inductive argument (C', D_2') and (C, D) are p-connected. □

Lemma 4.3 (common dual disk) *Let (C, D) and (C', D) be dual pairs for a genus-g handlebody V with $g \geq 3$. Then (C, D) and (C', D) are p-connected.*

Proof Cut V along D. Then we have a genus-$(g-1)$ handlebody V' with two copies D_+ and D_- of D on $\Sigma' = \partial V'$. The primitive curves C and C' are cut into arcs α and α' in Σ', respectively, such that for each arc one endpoint is in ∂D_+ and the other endpoint is in ∂D_-. Since $D_+ \cup \alpha' \cup D_-$ is homotopy equivalent to a point, we can take a dual pair (C'', D'') in V' disjoint from it. Then (C'', D'') can be regarded as a dual pair also in V, and (C'', D'') and (C', D) are p-connected. Since $(C'' \cup D'') \cap D = \varnothing$, by Lemma 4.2 (C, D) and (C'', D'') are p-connected. Hence (C, D) and (C', D) are p-connected. □

Now we are ready to prove Theorem 1.1 in the case of $g \geq 3$. Actually we show a slightly stronger version of it as follows.

Theorem 4.4 *Let (C, D) and (C', D') be dual pairs for a genus-g handlebody V with $g \geq 3$. Then (C, D) and (C', D') are p-connected.*

Proof As in the proof of Theorem 2.4, we may assume that D and D' are disjoint. By Lemma 2.3, there exists a sequence $C = C_1, C_2, \ldots, C_n = C''$ of primitive curves with a common dual disk D such that C_i and C_{i+1} are disjoint for each $i \in \{1, 2, \ldots, n-1\}$ and $|C'' \cap \partial D'| \leq 1$. Since (C_i, D) and (C_{i+1}, D) are p-connected for each $i \in \{1, 2, \ldots, n-1\}$ by Lemma 4.3, (C, D) and (C'', D) are p-connected. See Figure 21. In Figure 21, the dotted lines mean that (C_i, D) for each $i \in \{1, 2, \ldots, n\}$ and (C', D') are dual pairs, and the solid lines mean that $C_i \cap C_{i+1} = \varnothing$ for each $i \in \{1, 2, \ldots, n-1\}$ and $D \cap D' = \varnothing$. It remains to show that (C'', D) and (C', D') are p-connected.

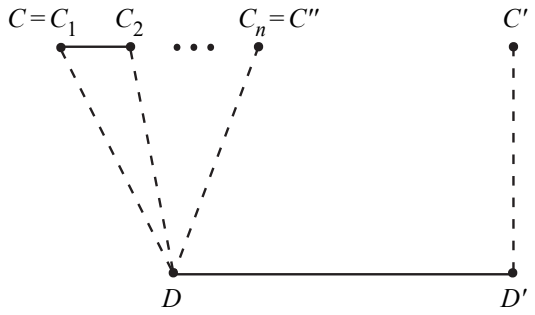


Figure 21: (C, D) and (C'', D) are p-connected and $|C'' \cap \partial D'| \leq 1$.

Suppose that $|C'' \cap \partial D'| = 1$. Then D' is a common dual disk of C'' and C'. By Lemma 4.3, (C'', D') and (C', D') are p-connected. Since (C'', D) and (C'', D') are p-connected by Lemma 4.1, (C'', D) and (C', D') are p-connected.

Now suppose that $|C'' \cap \partial D'| = 0$. Then since $(C'' \cup D) \cap D' = \varnothing$, by Lemma 4.2 (C'', D) and (C', D') are p-connected. □

Acknowledgements

The authors would like to thank the anonymous referee for valuable comments that improved the presentation significantly.

Cho is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2021R1F1A1060603 and RS-2024-00456645). Lee is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00275419), and supported by Global - Learning & Academic research institution for Master's · PhD students, and Postdocs (LAMP) Program of the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (RS-2024-00443714).

References

[1] **J L Harer**, *The virtual cohomological dimension of the mapping class group of an orientable surface*, Invent. Math. 84:1 (1986) 157–176 MR

[2] **B Wajnryb**, *Mapping class group of a handlebody*, Fund. Math. 158:3 (1998) 195–228 MR

[3] **A Zupan**, *The Powell conjecture and reducing sphere complexes*, J. Lond. Math. Soc. (2) 101:1 (2020) 328–348 MR

SANGBUM CHO scho@hanyang.ac.kr
Department of Mathematics Education, Hanyang University, Seoul, South Korea

JUNG HOON LEE junghoon@jbnu.ac.kr
Department of Mathematics and Institute of Pure and Applied Mathematics, Jeonbuk National University, Jeonju, South Korea

Received: February 1, 2024 Revised: December 15, 2024

msp

Algebraic & Geometric Topology **26**:3 (2026) 989–1035
DOI: 10.2140/agt.2026.26.989
Published: April 1, 2026

Extensions of finitely generated Veech groups

ELIOT BONGIOVANNI

Given a closed surface S with finitely generated Veech group G and its $\pi_1(S)$-extension Γ, there exists a hyperbolic space $\widehat{E}$ on which Γ acts isometrically and cocompactly. The space $\widehat{E}$ is obtained by collapsing some regions of the surface bundle over the convex hull of the limit set of G. Using the nice action of Γ on the hyperbolic space $\widehat{E}$, it is shown that Γ is hierarchically hyperbolic. These are generalizations of Dowdall–Durham–Leininger–Sisto, who assume in addition that G is a lattice. Because finitely generated Veech groups are among the most basic examples of subgroups of mapping class groups which are expected to qualify as geometrically finite, this result is evidence for the development of a broader theory of geometric finiteness.

1 Introduction

In the context of Kleinian groups, there is a well-defined notion of "convex cocompactness" and a generalization known as "geometric finiteness". For subgroups of mapping class groups, there is an analogous notion of convex cocompactness, but it is unclear what might be meant by geometric finiteness in this context. Finitely generated Veech groups are of interest here because they are subgroups of the mapping class group (those which stabilize Teichmüller disks) and Kleinian groups (since each of those Teichmüller disks is isometric to the hyperbolic plane). In general, a finitely generated Veech group is not convex cocompact—neither as a subgroup of a mapping class group nor as a Kleinian group—but it is geometrically finite as a Kleinian group. It is widely agreed upon that finitely generated Veech groups should qualify as geometrically finite as subgroups of mapping class groups; see Mosher's discussion in [23, Section 6]. For example, Tang [28, Theorem 1.4] shows that finitely generated Veech groups are parabolically geometrically finite in the sense of Dowdall–Durham–Leininger–Sisto [9], so these serve as a fundamental example for developing a theory of geometric finiteness in the context of mapping

MSC2020: primary 20F65, 20F67, 57M60; secondary 30F40, 30F60.

class groups. It is also known from Farb–Mosher and Hamenstädt that a subgroup of the mapping class group is convex cocompact if and only if its extension group is hyperbolic [11; 15]. Literature including that of Mosher [23, Problem 6.2], Dowdall–Durham–Leininger–Sisto [9, Section 1.4], and Russell [26, Section 1.1] suggests that some notion of geometric finiteness of subgroups of mapping class groups corresponds to hierarchical hyperbolicity of the extension group. This paper provides further evidence toward that conclusion.

The main result is a generalization of Dowdall–Durham–Leininger–Sisto [8, Theorem 1.1]. The following statement is identical, except that the term "lattice" has been generalized to "finitely generated". The $\pi_1(S)$ extension of G is precisely the group Γ that fits into the short exact sequence

$$1 \to \pi_1(S) \to \Gamma \to G \to 1,$$

and the vertex subgroups of Γ are those that stabilize the vertices of particular Bass–Serre trees upon which Γ acts isometrically.

Theorem 5.8 *Suppose $G < \mathrm{MCG}(S)$ is a finitely generated, nonelementary Veech group with extension group Γ and let $\Upsilon_1, \dots, \Upsilon_k < \Gamma$ be representatives of the conjugacy classes of vertex subgroups. Then Γ admits an isometric action on Gromov hyperbolic space $\widehat{E}$, quasi-isometric to the Cayley graph of Γ coned off along the cosets of $\Upsilon_1, \dots, \Upsilon_k$.*

Although this paper follows the basic construction and argument outline from [8], the details differ substantially. First, in both cases, the space $\widehat{E}$ is constructed from the hyperbolic plane bundle over the convex hull of the limit set of the associated Teichmüller disk. However, when G was also assumed to be a lattice, the convex hull of the limit set coincided with the Teichmüller disk. When G is only finitely generated, the convex hull of the limit set is generally some (strict) subset of the Teichmüller disk. As a consequence, the proof of the "fan lemma" (Lemma 6.22, which generalizes [8, Lemma 4.12]) is reformulated entirely. The fan lemma is the cornerstone for constructing sets that form slim triangles, which are used to prove hyperbolicity of $\widehat{E}$ via the guessing geodesics criterion. (The guessing geodesics criterion essentially says that the existence of *paths* that form slim triangles is sufficient evidence for the existence of *geodesics* that form slim triangles; see Proposition 5.1. The version used here is due to Masur–Schleimer [21] and Bowditch [5].) As in [8], this paper constructs the "guessed geodesics" by concatenating hyperbolic geodesic segments orthogonal to the fibers with saddle connections within fibers. When G was also assumed to be a lattice, the Veech dichotomy ensured that every saddle connection arose as a boundary component of a cylinder decomposition. It is therefore necessary to generalize the Veech dichotomy to the case of finitely generated Veech groups. Although this generalization is known to experts and follows quickly from other long-established results, it appears to be missing from the literature.

Theorem 3.3 (generalized Veech dichotomy) *Let (S, X, q) be a flat surface with finitely generated Veech group G. Every direction $\alpha \in \Lambda(G) \subset \partial D$ is either minimal and uniquely ergodic or completely periodic and invariant by a parabolic element of the maximal Veech group.*

Section 3, therefore, can be read independently of the rest of the paper. Its relevance here is to arrive at Corollary 3.9, which states that saddle connections either arise as boundary components of cylinder decompositions (as in the classical Veech dichotomy) or are associated to directions not in the limit set of G. Therefore, when G is only finitely generated, it is possible for the "guessed geodesics" to include saddle connections which are arbitrarily long — an issue not encountered when G was also assumed to be a lattice — and this complicates the process of proving that these paths form slim triangles.

It is important to note that some arguments in this paper require that the finitely generated Veech group G has at least one parabolic element. If G has no parabolic elements (equivalently, if G is convex cocompact) then it is already known that the extension group is hyperbolic by the aforementioned results of [11; 15]. However, the arguments in this paper provide a new proof in this special case with little extra work.

Theorem 6.29 (special case of [11; 15]) *Let $G < \mathrm{MCG}(S)$ be a finitely generated Veech group with extension group Γ. If G has no parabolic elements, then Γ is Gromov hyperbolic.*

After establishing the main result of Theorem 5.8, hierarchical hyperbolicity of the extension group is a natural next step.

Theorem 7.1 (hierarchical hyperbolicity) *Let $G < \mathrm{MCG}(S)$ be a finitely generated Veech group with extension group Γ. Then Γ is a hierarchically hyperbolic group.*

The proof of this theorem proceeds almost exactly as in [9], after patching one subcase.

One might also ask whether an extension of a finitely generated Veech group is quasi-isometrically rigid. Indeed, this is a result of Margolis [19]. Though [9] gives a stronger result for extensions of lattice Veech groups, their version does not generalize to extensions of finitely generated Veech groups.

Outline

Section 2 surveys the background information needed to approach this problem and refers to resources that explore these concepts in more depth. Section 3 is dedicated to a generalization of the classical Veech dichotomy, which lays the foundation for the rest of the paper. Section 4 constructs the space $\widehat{E}$ featured in Theorem 5.8. Section 5 proves that $\widehat{E}$ is hyperbolic (Theorem 5.7), and shows how the rest of the statement of Theorem 5.8 follows. A key result for the proof of hyperbolicity of $\widehat{E}$, Theorem 6.28, is deferred to and comprises the entirety of Section 6. Finally, Section 7 proves that an extension of a finitely generated Veech group is hierarchically hyperbolic. For the reader's convenience, a list of symbols and references to the pages on which they first appear begins on page 1032.

2 Background concepts

This section surveys concepts used throughout the paper. References for deeper reading are included at the beginning of each subsection.

2.1 Fundamental geometry

2.1.1 Paths in metric spaces See [6, Chapter I.1].

Let (X, d) be a metric space. A *path* from x to y in X is a continuous map c from an interval $[0, \ell] \subset \mathbb{R}$ to X with $c(0) = x$ and $c(\ell) = y$. The *length* of the path c is

$$\sup_P \sum_{i=1}^{n_P} d(c(t_{i-1}), c(t_i)),$$

where the supremum is taken over all partitions P of $[0, \ell]$ with $t_0 = 0$ and $t_{n_P} = \ell$. If the length of c is finite, then c is called *rectifiable*. The space X is called *rectifiably path connected* if any two points are connected by a rectifiable path. If for any $x, y \in X$ the distance $d(x, y)$ is precisely the infimum of the lengths of all paths from x to y, then (X, d) is called a *length space*.

A path c is called a *geodesic* if for all $t, t' \in [0, \ell]$,

$$d\left(c(t), c(t')\right) = |t - t'|.$$

The image of the map c is also sometimes referred to as a geodesic. The space X is called a *geodesic space* if any two points are connected by a geodesic. A geodesic between x and y in a geodesic space is typically denoted $[x, y]$, which is the notation used throughout this section.

2.1.2 Quotient spaces See [6, pp. 2, 64–70].

Let (X, d) be a metric space, and let $\sim$ be an equivalence relation on elements of X. A *chain* from x to y in X is a sequence $\{x_1, y_1, x_2, y_2, \dots, x_n, y_n\}$ with $y_i \sim x_{i+1}$. For $x', y' \in X' = X/\sim$ representing $x, y \in X$, respectively, the *quotient pseudometric* is defined by

$$d'(x', y') := \inf_C \sum_{i=1}^{n} d(x_i, y_i),$$

where the infimum is taken over all chains joining x to y. A *pseudometric* meets all of the requirements to be a metric except that it is perhaps not *positive-definite*, meaning that $d'(x', y') = 0$ need not imply $x' = y'$. The next lemma is used several times in this paper to verify that a quotient space is in fact a length space.

Lemma 2.1 [6, Lemma I.5.20] *Let (X, d) be a length space, let $\sim$ be an equivalence relation on X and let d' be the quotient pseudometric on $X' = X/\sim$. If d' is a metric then (X', d') is a length space.*

2.1.3 Gromov hyperbolicity See [6, Chapter III.H.1; 31].

Let (X, d) be a metric space with $x \in X$. The *Gromov product of $y, z \in X$ with respect to x* is

$$(y \cdot z)_x := \tfrac{1}{2}\left(d(y, x) + d(z, x) - d(y, z)\right).$$

For $\delta \geq 0$, X is called *δ-hyperbolic* (or *Gromov hyperbolic with hyperbolicity constant δ*) if

$$(x \cdot y)_w \geq \min\{(x \cdot z)_w, (y \cdot z)_w\} - \delta$$

for all $w, x, y, z \in X$. When X is a geodesic space, δ-hyperbolicity is equivalent to every geodesic triangle in X being δ'-slim for some $\delta' \geq 0$ depending on δ: A *geodesic triangle* consists of three points

$x, y, z \in X$ and choices of geodesics $[x, y]$, $[y, z]$, and $[x, z]$ and is δ-*slim* if each side is contained in the δ-neighborhood of the other two, i.e.,

$$[x, z] \subset N_\delta([x, y] \cup [y, z]),$$
$$[x, y] \subset N_\delta([y, z] \cup [x, z]),$$
$$[y, z] \subset N_\delta([x, z] \cup [x, y]),$$

where N_δ denotes the δ-neighborhood.

In this paper, any mention of hyperbolicity is Gromov hyperbolicity (with some hyperbolicity constant).

2.1.4 Coarse geometry See [6, pp. 138–144; 31].

Let (X, d_X) and (Y, d_Y) be metric spaces. Given $\lambda \geq 1$ and $\varepsilon \geq 0$, a map $f : X \to Y$ is called a (λ, ε)-*quasi-isometric embedding* if for all $x, y \in X$,

$$\frac{1}{\lambda} d_X(x, y) - \varepsilon \leq d_Y(f(x), f(y)) \leq \lambda d_X(x, y) + \varepsilon.$$

When the domain X is an interval in $\mathbb{R}$ or $\mathbb{Z}$, the map f is called a *quasi-geodesic*; the image of the map is also sometimes referred to as a quasi-geodesic. If in addition there is a constant $K \geq 0$ so that $N_K(f(X)) = Y$ — that is, if f is *coarsely surjective* — then f is called a *quasi-isometry*, and the spaces X and Y are said to be quasi-isometric. If f is a quasi-isometry, then there exists a *quasi-inverse* of f, which is a quasi-isometry $g : Y \to X$ so that $d_X\big(x, g(f(x))\big)$ and $d_Y\big(y, f(g(y))\big)$ are uniformly bounded for all $x \in X$ and $y \in Y$.

Gromov hyperbolicity of length spaces is an invariant of quasi-isometry: If (X, d_X) and (Y, d_Y) are length spaces, X is hyperbolic, and $f : X \to Y$ is a quasi-isometry, then Y is hyperbolic (perhaps with a different hyperbolicity constant from that of X) [31, p. 16].

2.1.5 Cayley graphs Given a finitely generated group G with generating set $\mathcal{A}$, the *Cayley graph of* G, denoted $\mathrm{Cay}(G)$, is the graph whose vertices are elements of G and whose edges connect g to ga for any $g \in G$, $a \in \mathcal{A}$. Defining all edges to be unit length makes the Cayley graph a metric space, where the induced metric is precisely the word metric with respect to the set $\mathcal{A}$. The group G equipped with the word metric is quasi-isometric to its Cayley graph, and any two Cayley graphs for G (obtained from two different generating sets) are quasi-isometric (see [6, pp. 139–141]).

Given a finite family of subgroups $\mathcal{H} = \{H_1, H_2, \ldots, H_n\}$ of G, the *Cayley graph of* G *with respect to* $\mathcal{H}$ or *coned-off Cayley graph*, denoted $\mathrm{Cay}(G, \mathcal{H})$ is the graph obtained from $\mathrm{Cay}(G)$ by adding a vertex V_{gH} for each left coset gH (with $g \in G$ and $H \in \mathcal{H}$) and attaching V_{gH} by an edge of length $\frac{1}{2}$ to each vertex of $\mathrm{Cay}(G)$ which is an element of the coset gH. The isometric left action of G on $\mathrm{Cay}(G)$ extends to an isometric action of G on $\mathrm{Cay}(G, \mathcal{H})$: For all $g, g' \in G$ and $H \in \mathcal{H}$, set $gV_{g'H} = V_{gg'H}$. The nontrivial vertex stabilizers of the action are conjugate to subgroups of $\mathcal{H}$. See [7], including the following variant of the Schwarz–Milnor lemma.

Theorem 2.2 [7, Theorem 5.1] *Let G be a finitely generated group and suppose that G admits a discontinuous (that is, with discrete orbits), cocompact, isometric action on a length space X. Let $\mathcal{H}$ denote a*

collection of subgroups of G consisting of exactly one representative of each conjugacy class of maximal isotropy subgroups for the action of G on X. Then $\mathcal{H}$ is finite and, for any finite generating set $\mathcal{A}$ of G, the coned-off Cayley graph $\mathrm{Cay}(G,\mathcal{H})$ is quasi-isometric to X. In particular, if X is a hyperbolic space then the coned-off Cayley graph is hyperbolic.

2.1.6 Mapping class groups See [10].

Let S be a closed, orientable surface of genus at least 2. The *mapping class group of S* is

$$\mathrm{MCG}(S) := \pi_0\left(\mathrm{Homeo}^+(S)\right).$$

That is, the mapping class group is the group of isotopy classes of orientation-preserving homeomorphisms on S.

Mark a point $*$ on S and denote the marked surface by $\dot{S}$. The mapping class group of the marked surface is then

$$\mathrm{MCG}(\dot{S}) := \pi_0\left(\mathrm{Homeo}^+(S,*)\right),$$

i.e., the isotopy classes of elements of $\mathrm{Homeo}^+(S)$ which fix the marked point $*$. These mapping class groups fit into the *Birman exact sequence*,

$$1 \to \pi_1(S) \to \mathrm{MCG}(\dot{S}) \to \mathrm{MCG}(S) \to 1, \tag{1}$$

where $\mathrm{MCG}(\dot{S}) \to \mathrm{MCG}(S)$ is the map which forgets the marked point on S [10; 4]. For a subgroup $G < \mathrm{MCG}(S)$, denote its preimage by $\Gamma_G < \mathrm{MCG}(\dot{S})$ and observe that it fits into a short exact sequence

$$1 \to \pi_1(S) \to \Gamma_G \to G \to 1,$$

which includes into the Birman exact sequence above. The group Γ_G is called the *$\pi_1(S)$-extension of G* (or simply the extension of G), and it is the fundamental group of an S-bundle with monodromy an isomorphism onto G.

2.2 Flat surfaces

See [12; 27].

Let S be a closed, connected, and oriented surface of genus at least 2. Equip S with a *complex structure X*, which is an atlas of charts $\{z_\alpha : U_\alpha \to \mathbb{C}\}$ whose transition functions $z_\beta^{-1} \circ z_\alpha$ are biholomorphic wherever the composition is defined. A *quadratic differential* for X, denoted q_X or more simply q, is a nonzero holomorphic section of the square of the canonical line bundle over (S,X). Note in particular that any nonzero quadratic differential has finitely many zeroes. The pair (X,q) are a *flat structure* on S, described as follows.

In a small disk neighborhood of a nonzero point p, choose a coordinate chart z so that p corresponds to $z(p)=0$ and pick a branch of $q^{1/2}(z)$. The *natural coordinate* or *preferred coordinate* in a neighborhood of p is given by

$$\zeta(z) = \int_0^z q^{1/2}(u)\,du.$$

In this coordinate, q is given by $q(z)\,dz^2 = d\zeta^2$. In the neighborhood of a zero of order $k \geq 1$ there are natural coordinates such that $q(z)\,dz^2 = \zeta^k d\zeta^2$.

Away from the zeroes of q, the transition functions for overlapping preferred coordinates of q are locally given by $z \mapsto \pm z + c$ for some $c \in \mathbb{C}$. Because the Euclidean metric is invariant under these transition functions, the Euclidean metric pulls back to a metric on S minus the zeroes of q. To complete the pullback metric, each zero of order $k-2$ is filled back in so that a neighborhood of the point is isometric to the image of k Euclidean half planes glued together in a cyclic pattern by identifying the positive real axis of one half plane with the negative real axis of another (and only one other) half plane. A point of S corresponding to a zero of q — that is, a singular point — of order $k-2$ is called a *cone point* with *cone angle* $k\pi$. The resulting metric is called a *flat metric*, which is also denoted by q. (Using the same notation for both the quadratic differential and the flat metric is a bit imprecise: The flat metric determines the quadratic differential up to multiplication by a nonzero complex number.)

Given a complex structure X on S and an associated flat metric q, the triple (S, X, q) is a *flat surface*. Where the structure and metric are implied, a flat surface is often denoted simply as S.

2.2.1 Directions Because the transition functions for overlapping preferred coordinates for q are locally given by $z \mapsto \pm z + c$, a line in the tangent space at any nonsingular point can be parallel translated (everywhere except the cone points) to produce a smooth line field, which corresponds to a line in the projective tangent space at any nonsingular point. In other words, any tangent line to the surface has a distinguishable direction, up to a rotation by π, which is consistent everywhere on the surface away from the cone points. This is known as the *space of directions*. It is sometimes denoted by $\P^1(q)$, but in this paper the notation ∂D (introduced in Section 2.3) is used instead.

2.2.2 Geodesics Because the flat metric q is Euclidean away from the cone points, (local) geodesics on S minus the cone points are straight lines. A geodesic containing a cone point locally consists of two straight line segments meeting at the cone point and forming angles of at least π on both sides. A geodesic between two cone points and with no cone points on its interior is called a *saddle connection*. In particular, a saddle connection σ determines a line in the tangent space at any of its interior points, and therefore determines a point in the space of directions which is denoted by $[\sigma]$. The length of a saddle connection σ is denoted by $\ell(\sigma)$.

2.2.3 Foliations The line field obtained by parallel translating a tangent line around S minus the cone points also corresponds to a foliation of S minus the cone points by geodesics, which extends to a singular foliation over all of S when the cone points are included back in. So for any direction α there is a corresponding (singular) foliation $\mathcal{F}(\alpha)$ in direction α.

It is possible that for particular α the foliation $\mathcal{F}(\alpha)$ defines a *cylinder decomposition* in which every nonsingular leaf is a closed geodesic, and the singular leaves are concatenations of saddle connections which separate S into a union of Euclidean cylinders. A description of some of the directions α for which this occurs is given by Theorem 3.3.

2.2.4 The universal cover Let $\widetilde{S}$ denote the universal cover of S. The complex structure X and quadratic differential q on S can be pulled back to $\widetilde{S}$, and the covering also gives a canonical identification of the directions on S with those on $\widetilde{S}$. The covering map from $\widetilde{S}$ to S sends cone points to cone points (and saddle connections to saddle connections) and is a local isometry. For simplicity, the same notation is used for the structure, quadratic differential (and associated metric), and space of directions for $\widetilde{S}$.

Because every cone point on S has cone angle greater than 2π, both S and $\widetilde{S}$ are nonpositively curved via Gromov's link condition (see, for instance, [6, Chapter II.5]). Therefore the pulled back metric on $\widetilde{S}$ is CAT (0), and so $\widetilde{S}$ is uniquely geodesic. The geodesics are analogous to those in S, consisting of concatenations of saddle connections (with perhaps a straight line segment at the beginning or end of the path). Given a Euclidean cylinder in S, the preimage is a union of strips, and the covering map restricts to a universal covering of the cylinder on each strip.

2.3 The Teichmüller space

See [13, Section 8.2; 20].

Two complex structures X and Y on S are called *equivalent* if there is a map $f : (S, X) \to (S, Y)$, biholomorphic in the coordinate charts, which is isotopic to the identity on S. The *Teichmüller space* of S, denoted $\mathcal{T}(S)$, is the space of equivalence classes of complex structures on S. The notation X is used both for a particular complex structure and its isotopy class $X \in \mathcal{T}(S)$. The Teichmüller space of S comes equipped with a metric known as the Teichmüller metric, which will not appear explicitly in this paper, but whose relevant features are described below.

Denoting by $\dot{S}$ the surface S with a marked point, the space $\mathcal{T}(\dot{S})$ is the space of isotopy classes of complex structures in which isotopies are also required to fix the marked point. A fibration of Teichmüller spaces called the *Bers fibration* is given by

$$\widetilde{S} \to \mathcal{T}(\dot{S}) \to \mathcal{T}(S), \tag{2}$$

obtained by forgetting the marked point, where the fiber over a point $X \in \mathcal{T}(S)$ is canonically identified with $\widetilde{S}$ [2; 18].

2.3.1 The Teichmüller disk Given a flat surface (S, X, q), where q has preferred coordinates ζ_i, a new complex structure can be obtained from any $A \in \mathrm{SL}_2(\mathbb{R})$ by applying A to the given atlas — that is, a new atlas $\{A \circ \zeta_i\}$, where A acts as a linear transformation of $\mathbb{R}^2 \cong \mathbb{C}$. The new complex structure is denoted $A \cdot (X, q) = (A \cdot X, A \cdot q)$. Note that this deformation preserves the zeroes (including their orders) of the original structure (X, q). The map $\mathrm{SO}(2)A \mapsto A \cdot X$ gives a homeomorphism from $\mathrm{SO}(2)\backslash\mathrm{SL}_2(\mathbb{R})$ to the image of the orbit $D \subset \mathcal{T}(S)$ of (X, q), since $\mathrm{SO}(2)$ preserves the underlying complex structure. The disk D is called the *Teichmüller disk* of q, on which the Teichmüller metric is the push-forward of the Poincaré metric (by Teichmüller's theorem). As a consequence,

$$\mathrm{Isom}^+(D) \cong \mathrm{PSL}_2(\mathbb{R}),$$

where the latter is the group of orientation-preserving isometries of the hyperbolic plane. It is often convenient to think of D as the Poincaré disk model of the hyperbolic plane, where each point represents a complex structure on S and traversing a geodesic between two points in D corresponds to varying the underlying flat structure by affine deformations.

More specifically, a geodesic through $X \in D$ is the map $t \mapsto A_t \cdot (X, q)$, where $\{A_t\}_{t\in\mathbb{R}}$ is a symmetric, 1-parameter hyperbolic subgroup of $\mathrm{SL}_2(\mathbb{R})$ — that is, given by a matrix conjugate to

$$\pm\begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}$$

by an element of $\mathrm{SO}(2)$. All geodesics in D can be obtained by $\mathrm{SL}_2(\mathbb{R})$ conjugates of the same 1-parameter family. The unit eigenvectors of such a transformation are orthogonal, leading to an identification of the boundary circle ∂D with the space of directions $\P^1(q)$ by associating the endpoint of the positive ray with the direction of the contracting eigenvector. As the ray approaches the boundary point associated to direction α, the length of any saddle connection in direction α (if one exists) shrinks exponentially. Since this occurs for a ray based at any point and ending at the boundary point associated to direction α, any horocycle based at the boundary point is a level set for the length of a saddle connection in direction α.

2.4 Veech groups

See [13, Chapters 7–8; 16; 30].

For a flat surface (S, X, q) and its associated Teichmüller disk $D \subset \mathcal{T}(S)$, a *Veech group* of q is a subgroup of the stabilizer of D. (Note also that the stabilizer of D is a subgroup of the mapping class group of S.) In this paper, a choice of Veech group of q is denoted by G. Equivalently, a Veech group is a subgroup of the affine homeomorphisms of (S, X, q) in preferred coordinates for q which fix the cone points, projected into the mapping class group.

Any affine homeomorphism in the Veech group has a derivative in preferred coordinates which is well defined up to sign and determines an element of $\mathrm{PSL}_2(\mathbb{R})$, the group of orientation-preserving isometries of D — or, equivalently, of the hyperbolic plane. The action of the Veech group G on D is conjugate to an action on $\mathbb{H}$ via the derivative (an element of $\mathrm{PSL}_2(\mathbb{R})$). Since these two actions are essentially the same, both perspectives are employed throughout this paper. Elements of G are referred to as *parabolic*, *hyperbolic*, or *elliptic* according to whether their images under the derivative homomorphism are parabolic, hyperbolic, or elliptic isometries of $\mathbb{H}$, respectively. Finally, because $\mathrm{MCG}(S)$ acts properly discontinuously on $\mathcal{T}(S)$, a Veech group G acts properly discontinuously on $\mathbb{H}$. Therefore the image of G under the derivative homomorphism is a *Fuchsian group*, that is, a discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$.

In this paper, all Veech groups are assumed to be finitely generated.

2.4.1 Limit points The *limit set* of G, denoted $\Lambda(G)$, is the set of all possible limit points of some G-orbit, Gz, for $z \in D$. A Fuchsian group G is called *nonelementary* if $|\Lambda(G)| > 2$, which implies that G is not virtually cyclic. The limit set $\Lambda(G)$ is contained in the boundary circle ∂D [17, Corollary 2.2.7].

By identifying the boundary circle ∂D with $\P^1(q)$, each $\alpha \in \Lambda(G)$ corresponds to a direction $\alpha \in \P^1(q)$. Throughout this paper, limit points and their associated directions are referred to interchangeably.

A limit point is called a *parabolic fixed point* if it is the fixed point of a parabolic element of G. A limit point $\alpha \in \partial D$ is a *conical limit point* if for any ray R ending at α there is a point $Y \in D$, a sequence $\{g_i\}_{i=1}^{\infty}$ of elements of G, and an $\epsilon > 0$ such that $\{g_i Y\}_{i=1}^{\infty}$ converges to α within the ϵ-neighborhood of R in D [25, p. 617]. Since a finitely generated Veech group is Fuchsian, the following characterization of limit points applies.

Theorem 2.3 [1, Theorem 10.2.5] *A Fuchsian group is finitely generated if and only if each limit point is either a parabolic fixed point or a conical limit point.*

The precise definition of a conical limit point is only necessary for a proof of the generalized Veech dichotomy in Section 3.

2.4.2 Convex hull and convex core The *convex hull of* G, denoted $\text{hull}(G)$, is the intersection of all hyperbolic half planes whose closures in $D \cup \partial D$ contain $\Lambda(G)$. The convex hull is the minimal closed, convex, G-invariant subset of D [25, p. 637]. Consequently, the *convex core* of D/G, denoted $\text{core}(D/G) := \text{hull}(G)/G$, is the smallest closed convex subset of D/G for which the inclusion is a homotopy equivalence. When G is finitely generated, $|\Lambda(G)| \leq 1$ implies that $\text{core}(D/G) = \varnothing$, otherwise $\text{core}(D/G)$ is nonempty [25, p. 637]. Moreover, when the group G is finitely generated, the convex core has finite area, and truncating the convex core along its cusps results in a compact space [1, Theorem 10.1.2; 29, Proposition 8.4.3].

2.4.3 Lattices If the quotient space of D under the action of G has finite area, then the Veech group G is called a *lattice*, and the associated flat surface (S, X, q) is called a *lattice surface*. In particular, if G is a lattice then $\text{hull}(G) = D$.

3 Generalized Veech dichotomy

The Veech dichotomy characterizes foliations on a lattice surface. A generalization to foliations in certain directions on any flat surface follows quickly from several well-known results, but appears to be missing from the literature. The result most relevant to this paper is Corollary 3.9 at the end of this section. Otherwise, this section is self-contained.

Definition 3.1 Let $\mathcal{F}_\alpha$ denote a foliation of (S, X, q) in the direction α. If each of the leaves of $\mathcal{F}_\alpha$ is dense in (S, X, q), then $\mathcal{F}_\alpha$ is called *minimal*; if in addition the transverse measure is unique up to scalar multiplication, then it is called *uniquely ergodic*. If, on the other hand, each of the leaves of $\mathcal{F}_\alpha$ is either closed or a saddle connection, then $\mathcal{F}_\alpha$ is called *completely periodic*.

A saddle connection is considered a leaf, so when $\mathcal{F}_\alpha$ is minimal, there are no saddle connections on S in the direction α. When $\mathcal{F}_\alpha$ is completely periodic, this means that there is a cylinder decomposition of S in the direction α.

Theorem 3.2 (classical Veech dichotomy[1]) *Let (S, X, q) be a lattice surface. Every direction $\alpha \in \partial D$ is either minimal and uniquely ergodic or completely periodic and invariant by a parabolic element of the maximal Veech group.*

When the maximal Veech group G is not a lattice, there may be directions on S which do not correspond to points in the limit set of G. The general statement considers only directions corresponding to limit points for a finitely generated Veech group.

Theorem 3.3 (generalized Veech dichotomy) *Let (S, X, q) be a flat surface with finitely generated Veech group G. Every direction $\alpha \in \Lambda(G) \subset \partial D$ is either minimal and uniquely ergodic or completely periodic and invariant by a parabolic element of the maximal Veech group.*

Remark 3.4 When the Veech group is a lattice, the limit set is all of ∂D. Therefore the classical Veech dichotomy appears as a special case of this theorem.

Definition 3.5 [22, p. 1033] The *Teichmüller geodesic flow* is the one parameter subgroup of $\mathrm{SL}_2(\mathbb{R})$ given by

$$g_t = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}$$

acting on the space of all quadratic differentials. A flat structure (X, q) is called *divergent* if $g_t \cdot (X, q)$ eventually exits every compact set in the moduli space $\mathcal{T}(S)/\mathrm{MCG}(S)$ as $t \to \infty$.

Similar to the proof of the classical Veech dichotomy, the proof of the generalized Veech dichotomy will require Masur's criterion.

Theorem 3.6 (Masur's criterion; [22, Theorem 3.8]) *Let α be the vertical direction. If the foliation $\mathcal{F}_\alpha$ of (S, X, q) is minimal but not uniquely ergodic, then (X, q) is divergent.*

Lemma 3.7 (from proof of [22, Theorem 1.8]) *If there is no saddle connection in direction α, then the foliation $\mathcal{F}_\alpha$ is minimal.*

Lemma 3.8 *Let $\alpha \in \Lambda(G) \subset \partial D$ and let R be a ray in D ending at α. If the image of R under $D \to D/G$ returns to a compact set infinitely often, there is no saddle connection on (S, X, q) in direction α.*

Proof Note that for all $g \in G$ and $(S, X, q) \in D$, $g \cdot (S, X, q)$ is isometric to (S, X, q) — so points of D/G are flat surfaces up to isometry. The length of the shortest saddle connection in direction α is a continuous function on D/G, so over any compact subset of D/G there is a lower bound on the length of the shortest saddle connection in direction α.

Suppose that there is a saddle connection on (S, X, q) in direction α. Without loss of generality assume that α is the vertical direction, so that R is the image of the Teichmüller geodesic flow g_t. Then the length of the shortest saddle connection in the vertical direction tends to zero. Therefore the image of R in D/G cannot return to a compact set of D/G infinitely often. □

[1] This statement is primarily based on [22, Theorem 5.10]. An alternate statement and proof (in the language of flows rather than foliations) is [16, Theorem 1]. Both papers note the proof of [33, Theorem 3.4] as another resource. The original statement is due to [32].

Proof of generalized Veech dichotomy Because G is finitely generated, α is either a conical limit point or parabolic fixed point (Theorem 2.3). When α is a parabolic direction the foliation in direction α is completely periodic; see, for example, [30, Section 6].

Suppose that α is conical, and let R be a ray based at a point $X \in \text{hull}(G)$ and ending at α. Then there exist a sequence $\{g_i\}_{i=1}^{\infty}$ of elements of G and an $\epsilon > 0$ large enough that $\{g_i X\}_{i=1}^{\infty}$ converges to α within the ϵ-neighborhood of R in D. Then R intersects the closed ϵ-neighborhood of X and its images under each of the isometries g_i (i.e., the translates of the neighborhood around X), so R must return to a compact set in D/G infinitely often. By Lemma 3.8 there is no saddle connection in direction α, so the foliation $\mathcal{F}_\alpha$ is minimal (Lemma 3.7). Finally, because $\mathcal{F}_\alpha$ is minimal and X is not divergent, the contrapositive to Masur's criterion (Theorem 3.6) implies that $\mathcal{F}_\alpha$ must be uniquely ergodic. □

Corollary 3.9 *For any saddle connection σ, the associated direction $[\sigma] \in \partial D$ is either parabolic or lies outside of $\Lambda(G)$.*

4 Construction

Let S be a closed, connected, oriented surface of genus at least 2. Fix a complex structure X_0 on S and a flat metric q_0 for X_0 so that (S, X_0, q_0) is a flat surface. Let D be the Teichmüller disk of q_0, let ρ be the Poincaré metric on D, and let G be a finitely generated Veech group of q_0.

See [8] for the motivating construction in the case where G is not only finitely generated but also a lattice. Their construction of analogous spaces E and $\widehat{E}$ differs from the construction given here. However, several of their arguments continue to apply in this more general setting essentially verbatim, and their work is cited wherever this is the case.

The construction is illustrated in Figure 2 on page 1005.

4.1 Base spaces

4.1.1 Horoballs and horopoints Denote by $\mathscr{D}$ the set of all directions of all saddle connections on (S, X_0, q_0). Consider $\mathscr{D}$ to be a subset of ∂D as described in Section 2.2.1. For a saddle connection σ, denote its direction by $[\sigma] \in \mathscr{D}$. The notation $\alpha \in \mathscr{D}$ is sometimes used to refer to a direction without a specific choice of saddle connection.

Because G is assumed to be finitely generated and not necessarily a lattice, the direction of any saddle connection is associated to either a parabolic limit point or to a point outside the limit set of G (Corollary 3.9), which inspires the definitions below.

Definition 4.1 Let σ be a saddle connection with direction $[\sigma] \in \mathscr{D}$.

(i) If $[\sigma]$ is a parabolic limit point, then it is called a *parabolic direction* and σ is called a *parabolic saddle connection*. For each such $[\sigma]$, make a choice of closed horoball which is invariant by the maximal parabolic subgroup of G corresponding to $[\sigma]$. Choose the horoball to be small enough so that in any fiber over a point in its interior, the length of a saddle connection in direction $[\sigma]$ is no more than one third the

length of a saddle connection in any other direction. (This ensures that the horoballs are 1-separated.) Also choose the horoball to be small enough that its distance to $\partial\,\mathrm{hull}\,(G)$ is at least one. Finally, choose the horoballs so that the set of all horoballs (for all parabolic directions) is G-invariant. For each parabolic $[\sigma]$, fix this choice of horoball and call it the *horoball* for $[\sigma]$, denoted $B_{[\sigma]}$.

(ii) If $[\sigma]$ is not a parabolic limit point, then it is called a *nonparabolic direction* and σ is called a *nonparabolic saddle connection*. For each such $[\sigma]$, define the *horopoint* for $[\sigma]$, denoted $B_{[\sigma]}$, to be the ρ-closest orthogonal projection of $[\sigma]$ to $\partial\,\mathrm{hull}\,(G)$ — that is, the unique point $x \in \partial\,\mathrm{hull}\,(G)$ such that the (unique, hyperbolic) geodesic through x and $[\sigma]$ meets $\partial\,\mathrm{hull}\,(G)$ orthogonally. Note that the set of all horopoints (for all nonparabolic directions) is necessarily G-invariant.

See Figure 1 (blue) for examples. The notation B_* is used in both definitions because these objects function analogously later, when it is sometimes helpful to refer to them interchangeably.

The set of horopoints is not necessarily pairwise separated by a fixed constant, although by construction the horopoints are separated by a fixed constant from the set of horoballs associated to parabolic saddle connections. For each $\alpha \in \mathscr{D}$ fix a point $X_\alpha \in \partial B_\alpha$; for nonparabolic directions the only possible choice is $X_\alpha = B_\alpha$, the horopoint for α.

4.1.2 Convex hull and truncated convex hull $(\overline{D}, \bar{\rho})$ Let $\mathrm{hull}\,(G)$ denote the convex hull of the limit set of G as defined in Section 2.4.2. Because $\mathrm{hull}\,(G)$ is a subset of D and (D, ρ) is a length space, the space $(\mathrm{hull}\,(G), \rho|_{\mathrm{hull}(G)})$ is also a length space.

The *truncated convex hull* of G, denoted $\overline{D}$, is obtained from $\mathrm{hull}\,(G)$ by deleting the interiors of the horoballs B_α. That is,

$$\overline{D} := \mathrm{hull}\,(G) \setminus \bigcup_{\alpha\in\mathscr{D}} B_\alpha^\circ,$$

where B_α° denotes the interior of B_α. (Note that for nonparabolic $\alpha \in \mathscr{D}$, B_α is a point and $B_\alpha^\circ = \varnothing$.) By [6, Theorem II.11.27], the truncated disk $D \setminus \bigcup_{\alpha\in\mathscr{D}} B_\alpha^\circ$ with the induced path metric is $\mathrm{CAT}\,(0)$. Therefore the restriction to $\overline{D}$ with the associated restricted metric $\bar{\rho}$ is also a $\mathrm{CAT}\,(0)$ space. Because G is finitely generated, it acts cocompactly on $\overline{D}$ (refer to Section 2.4.2).

Denote by $\mathrm{proj}_{\overline{D}} : D \to \overline{D}$ the ρ-closest-point projection of D onto $\overline{D}$.

4.1.3 Collapsed convex hull $(\widehat{D}, \hat{\rho})$ Let $\widehat{D}$ be the *collapsed convex hull*, the quotient space obtained from $\overline{D}$ by collapsing the boundary of each horoball B_α to a point. The quotient pseudometric $\hat{\rho}$ is positive-definite, therefore a metric, and so the space $(\widehat{D}, \hat{\rho})$ is a length space (Lemma 2.1).

Analogous to the map $\bar{p} : \overline{D} \to \widehat{D}$ which collapses each horoball to a point, there is also a map $p : \mathrm{hull}\,(G) \to \widehat{D}$ which collapses the *interior* of each horoball to a point.

4.2 Bundles

4.2.1 Total space E Let $\pi : E \to \mathrm{hull}\,(G)$ be the pullback bundle of the Bers fibration (2) via the inclusion $\mathrm{hull}\,(G) \subset D \subset \mathcal{T}(S)$, identifying $E \subset \mathcal{T}(\dot{S})$. Let $\Gamma < \mathrm{MCG}(\dot{S})$ be the $\pi_1 S$-extension of G (that is, the group fitting into the Birman exact sequence (1)), which acts on E.

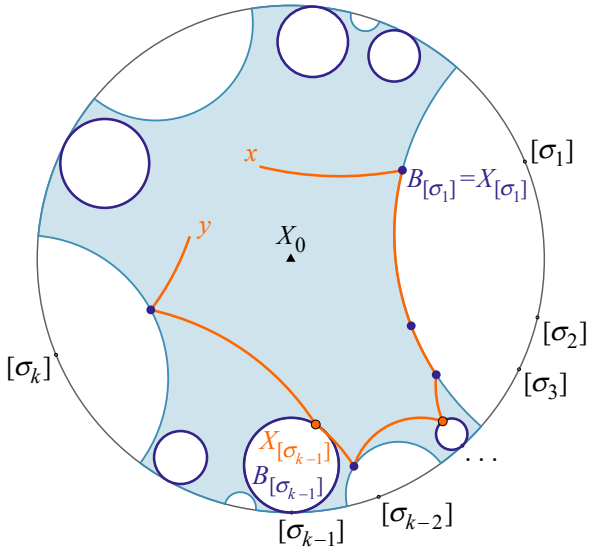


Figure 1: Examples of the constructions from Definitions 4.1 and 5.2, projected to D. The region shaded in blue, *including* its boundary in D, is the truncated convex hull $\overline{D}$. Some saddle connection directions are labeled along ∂D. For nonparabolic directions $[\sigma_*]$ — that is, those outside hull (G) — the associated "horopoint" $B_{[\sigma_*]}$ (blue dot) is the ρ-closest projection of $[\sigma_*]$ onto ∂ hull (G). For the remaining parabolic directions $[\sigma_*]$, the associated "horoball" $B_{[\sigma_*]}$ (blue circle and its interior) is a closed horoball in hull (G) based at $[\sigma_*]$, with the choice of horoball made according to the technical requirements of Definition 4.1. Finally, for each σ_* there is a choice of point $X_{[\sigma_*]}$: When σ_* is nonparabolic $X_{[\sigma_*]} := B_{[\sigma_*]}$ (blue dot), and when σ_* is parabolic $X_{[\sigma_*]} \in \partial B_{[\sigma_*]}$ (orange dot).

To construct the preferred path from x to y, $\varsigma(x, y)$, as in Definition 5.2, first observe that the geodesic between $f(x)$ and $f(y)$ in the fiber over X_0 (black triangle) consists of saddle connections σ_1 through σ_k, in order. The fact that their associated directions appear in a nice order along ∂D is not a coincidence; see Lemma 6.19. The preferred path is constructed from concatenations of hyperbolic geodesics through the horizontal fibers of $\overline{E}$ (orange curves) with saddle connections traversed in the fibers over the chosen points X_* (blue dots or orange dots). Because this diagram shows only the projection of $\varsigma(x, y)$ to D, it is important to keep in mind that the horizontal pieces (orange curves) all belong to different horizontal fibers in $\overline{E}$ and that the saddle connections (blue dots or orange dots) in fact have positive length. The saddle connections associated to parabolic directions (orange dots only) will have zero length after collapsing the bundle as in Section 4.3.2.

The space E is a surface bundle over hull (G): By construction the fiber over $X \in$ hull (G), denoted $E_X := \pi^{-1}(X)$, is canonically identified with $\widetilde{S}$ equipped with the pulled back complex structure X and flat metric q_X. Recall that there was a fixed choice of a complex structure X_0 (which is necessarily in hull (G)), and denote the fiber over X_0 by E_0.

It will be necessary to have maps between fibers. For $X, Y \in$ hull (G), let $f_{X,Y} : E_Y \to E_X$ be the lift of the Teichmüller map, i.e., the map that sends $y \in E_Y$ to the unique point $f(y) \in E_X$ along the lift of the geodesic in D connecting $\pi(x)$ and $\pi(y)$. This map is affine with respect to the flat metrics q_X and q_Y

and is $e^{\rho(X,Y)}$-bilipschitz by construction. For any $X, Y, Z \in \operatorname{hull}(G)$, the composition $f_{X,Y} \circ f_{Y,Z}$ agrees with $f_{X,Z}$. For any $X \in \operatorname{hull}(G)$, define $f_X : E \to E_X$ by $f_X|_{E_Y} = f_{X,Y}$ for $Y \in \operatorname{hull}(G)$. When $X = X_0$, the map is denoted $f = f_{X_0} : E \to E_0$.

In fact, E is a product: It admits a product structure $E \cong \operatorname{hull}(G) \times \widetilde{S}$. The map π is the projection onto the first factor $\operatorname{hull}(G)$. The projection onto $\widetilde{S}$, the universal cover of S, is any map f_X for $X \in \operatorname{hull}(G)$. Define a metric d on E as the orthogonal direct sum of the Poincaré metric ρ in each horizontal fiber D_x and the pulled back flat metric q_X in each fiber $E_X \cong \widetilde{S}$.

For any $X \in \operatorname{hull}(G)$, denote by $\Sigma_X \subset E_X$ the set of cone points of the flat structure q_X on E_X, and define

$$\Sigma = \bigcup_{X \in \operatorname{hull}(G)} \Sigma_X.$$

The quotient E/Γ is generally a noncompact S-bundle over the convex core $\operatorname{hull}(G)/G$. If the quotient $\operatorname{hull}(G)/G$ is compact and hence E/Γ is compact, then this represents a special case in which there are no parabolic directions for G — that is, the Veech group G is convex cocompact and therefore the extension group Γ is hyperbolic (see [11; 15] or Theorem 6.29). In the general case, a Γ-equivariant quotient of E is constructed from the G-equivariant quotient $\operatorname{hull}(G) \to \widehat{D}$ as described in the following several subsections.

4.2.2 Horizontal fibers D_x For any $X \in \operatorname{hull}(G)$ and $x \in E_X$, the horizontal fiber $D_x = f_X^{-1}(x)$ is the unique lift of $\operatorname{hull}(G)$ to the Teichmüller disk in $\mathcal{T}(\dot{S})$ through x that covers $\operatorname{hull}(G)$ via the projection π.

Remark 4.2 A more appropriate choice of notation might be $\operatorname{hull}(G)_x$, to emphasize that this is a lift of only the convex hull rather than the entire Teichmüller disk D. Because there will not be any objects lifted to the complement of the hull, the notation D_x is used for simplicity.

4.2.3 Pullback bundle over truncated convex hull $\overline{E}$ Define the pullback bundle over the truncated convex hull by

$$\overline{E} = \pi^{-1}(\overline{D}) = E \setminus \bigcup_{\alpha \in \mathscr{D}} \mathcal{B}_\alpha^\circ,$$

where $\mathcal{B}_\alpha^\circ = \pi^{-1}(B_\alpha^\circ)$ is the interior of the horoball preimage $\mathcal{B}_\alpha = \pi^{-1}(B_\alpha)$. (As before, if α is nonparabolic then B_α is a point and $B_\alpha^\circ = \varnothing$, so $\mathcal{B}_\alpha^\circ = \varnothing$ also.)

The truncated bundle $\overline{E}$ is equipped with the length metric $\bar{d}$ induced from (E, d). Because G acts isometrically and cocompactly on $(\overline{D}, \bar{\rho})$ and $\pi_1(S)$ acts isometrically and cocompactly on $(\widetilde{S}, \widetilde{q_0})$, Γ acts isometrically and cocompactly on $(\overline{E}, \bar{d})$. Therefore the space $(\overline{E}, \bar{d})$ is quasi-isometric to the group Γ equipped with the word metric. Then the quotient $\overline{E}/\Gamma$ is a compact S-bundle over $\overline{D}/G$,

$$S \to \overline{E}/\Gamma \to \overline{D}/G.$$

As in [8], compactness of $\overline{E}/\Gamma$ gives that Σ is r-dense in $\overline{E}$ for some $r > 0$. By an application of the Arzelà–Ascoli theorem, $(\overline{E}, \bar{d})$ is a geodesic space.

4.3 Electrified space $\widehat{E}$

4.3.1 Bass–Serre trees Let α be a parabolic direction. The $\mathbb{R}$-tree dual to the foliation of E_{X_α} in the direction α, denoted T_α, is a weighted Bass–Serre tree when equipped with the metric defined by the transverse measure on the foliation of E_{X_α} in direction α. (See, for instance, the construction in [3, Subection 4.4].) Define a map

$$t_\alpha : E \to T_\alpha$$

to be the composition of $f_{X_\alpha} : E \to E_{X_\alpha}$ followed by the projection $E_{X_\alpha} \to T_\alpha$. Recall from Section 4.1.1 that there was a choice of fixed point $X_\alpha \in \partial B_\alpha$, and note that the tree T_α is independent of the choice of $X_\alpha \in \partial B_\alpha$. Therefore the map t_α is also independent of this choice. Because there are only finitely many Γ-orbits of edges of these trees, every edge of every tree has length uniformly bounded above and below.

4.3.2 Collapsed "bundle" $\widehat{E}$ Define $\widehat{E}$ to be the quotient of E obtained by collapsing each $\mathcal{B}_\alpha$ onto T_α via the restriction $t_\alpha|_{\mathcal{B}_\alpha}$, and denote this map by $P : E \to \widehat{E}$. Note that the quotient is Γ-equivariant because it is constructed from the G-equivariant quotient $\text{hull}(G) \to \widehat{D}$. Finally, $\widehat{E}$ is equipped with the quotient pseudometric $\hat{d}$ obtained from $\bar{d}$ by the restriction $\overline{P} := P|_{\overline{E}}$. As in [8, Lemma 3.2], $\hat{d}$ is positive-definite, therefore a metric. Then by Lemma 2.1, the space $(\widehat{E}, \hat{d})$ is a length space. The following fact is used throughout the paper.

Lemma 4.3 [8, Lemma 3.3] *The map P is 1-Lipschitz.*

4.4 Illustration of the construction

Figure 2 is a modified copy of [8, Figure 1], which illustrates the construction of $\widehat{E}$ when the Veech group G is not a lattice. The base is a copy of the Teichmüller disk for the flat structure (S, X_0), which is isometric to the hyperbolic plane. Over each point in the base disk is a copy of the universal cover of S with the lift of the associated flat metric. The truncated convex hull $\overline{D}$ is shaded. The structures illustrated are:

(a) Some points of ∂D associated to nonparabolic directions. For each nonparabolic direction η, the horopoint B_η is the nearest-point projection of η to $\text{hull}(G)$. The point X_η is chosen to be identical to B_η. The fiber over X_η is quasi-isometric to E_0, the fiber over X_0.

(b) Some horoballs based at points of ∂D associated to parabolic directions. For each parabolic direction α, the associated horoball B_α is a sublevel set for the length of a saddle connection in direction α in a fiber over a point of B_α. The point X_α is a fixed choice in ∂B_α. By construction, the horoballs are pairwise 1-separated and also 1-separated from the set of horopoints associated to nonparabolic directions.

(c) For any parabolic direction β, there is a multitwist $\tau_\beta \in G$ which takes X_β to other points in ∂B_β. As τ_β moves X_β along ∂B_β in D, it induces shearing in the fibers, which can be seen in E_{X_β} and $E_{\tau_\beta(X_\beta)}$.

(d) The fiber E_0 over the basepoint X_0 is the universal cover of (S, X_0). The highlighted path γ is a flat geodesic between cone points, which consists of concatenated saddle connections.

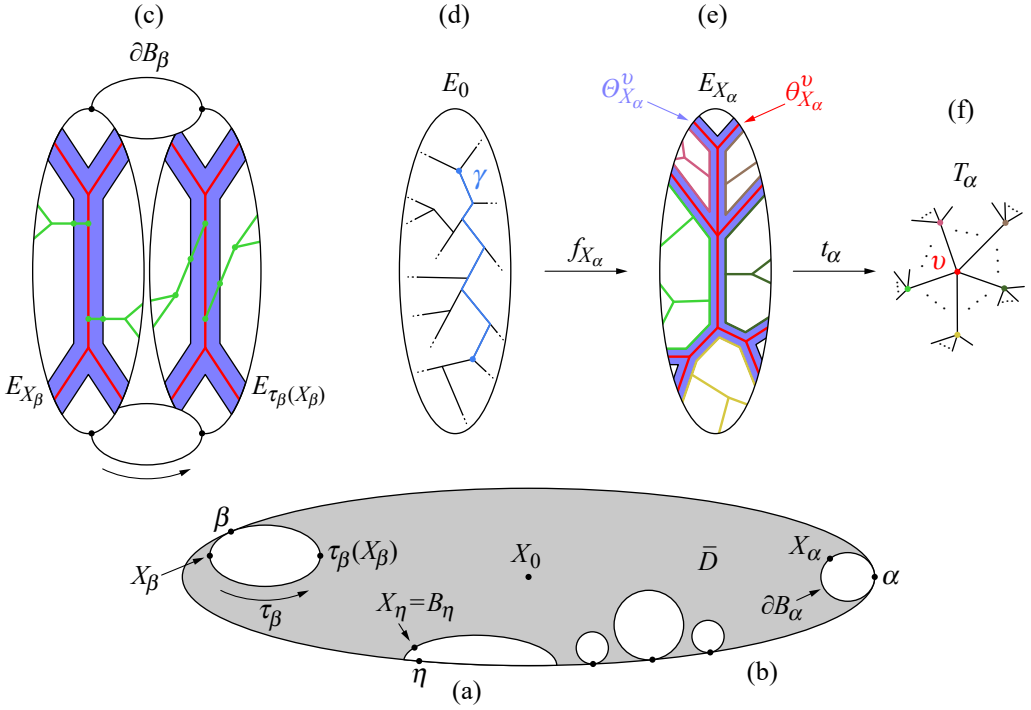


Figure 2: A copy of [8, Figure 1] which has been modified to reflect the case in which the Veech group G is not a lattice.

(e) The fiber E_{X_α} over X_α is the universal cover of (S, X_α). It can also be obtained from E_0 by mapping via f_{X_α}, the lift of the Teichmüller map between X_0 and X_α in D. Each collection of colored segments, for example $\theta^v_{X_\alpha}$ in red, represents a connected set of saddle connections in direction α. These are called spines; see Section 5.2. (The lavender region $\Theta^v_{X_\alpha}$ is the "thickened spine neighborhood", which only appears in this paper once when referring to results from [9].) The complement of the saddle connections is a cylinder decomposition of E_{X_α} in the direction α.

(f) The Bass–Serre tree T_α associated to the direction α. Each vertex is associated to a spine in E_{X_α}. If two spines in E_{X_α} bound a cylinder, their associated vertices in T_α are connected by an edge with length corresponding to the transverse measure of the cylinder.

5 Guessing geodesics

Hyperbolicity of $\widehat{E}$ will be proven using the "guessing geodesics" criterion of Bowditch [5, Proposition 3.1] and Masur–Schleimer [21, Theorem 3.15], as formulated by [8].

Proposition 5.1 (guessing geodesics; [8, Proposition 2.2]) *Suppose Ω is a length space, $\Upsilon \subset \Omega$ an R-dense subset for some $R > 0$, and $\delta \geq 0$ a constant such that for all pairs $x, y \in \Upsilon$ there are rectifiably path-connected sets $L(x, y) \subset \Omega$ containing x, y satisfying these properties*:

(1) *The $L(x, y)$ form δ-slim triangles.*

(2) *If $x, y \in \Upsilon$ have $\hat{d}(x, y) \leq 3R$, then the diameter of $L(x, y)$ is at most δ.*

Then Ω is hyperbolic.

Saying the $L(x, y)$ form δ-slim triangles means that for all $x, y, z \in \Upsilon$,

$$L(x, y) \subset N_\delta \left(L(x, z) \cup L(z, y) \right),$$

where N_δ denotes the δ-neighborhood. This is similar to the usual slim triangles condition for hyperbolicity from Section 2.1.3, except that the $L(x, y)$ need not be geodesics.

This section will verify most of the hypotheses required to apply the guessing geodesics criterion. Many of the arguments made by [8] continue to hold in this case. However, the proof that the sets $L(x, y)$ (constructed in Section 5.1) form slim triangles differs considerably and composes the entirety of Section 6.

5.1 Preferred paths and collapsed preferred paths

The sets $L(u, v)$ in the guessing geodesics criterion will be constructed from the following preferred paths in E which connect cone points by concatenating geodesics in the horizontal and vertical fibers.

Definition 5.2 For two cone points $x, y \in \Sigma$ the *preferred path* from x to y, denoted $\varsigma(x, y)$, is constructed as follows.

(1) In E_0, $f(x)$ and $f(y)$ are cone points with respect to the flat metric q_0, and they are connected by a geodesic segment which is a concatenation of saddle connections in E_0. That is,

$$[f(x), f(y)] = \sigma_1 \sigma_2 \cdots \sigma_k.$$

Each saddle connection σ_i is considered to be oriented with initial point $\sigma_i^- := \sigma_i \cap \sigma_{i-1}$ for $i = 2, \ldots, k$ (and $\sigma_1^- = f(x)$) and terminal point $\sigma_i^+ := \sigma_i \cap \sigma_{i+1}$ for $i = 1, \ldots, k-1$ (and $\sigma_k^+ = f(y)$). In the preferred path, each saddle connection σ_i will be traversed in the fiber over the associated point $X_{[\sigma_i]} \in B_{[\sigma_i]}$ as fixed in Section 4.1.1. More precisely, define the i-th *saddle piece* to be

$$\gamma_i := f_{X_{[\sigma_i]}}(\sigma_i),$$

which is a segment of the preferred path for all $1 \leq i \leq k$. Each γ_i is considered to be oriented with initial point

$$\gamma_i^- := f_{X_{[\sigma_i]}}\left(\sigma_i^-\right)$$

and terminal point

$$\gamma_i^+ := f_{X_{[\sigma_i]}}\left(\sigma_i^+\right).$$

(2) The *horizontal pieces* h_i are chosen to make the preferred path continuous. Specifically, for $1 \leq i \leq k-1$, h_i is the geodesic in $D_{\gamma_i^+} = D_{\gamma_{i+1}^-}$ which connects γ_i^+ to γ_{i+1}^-. The first horizontal piece, denoted h_0, is the geodesic in $D_x = D_{\gamma_1^-}$ connecting x to γ_1^-. The last horizontal piece, denoted h_k, is the geodesic in $D_{\gamma_k^+} = D_y$ connecting γ_k^+ to y.

(3) Define

$$\varsigma(x, y) := h_0\gamma_1 h_1\gamma_2 h_2 \cdots \gamma_k h_k.$$

See Figure 1 for an example. The *collapsed preferred paths*, denoted $\hat{\varsigma}(x, y)$, are the images of the preferred paths $\varsigma(x, y)$ under the map $P : E \to \widehat{E}$; see Section 4.3.2.

5.2 The sets $L(u, v)$

Let $\mathcal{V}$ be the set of all vertices of all Bass–Serre trees in $\widehat{E}$. Given $X \in \text{hull}(G)$ and a saddle connection σ, the union of the saddle connections in E_X with direction $[\sigma]$ is precisely the preimage of the vertices of the Bass–Serre tree $T_{[\sigma]}$ under the map

$$E_X \xrightarrow{f_{X_{[\sigma]},X}} E_{X_{[\sigma]}} \xrightarrow{P} T_{[\sigma]}, \tag{3}$$

which is simply the intersection of E_X with the preimage of $T_{[\sigma]}$ under $t_{[\sigma]}$ (see Section 4.3.1). For any vertex $v \in \mathcal{V}$ and $X \in \text{hull}(G)$, the preimage of v under this composition (3) is called the *v-spine in E_X* and is denoted θ^v_X. Note that a choice of $v \in \mathcal{V}$ determines the Bass–Serre tree $T_{[\sigma]}$ to which v belongs, and denote the union of the v-spines over all $X \in \partial B_{[\sigma]}$ by θ^v.

Given $u, v \in \mathcal{V}$, define

$$L(u, v) = \bigcup \hat{\varsigma}(x, y),$$

where the union is taken over all $x \in \theta^u \cap \Sigma$ and $y \in \theta^v \cap \Sigma$. Since $\varsigma(x, y)$ is a finite length path (by construction, each saddle piece and horizontal piece has finite length), P is 1-Lipschitz (Lemma 4.3), and all $\hat{\varsigma}(x, y)$ in $L(u, v)$ connect u to v, it follows that the set $L(u, v)$ is a path connected, rectifiable set containing both u and v.

5.3 Hyperbolicity

The statement of the following theorem is identical to [8, Theorem 4.2], but the proof depends on results which require considerable reformulation in the case that G is only finitely generated.

Theorem 6.28 (collapsed preferred paths form slim triangles) *There exists $\delta > 0$ so that collapsed preferred paths form δ-slim triangles. That is, for any $x, y, z \in \Sigma$,*

$$\hat{\varsigma}(x, y) \subset N_\delta\left(\hat{\varsigma}(x, z) \cup \hat{\varsigma}(y, z)\right).$$

As the numbering suggests, the entirety of Section 6 is dedicated to the proof of this theorem. For now, Theorem 6.28 is assumed and used to prove hyperbolicity of $\widehat{E}$ in Theorem 5.7. Besides this, the last result needed is Corollary 5.6, which is stated after a few more constructions below.

Definition 5.3, Lemma 5.4, and Claim 5.5 do not reappear in this paper until Section 7; since the proofs are purely technical rather than geometrically intuitive, the reader may want to skip these on a first pass.

Definition 5.3 A *horizontal jump* in $\widehat{E}$ is the image under $\overline{P}$ of a geodesic in $\overline{D}_z$, for some $z \in \Sigma$, that either is a subsegment of a component of $\partial\overline{D}_z$ or connects two components of $\partial\overline{D}_z$ with interior

disjoint from $\partial\overline{D}_z$. A *combinatorial path* in $\widehat{E}$ is a concatenation of horizontal jumps and (perhaps zero) nonparabolic saddle connections; in particular, it is a concatenation of the collapsed preferred paths from Definition 5.2.

When the Veech group is a lattice, horizontal jumps and combinatorial paths coincide with the objects of the same names in [8, Definition 3.7]. Both definitions have been changed in this paper to fit the next statement, which is adapted from [8, Lemma 3.8]; their argument is sketched here and supplemented as necessary.

Lemma 5.4 *There is a constant $C > 0$ such that any pair of points $x, y \in \mathcal{V}$ may be connected by a combinatorial path of length at most $C\hat{d}(x, y)$. In particular, there is a constant $C' > 0$ such that this combinatorial path consists of at most $C'\hat{d}(x, y)$ horizontal jumps and saddle connections.*

Proof sketch Define $\Sigma_\alpha := \Sigma \cap \partial\mathcal{B}_\alpha$. There exists some $K \geq 30$ so that for each $\alpha \in \mathscr{D}$ and $z \in \partial\mathcal{B}_\alpha$ there is $w \in \Sigma_\alpha$ with $\bar{d}(z, w) \leq K/30$ (in particular, K is chosen to be $30M$, where M is from [8, Lemma 3.4(3)]). The next step in the argument is to describe how to traverse between points of $\overline{P}(\Sigma_\alpha)$. Recall from Section 4.3.1 that T_α is the $\mathbb{R}$-tree dual to the foliation of E_{X_α}, and denote the length metric on T_α by ℓ_α. The statement of the following claim is identical to [8, Claim 3.9]; this is the only part of the argument for which the proof does not extend verbatim to the nonlattice case.

Claim 5.5 *There exists $R' > 0$ such that any pair of points $v_1, v_2 \in \overline{P}(\Sigma_\alpha) \subset \mathcal{V}$ may be connected by a combinatorial path of length at most $R'\ell_\alpha(v_1, v_2)$.*

Proof It suffices to assume that v_1 and v_2 are adjacent vertices of T_α. Choosing any $X \in \partial B_\alpha$, the preimages of v_1 and v_2 under the restriction $\overline{P}|_{E_X} : E_X \to T_\alpha$ are adjacent spines $\theta_X^{v_1}$ and $\theta_X^{v_2}$ which are separated by a strip of uniformly bounded width [8, Lemma 3.4]. Then there exists a saddle connection $\sigma \subset E_X$ of bounded length which joins cone points $y_1 \in \theta_X^{v_1}$ and $y_2 \in \theta_X^{v_2}$. Denote the direction of σ by $[\sigma] \in \mathscr{D}$; unlike the case of [8], this direction may be nonparabolic.

Because σ has bounded length, $X \in \partial B_\alpha$ lies within a bounded distance of $Y \in \partial B_{[\sigma]}$. Let $z_i := f_{Y,X}(y_i)$. Let h_i be the horizontal geodesic in D_{y_i} from y_i to z_i, and note that the $\overline{P}$-image of each component of $h_i \cap \overline{D}_{y_i}$ is a horizontal jump in $\widehat{E}$. In particular, each $h_i \cap \overline{D}_{y_i}$ must have total length at most $\rho(X, Y)$, which is uniformly bounded. The saddle connection $f_{Y,X}(\sigma) \subset E_Y$ has length bounded by $e^{\rho(X,Y)}$ times the length of σ, which are both bounded. Therefore following the horizontal jumps along h_1, traversing $f_{Y,X}(\sigma)$, and then following the horizontal jumps along h_2 gives a bounded length combinatorial path in $\widehat{E}$ from v_1 to v_2. The claim follows from the fact that an edge in T_α must be at least some minimal length, so $\ell_\alpha(v_1, v_2)$ is uniformly bounded below. □

Note that there exists a constant R_0 so that $\mathcal{V}$ is R_0-dense in $\widehat{E}$ because $\mathcal{V}$ is Γ-invariant and $\widehat{E}/\Gamma$ is compact (since Γ acts cocompactly on $\overline{E}$ and $\widehat{E}/\Gamma$ is the continuous image of $\overline{E}/\Gamma$ under the descent of $P : \overline{E} \to \widehat{E}$; [8, Lemma 3.6]).

Now it can be shown that if $x, y \in \mathcal{V}$ and $r = \hat{d}(x, y) > 0$, then x and y can be connected by a combinatorial path of length $\leq 4R'e^{Kr}r$ by [8, Claim 3.10]. Set $C := 36R'e^{3KR_0}$. Then, if $r := \hat{d}(x, y) \leq 3R_0$,

there is a combinatorial path joining x and y of length at most

$$4R'e^{3KR_0}\hat{d}(x,y) = \tfrac{1}{9}C\hat{d}(x,y),$$

satisfying the lemma. Otherwise if $r > 3R_0$, x and y can be joined by a path γ of length at most $2r$, which can be subdivided into $n = \lceil \text{length}(\gamma)/R_0 \rceil$ equal-length subsegments of length at most R_0. Since $\mathcal{V}$ is R_0-dense, there is a sequence $\{x_i\} \subset \mathcal{V}$ with $x_0 = x$, $x_n = y$, and $\hat{d}(x_i, x_{i+1})$. Each pair x_i, x_{i+1} can then be connected by a combinatorial path of length at most $CR_0/3$, and so there is a combinatorial path from x to y of length at most $CR_0n/3$. Finally, because $R_0n < \text{length}(\gamma) + R_0 \leq 3\hat{d}(x,y)$, the first statement of the lemma holds.

The number of horizontal jumps and saddle connections in the combinatorial path joining x and y can be bounded as follows. Each horizontal jump between different components of the hull boundary (see Definition 5.3) has length at least 1 because the horoballs associated to parabolic directions were constructed to be 1-separated from each other and from $\partial\,\text{hull}\,(G)$ (see Definition 4.1). Therefore the combinatorial path may contain at most $C\hat{d}(x,y)$ horizontal jumps between different components of the hull boundary. By Proposition 6.3, there are no short saddle connections in the fibers over $\overline{D}$; supposing that $A' < 1$ is a uniform positive lower bound on the lengths of the saddle connections, the combinatorial path may contain at most $2C\hat{d}(x,y)/A'$ nonparabolic saddle connections and horizontal jumps which are subsegments of components of the hull boundary. Then the combinatorial path joining x and y contains at most $3C\hat{d}(x,y)/A'$ horizontal jumps and saddle connections. Write $C' := 3C/A'$ to simplify the notation. □

Combined with Theorem 6.28, this bound on $\hat{d}(x,y)$ allows for a bound on the diameter of the sets $L(u,v)$. The following statement is identical to [8, Lemma 4.4]; the proof is nearly identical except for some changes in the language regarding combinatorial paths, which were given a different definition in this paper (see Definition 5.3 above).

Corollary 5.6 *There exists a constant $C > 0$ so that if $u, v \in \mathcal{V}$ with $\hat{d}(u,v) \leq 3R_0$ (where R_0 is a constant so that $\mathcal{V}$ is R_0-dense in $\widehat{E}$, as in the proof of Claim 5.5), then* $\text{diam}(L(u,v)) \leq C$.

Proof Pick $x \in \theta^u \cap \Sigma$ and $y \in \theta^v \cap \Sigma$. By Lemma 5.4 there is a combinatorial path from $P(x) = u$ to $P(y) = v$ of bounded length which is a concatenation of n horizontal jumps and saddle connections, where the bound on the length and the number n depend only on R_0. By repeatedly applying Theorem 6.28, $\hat{\varsigma}(x,y)$ is in the $n\delta$-neighborhood of the combinatorial path joining $P(x) = u$ to $P(y) = v$. Therefore $\hat{\varsigma}(x,y)$ has uniformly bounded diameter. As in [8, Lemma 4.3], $L(u,v)$ is contained in the 2δ-neighborhood of $\hat{\varsigma}(x,y)$, where $\delta > 0$ is the constant from Theorem 6.28. Therefore $L(u,v)$ has uniformly bounded diameter also. □

Finally, it is possible to establish hyperbolicity of $\widehat{E}$.

Theorem 5.7 *The space $(\widehat{E}, \hat{d})$ is hyperbolic when the limit set $\Lambda(G)$ contains at least one parabolic point.*

Proof Hyperbolicity of $\widehat{E}$ will be proven by the guessing geodesics criterion (Proposition 5.1). It was established during construction that $(\widehat{E}, \hat{d})$ is a length space (see Section 4.3.2 and [8, Lemma 3.2]). Let $\mathcal{V} \subset \widehat{E}$ be the collection of all vertices of Bass–Serre trees in $\widehat{E}$. Let $R_0 > 0$ be as in the proof of Claim 5.5, so that $\mathcal{V}$ is R_0-dense in $\widehat{E}$. For any $u, v \in \mathcal{V}$, $L(u, v)$ is a rectifiably path-connected set containing u and v (see Section 5.2). Let $\delta > 0$ be the constant from Theorem 6.28. As in [8, Lemma 4.3], each $L(u, v)$ is contained in the 2δ-neighborhood of $\hat{\varsigma}(x, y)$ for any $x \in \theta^u \cap \Sigma$ and $y \in \theta^v \cap \Sigma$; because triangles of collapsed preferred paths in $\widehat{E}$ are δ-slim (Theorem 6.28), the $L(u, v)$ also form 3δ-slim triangles, satisfying condition (1) of the guessing geodesics criterion. Finally, by Corollary 5.6 the sets $L(u, v)$ have diameter bounded by a constant $C > 0$ whenever $\hat{d}(u, v) \leq 3R_0$, satisfying condition (2) of the guessing geodesics criterion. Therefore $(\widehat{E}, \hat{d})$ is hyperbolic. □

The action of Γ on the total space E is isometric by construction but noncocompact in general. The truncated convex hull $\overline{E}$ was constructed as a subspace on which Γ acts cocompactly in addition to isometrically. By a direct application of the Schwarz–Milnor lemma, $\overline{E}$ and Γ are quasi-isometric. However, neither is hyperbolic unless there are no parabolic directions on S. To construct a hyperbolic space from E, for parabolic saddle connections σ the sets $\mathcal{B}_{[\sigma]}$ are collapsed to Bass–Serre trees. In Γ, this corresponds to coning off by the *vertex subgroups*, the stabilizers of the vertices of the Bass–Serre trees under the isometric action of Γ. Then when the limit set $\Lambda(G)$ contains at least one parabolic point, the main theorem follows. The proof is essentially identical to that of [8, Corollary 4.5], but it is included here for the sake of completeness.

Theorem 5.8 *Suppose $G < \mathrm{MCG}(S)$ is a finitely generated, nonelementary Veech group with extension group Γ and let $\Upsilon_1, \dots, \Upsilon_k < \Gamma$ be representatives of the conjugacy classes of vertex subgroups. Then Γ admits an isometric action on Gromov hyperbolic space $\widehat{E}$, quasi-isometric to the Cayley graph of Γ coned off along the cosets of $\Upsilon_1, \dots, \Upsilon_k$.*

Proof Recall from Section 4.3.2 that $\widehat{E}$ is a length space. By construction, the action of Γ on $\overline{E}$ is isometric and cocompact, and therefore so is the action of Γ on $\widehat{E}$. Let $\Upsilon_1, \dots, \Upsilon_k < \Gamma$ be representatives of the conjugacy classes of vertex subgroups stabilizing the vertices of the Bass–Serre trees. Any point-stabilizers for the action of Γ on $\widehat{E}$ are trivial or conjugate into one of the Υ_*, and therefore any point in $\widehat{E}$ has a discrete orbit under Γ. By Theorem 2.2 it follows that, for any finite generating set $\mathcal{S}$ of Γ, the coned-off Cayley graph $\mathrm{Cay}(\Gamma, \mathcal{S} \cup \bigcup \Upsilon_i)$ is quasi-isometric to $\widehat{E}$. Finally, because $\widehat{E}$ is hyperbolic by Theorem 5.7, so too is the coned-off Cayley graph of Γ. □

6 Slim triangles of collapsed preferred paths

Section 5 verified all but one condition for the guessing geodesics criterion (Proposition 5.1). This section will verify the remaining condition, which is that all triangles of collapsed preferred paths in $\widehat{E}$ are slim (Theorem 6.28).

The following definitions are comparable to those in [8, Section 4.3].

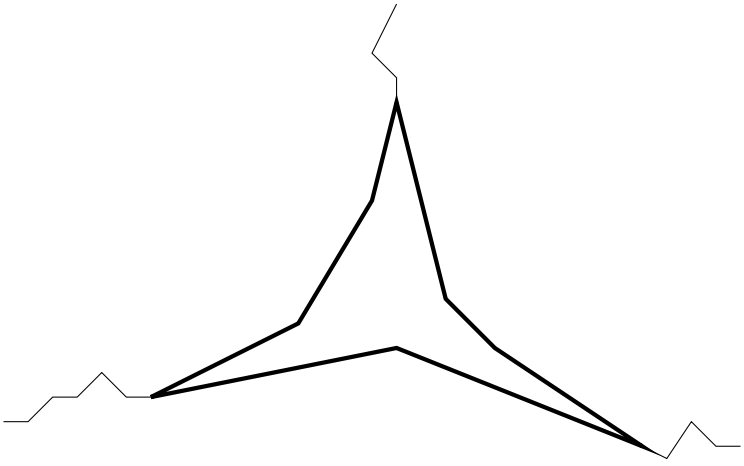


Figure 3: Illustration from [8, Figure 2] of a degenerate triangle in E_0 with the nondegenerate subtriangle in bold.

Definition 6.1 Given $x, y, z \in \Sigma$, define the associated geodesic triangle in E_0 by

$$\Delta(x, y, z) := [f(x), f(y)] \cup [f(y), f(z)] \cup [f(z), f(x)],$$

the triangle of preferred paths in E by

$$\Delta^{\varsigma}(x, y, z) := [\varsigma(x, y)] \cup [\varsigma(y, z)] \cup [\varsigma(z, x)],$$

and the triangle of collapsed preferred paths in $\widehat{E}$ by

$$\Delta^{\hat{\varsigma}}(x, y, z) := [\hat{\varsigma}(x, y)] \cup [\hat{\varsigma}(y, z)] \cup [\hat{\varsigma}(z, x)].$$

Equivalently, $\Delta^{\hat{\varsigma}}(x, y, z)$ is the image of $\Delta^{\varsigma}(x, y, z)$ under the map P. If any pair of the sides of $\Delta(x, y, z)$ intersect at one or more (nontrivial) saddle connections, then $\Delta(x, y, z)$ is called *degenerate*. Otherwise, if each pair of sides of $\Delta(x, y, z)$ intersect only at the vertices, then $\Delta(x, y, z)$ is called *nondegenerate*. The triangles $\Delta^{\varsigma}(x, y, z)$ and $\Delta^{\hat{\varsigma}}(x, y, z)$ are described as degenerate (or nondegenerate) if the triangle $\Delta(x, y, z)$ is degenerate (or nondegenerate).

Recall from the construction that $E_0 \cong \widetilde{S}$ (equipped with the lift of the flat metric q_0). Since E_0 is complete, simply connected, and nonpositively curved, it is CAT (0) (as a result of the Cartan–Hadamard theorem) and therefore it is uniquely geodesic. Then the nontrivial intersection of two sides of a degenerate triangle must be a concatenation of saddle connections based at the shared vertex. As a consequence, every degenerate triangle — whether in E_0, E, or $\widehat{E}$ — contains a subtriangle which is nondegenerate. See Figure 3.

To prove that all triangles are slim, it suffices to prove that all nondegenerate triangles are slim (see [8, Lemma 4.6] for a detailed proof). Therefore the rest of this section assumes that all triangles are nondegenerate.

6.1 Euclidean triangles

Definition 6.2 Given $x, y, z \in \Sigma \cap E_0$, the triangle $\Delta(x, y, z)$ is called *Euclidean* if it is nondegenerate and each side consists of a single saddle connection. When the information about x, y, z is clear from context, a Euclidean triangle is sometimes denoted simply by T.

Proposition 6.3 (no short saddle connections condition) *For any $X \in \overline{D}$, there is a positive lower bound on the lengths of all saddle connections in (S, X).*

Proof Recall from Definition 4.1 that for parabolic saddle connections σ the horoballs $B_{[\sigma]}$ were chosen so that over $\partial B_{[\sigma]}$ the saddle connection σ would have (positive) length less than any saddle connection in a different direction. (The lengths of the nonparabolic saddle connections are necessarily positive over the convex hull.) Because the length of the shortest saddle connection on (S, X) is a continuous function on D/G and $\overline{D}/G$ is compact, there must be a positive lower bound on the lengths of all saddle connections in the fibers over $\overline{D}$. □

It is also true that Euclidean triangles cannot be too large. This is known for lattice surfaces (see, for instance, [33]), but requires a new proof for general flat surfaces.

Proposition 6.4 (no large triangles condition) *There exists $A > 0$ so that for any $X \in D$ the area of any Euclidean triangle in (S, X, q) is at most A.*

Proof Any Euclidean triangle minus its vertices is embedded in the flat surface (S, X, q) [24, Lemma 2.1]. Choose A to be the area of (S, X, q). Then any Euclidean triangle in (S, X, q) has area at most A. Recalling a fact from Section 2.3, the choice of A is uniform for all $X \in D$ because the area of (S, X, q) is preserved by the action of $\mathrm{SL}_2(\mathbb{R})$. □

Corollary 6.5 *Every Euclidean triangle is uniformly slim.*

Proof The inradius of any Euclidean triangle is bounded above by the inradius of an equilateral triangle of area A (where A is the constant from Proposition 6.4), so every Euclidean triangle is $2\sqrt{A}/3^{3/4}$-slim. □

Throughout this section, it will be necessary to relate Euclidean triangles in E_0 to particular ideal triangles in D, described below.

Definition 6.6 Let $x, y, z \in \Sigma$ so that $T = \Delta(x, y, z)$ is a Euclidean triangle consisting of saddle connections σ_x, σ_y, and σ_z. The *ideal triangle associated to T*, denoted $\widetilde{T}$, is the ideal triangle in D with vertices $[\sigma_x]$, $\left[\sigma_y\right]$, and $[\sigma_z]$.

These are studied in depth by [24].

Definition 6.7 Let $T = \Delta(x, y, z)$ be a Euclidean triangle. The *balance point of T* (or of its associated ideal triangle $\widetilde{T}$) is the unique point $b \in D$ such that $f_b(T)$ is an equilateral triangle. The *adjusted balance point of T* (or of its associated ideal triangle $\widetilde{T}$), denoted $\bar{b}$, is the projection of b onto $\mathrm{hull}(G)$. If $b \in \mathrm{hull}(G)$, then b and $\bar{b}$ coincide.

Remark 6.8 Existence and uniqueness of the balance point is given by [8, Lemma 4.9]. Uniqueness of the adjusted balance point follows from uniqueness of the closest-point projection.

Proposition 6.9 *Let $T = \Delta(x, y, z)$ be a Euclidean triangle with sides consisting of saddle connections σ_x, σ_y, and σ_z. Then the balance point b is the incenter of the ideal triangle $\widetilde{T}$ (that is, the center of the largest circle inscribed in $\widetilde{T}$).*

Proof Without loss of generality assume that $[\sigma_x]$ is the vertical direction (perhaps after acting on D with an element of SO(2); see Section 2.3). Changing the coordinate chart by a rotation of $\pm\pi/3$ makes $\left[\sigma_y\right]$ or $[\sigma_z]$ vertical. Therefore the three geodesic rays in D starting at b and ending at $[\sigma_x]$, $\left[\sigma_y\right]$, or $[\sigma_z]$ diverge from b at equal angles. In particular, changing the coordinate chart by a rotation of $\pm\pi/3$ preserves $\widetilde{T} \subset D$. It also preserves the three geodesic segments from b to the nearest point on each side of $\widetilde{T}$, so b must be the incenter of $\widetilde{T}$. □

Corollary 6.10 *Let T be a Euclidean triangle, and let b be the balance point of T. Because the area of $f_b(T)$ is bounded above by A (Proposition 6.4), each side of $f_b(T)$ has length at most $2\sqrt{A}$.*

For many of the following proofs it is important to recall from Section 2 that given a saddle connection σ with direction $[\sigma] \in \partial D$, the horoballs based at $[\sigma] \in D$ are sublevel sets for the length of σ, where the length of σ shrinks exponentially along any geodesic ending at $[\sigma]$. The following technical lemma states this more precisely.

Throughout Lemmas 6.11 and 6.12, a geodesic between $[\sigma_x]$ and $\left[\sigma_y\right]$ is denoted $\overleftrightarrow{[\sigma_x], \left[\sigma_y\right]}$.

Lemma 6.11 *Let $T = \Delta(x, y, z)$ be a Euclidean triangle with sides consisting of saddle connections σ_x, σ_y, and σ_z. Then for any segment $c \subset \overleftrightarrow{[\sigma_x], \left[\sigma_y\right]}$ of length $L > 0$,*

$$\min_{t \in c} \left(\min \left\{ \ell\left(f_t(\sigma_x)\right), \ell\left(f_t(\sigma_y)\right) \right\} \right) \leq 2\sqrt{3A}e^{-L/2},$$

where A is the constant from Proposition 6.4.

Proof At the balance point b for T, the lengths of $f_b(\sigma_x)$ and $f_b(\sigma_y)$ are at most $2\sqrt{A}$ (Corollary 6.10). Denote by t' the point of intersection of $\overleftrightarrow{[\sigma_x], \left[\sigma_y\right]}$ with the inscribed circle of $\widetilde{T}$. The distance between b and t' is $\ln\sqrt{3}$ (as this is the inradius of an ideal triangle — see Proposition 6.9), so the lengths of $f_{t'}(\sigma_x)$ and $f_{t'}(\sigma_y)$ are at most $2\sqrt{A}e^{\ln\sqrt{3}} = 2\sqrt{3A}$. Traversing $\overleftrightarrow{[\sigma_x], \left[\sigma_y\right]}$ towards $[\sigma_x]$ (or towards $\left[\sigma_y\right]$) shrinks $\ell(\sigma_x)$ (or $\ell\left(\sigma_y\right)$) exponentially as a function of the distance traversed. Since one of the endpoints has distance at least $L/2$ from t', it follows that $\min\left(\ell(\sigma_x), \ell\left(\sigma_y\right)\right) \leq 2\sqrt{3A}e^{-L/2}$ on c. □

Lemma 6.12 *Let $T = \Delta(x, y, z)$ be a Euclidean triangle consisting of saddle connections σ_x, σ_y, and σ_z; let $\widetilde{T} \subset D$ be the ideal triangle associated to directions $[\sigma_x]$, $\left[\sigma_y\right]$, and $[\sigma_z]$; and let $\mathrm{proj}_{\overline{D}}(\widetilde{T})$ denote the closest-point projection of $\widetilde{T}$ onto $\overline{D}$ (where this map is the same as in Section 4). Then the perimeter of $\mathrm{proj}_{\overline{D}}(\widetilde{T})$ is uniformly bounded. As a consequence, the diameter of $\mathrm{proj}_{\overline{D}}(\widetilde{T})$ is uniformly bounded.*

Proof Denote the balance point for $\widetilde{T}$ by b and recall that the lengths of σ_x, σ_y, and σ_z are all uniformly bounded at b. They are also uniformly bounded at the point closest to b on the side of $\widetilde{T}$ opposite (without

loss of generality) $[\sigma_x]$ (as in the proof of Lemma 6.11). Then, traversing the geodesic from this point toward (without loss of generality) $\left[\sigma_y\right]$, $\ell(\sigma_y) \to 0$. Because there are no short saddle connections in $\overline{D}$ (Proposition 6.3), there is a bound on the length of the intersection of each side of $\widetilde{T}$ with $\overline{D}$.

If $[\sigma]$ is a parabolic vertex of $\widetilde{T}$, then $\mathrm{proj}_{\overline{D}}(\widetilde{T} \cap B_{[\sigma]}) = \widetilde{T} \cap \partial B_{[\sigma]}$, which has uniformly bounded length due to the choice of horoballs (see Definition 4.1).

If the distance between $\widetilde{T}$ and $\mathrm{hull}\,(G)$ is at least 1, then the projection of $\widetilde{T}$ onto any component of $\partial\,\mathrm{hull}\,(G)$ is bounded. More generally, any component of $\widetilde{T}$ which lies outside of the 1-neighborhood of $\mathrm{hull}\,(G)$ in D projects to a bounded subsegment of $\partial\,\mathrm{hull}\,(G)$.

It only remains to consider the portions of $\widetilde{T}$ contained in the 1-neighborhood of $\mathrm{hull}\,(G)$. By convexity of $\mathrm{hull}\,(G)$, the intersection of a side of $\widetilde{T}$ with the 1-neighborhood of $\mathrm{hull}\,(G)$ consists of at most one component. Fix some $L > 0$ and assume (without loss of generality) that there is such a component $c \subset \overleftrightarrow{[\sigma_x], \left[\sigma_y\right]}$ of length $L > 0$ contained in the 1-neighborhood of a component of $\partial\,\mathrm{hull}\,(G)$. Then the lengths of σ_x and σ_y in the fibers over points of $\mathrm{proj}_{\overline{D}}(c)$ are bounded by $2\sqrt{3A}e^{-L/2+1}$ (Lemma 6.11). Again, because there are no short saddle connections in $\overline{D}$ (Proposition 6.3), this implies that the projection of c to $\partial\,\mathrm{hull}\,(G)$ has bounded length.

Because each of these cases can occur at most three times for any $\widetilde{T}$, there is a bound on the perimeter of $\mathrm{proj}_{\overline{D}}(\widetilde{T}) \cap \overline{D}$. Therefore the diameter of $\mathrm{proj}_{\overline{D}}(\widetilde{T}) \cap \overline{D}$ is also bounded. □

The next lemma shows that for any ideal triangle that intersects $\mathrm{hull}\,(G)$, the balance point and adjusted balance point are close to each other, to the horopoints associated to any nonparabolic vertices, and to the horoballs associated to any parabolic vertices. This statement functions similarly to [8, Lemma 4.10], which is a consequence of [33]; both assume in addition that G is a lattice, which is not necessary for the result below.

Lemma 6.13 *Let $T = \Delta(x, y, z)$ be a Euclidean triangle consisting of saddle connections σ_x, σ_y, and σ_z so that $\widetilde{T} \cap \mathrm{hull}\,(G) \neq \varnothing$. Then the balance point b of $\widetilde{T}$, the adjusted balance point $\bar{b}$ of $\widetilde{T}$, the horoballs associated to any parabolic vertices of $\widetilde{T}$, and the horopoints associated to any nonparabolic vertices of $\widetilde{T}$ are all uniformly close to one another. See Figure 4.*

Proof Because the projection of $\widetilde{T}$ to $\mathrm{hull}\,(G)$ has bounded diameter (Lemma 6.12), the adjusted balance point $\bar{b}$ and the horopoints associated to any nonparabolic vertices of $\widetilde{T}$ must be uniformly close. Also, the adjusted balance point $\bar{b}$ must be uniformly close to the horoballs associated to any parabolic vertices of $\widetilde{T}$ because G acts cocompactly on the truncated hull $\overline{D}$, meaning that $\bar{b}$ is in some uniformly thick part of $\mathrm{hull}\,(G)$. Lastly, if $b \in \mathrm{hull}\,(G)$, then $b \in \overline{D}$ (by the choice of horoballs made during construction; see Definition 4.1) and therefore $b = \bar{b}$.

It only remains to show that b and $\bar{b}$ are uniformly close even when b lies outside of $\mathrm{hull}\,(G)$. Choose a saddle connection σ in T for which $[\sigma] \in \partial D$ and b are in different components of the closure of $D - \mathrm{hull}\,(G)$. Denote by c the geodesic from b to $[\sigma]$. The length of σ decreases along c as it approaches $[\sigma]$. The length of σ is uniformly bounded above at b (Corollary 6.10) and uniformly bounded below at a point $t \in c \cap \partial\,\mathrm{hull}\,(G)$ (Proposition 6.3). Therefore the subsegment of c between b and t has

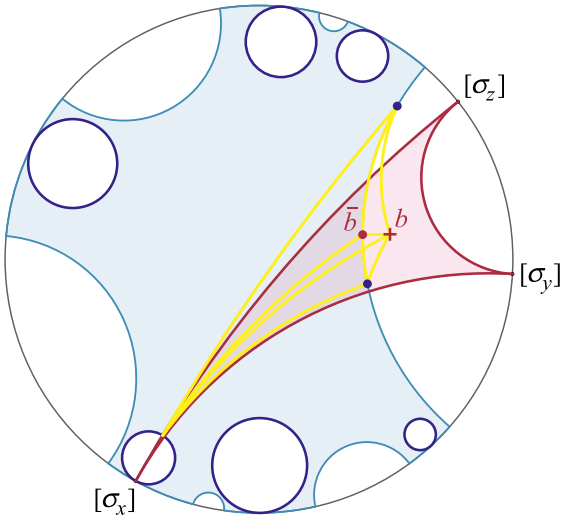


Figure 4: An example of Lemma 6.13: When $\widetilde{T}$ (red) intersects hull (G) (blue), there is a uniform bound on the distances (yellow curves) between the balance point b of $\widetilde{T}$ (red $+$), the adjusted balance point $\bar{b}$ of $\widetilde{T}$ (red dot), the horoball B_x associated to the parabolic direction $[\sigma_x]$ (blue circle), and the horopoints B_y and B_z associated to the nonparabolic directions $[\sigma_y]$ and $[\sigma_z]$ (blue dots).

uniformly bounded length. Since $\bar{b}$ is the closest point projection of b to ∂ hull (G), the distance between b and $\bar{b}$ is no greater than the distance between b and t, so b and $\bar{b}$ are uniformly close. □

Corollary 6.14 *Let $T = \Delta(x, y, z)$ be a Euclidean triangle consisting of saddle connections σ_x, σ_y, and σ_z so that $\widetilde{T} \cap$ hull $(G) \neq \varnothing$. If σ_x is nonparabolic, then it has uniformly bounded length at its associated horopoint. That is, the length of*

$$f_{X_{[\sigma_x]}}(\sigma_x)$$

is uniformly bounded.

The next definition is from [8, Section 4.6]. The following corollary is comparable to [8, Corollary 4.11], which is given a new proof.

Definition 6.15 Two saddle connections in a common fiber *span a triangle* if they share an endpoint and the geodesic joining their other endpoints is a single (possibly degenerate) saddle connection. Write $\mathcal{P}(\sigma)$ for the set of saddle connections that span a triangle with σ. Denote

$$B(\sigma) = \bigcup_{\sigma' \in \mathcal{P}(\sigma)} B_{[\sigma']}$$

for the union of horoballs or horopoints associated to the saddle connections that span a triangle with σ. Taking preimages of the horoballs or horopoints, denote

$$\mathcal{B}(\sigma) = \bigcup_{\sigma' \in \mathcal{P}(\sigma)} \mathcal{B}_{[\sigma']}.$$

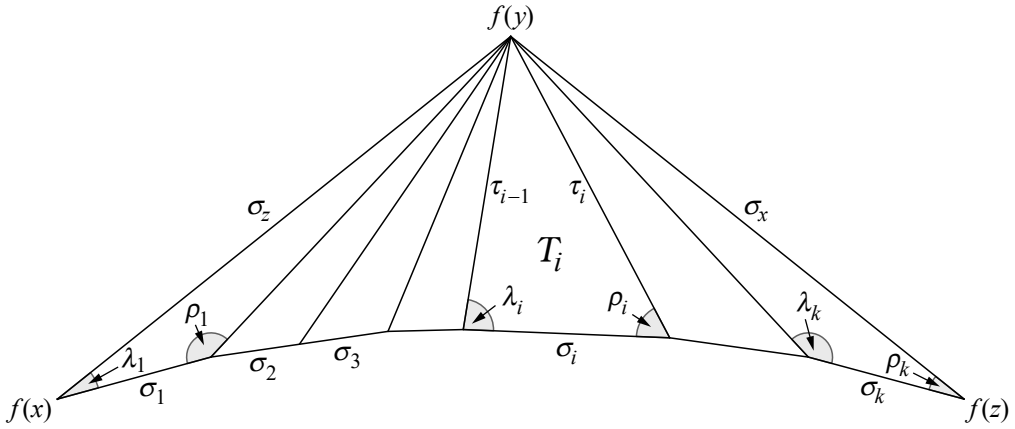


Figure 5: An example of a fan $\Delta(x, y, z)$ in E_0 in which $[f(x), f(y)]$ and $[f(y), f(z)]$ are each a single saddle connection, labeled with the notation used in the proof of Lemma 6.19 and throughout the rest of the paper. In this example, the saddle connections σ_2 and σ_3 are parallel, i.e., $[\sigma_2] = [\sigma_3]$.

Corollary 6.16 *For any $\sigma' \in \mathcal{P}(\sigma)$, the sets $B_{[\sigma]}$ and $B_{[\sigma']}$ are uniformly close in D.*

Proof Suppose saddle connections σ and σ' span a triangle T, and let $\widetilde{T}$ be the associated ideal triangle. Then $B_{[\sigma]}$ and $B_{[\sigma']}$ are close by Lemma 6.12. □

6.2 Fans

Definition 6.17 Let $x, y, z \in \Sigma$. The triangle $\Delta(x, y, z) \subset E_0$ is called a *fan* if it is nondegenerate and at least two sides consist of a single saddle connection. When the information about x, y, z is clear from context, a fan is sometimes denoted simply by F.

A fan canonically decomposes into a finite union of Euclidean triangles sharing a common vertex. See Figure 5 for an illustration and notation used throughout this section.

Definition 6.18 Let $F = \Delta(x, y, z) \subset E_0$ be a fan as in Figure 5. The *ideal fan associated to* $\Delta(x, y, z)$, denoted $\widetilde{F}$, is the union of the ideal triangles $\widetilde{T_i} \subset D$ associated to the Euclidean triangles $T_i \subset \Delta(x, y, z)$.

Figure 6 shows the ideal fan associated to the fan from Figure 5. For a fan F consisting of k Euclidean triangles, the associated ideal fan $\widetilde{F}$ consists of k ideal triangles with disjoint interiors formed from at most $2k + 1$ vertices; of these vertices, $k - 1$ vertices each belong to exactly two ideal triangles. The following lemma describes that these $k - 1$ vertices are contained in a connected component of the boundary at infinity.

Lemma 6.19 (structure lemma) *Let $F = \Delta(x, y, z)$ be a fan consisting of k Euclidean triangles $\{T_i\}_{1 \le i \le k}$ as in Figure 5. Then the vertices of the associated ideal fan $\widetilde{F}$ appear in cyclic order (counterclockwise) in ∂D:*

$$[\sigma_z] < [\tau_1] < \cdots < [\tau_{k-1}] < [\sigma_x] < [\sigma_k] \le \cdots \le [\sigma_1] < [\sigma_z].$$

Moreover, the ideal triangles $\{\widetilde{T_i}\}_{1 \le i \le k}$ have disjoint interiors. See Figure 6.

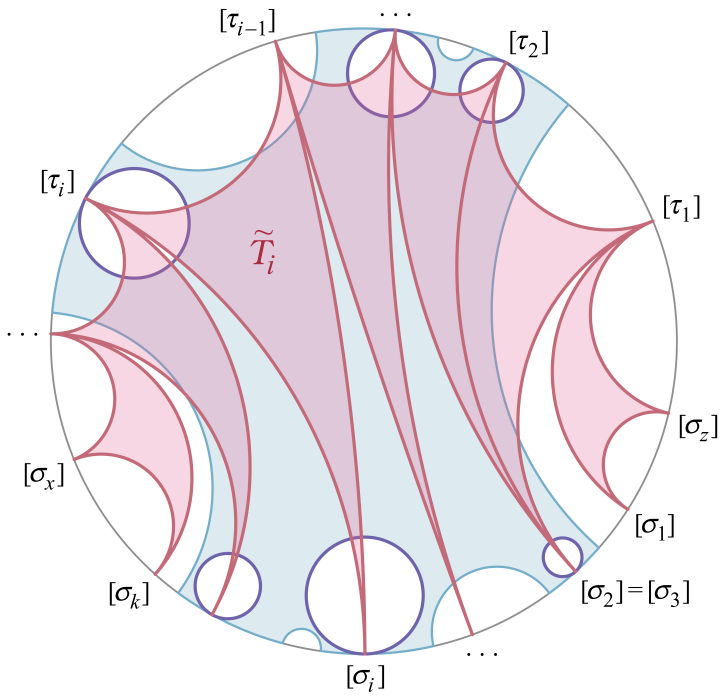


Figure 6: An example of an ideal fan (red) in D (with ∂D represented by the gray circle) corresponding to the Euclidean fan from Figure 5. An example of hull (G) is in blue. As given by the structure lemma (Lemma 6.19) each vertex $[\tau_*]$ belongs to exactly two ideal triangles, and $[\sigma_x]$ and $[\sigma_z]$ each belong to exactly one ideal triangle. The remaining vertices may belong to one or more triangles: In this example, the vertex $[\sigma_2]=[\sigma_3]$ belongs to two triangles because σ_2 and σ_3 were parallel in the Euclidean fan. Most importantly, the interiors of the ideal triangles are pairwise disjoint. Notice that some triangles in the ideal fan do not intersect hull (G), which is important for Corollary 6.20.

Proof Refer to Figures 5 and 6 for a Euclidean fan, an associated ideal fan, and related notation. Develop F into the Euclidean plane; it can be assumed that, after rotating, σ_1 is in the horizontal direction. Because $\sigma_1\sigma_2\cdots\sigma_k$ is a geodesic in E_0, $\rho_i+\lambda_{i+1}\geq\pi$, and therefore

$$[\sigma_k]\leq[\sigma_{k-1}]\leq\cdots\leq[\sigma_1]\,.$$

Note that for each T_i

$$[\tau_{i-1}]<[\tau_i]<[\sigma_i]\,,$$

and therefore

$$[\sigma_z]=[\tau_0]<[\tau_1]<[\tau_2]<\cdots<[\tau_{k-2}]<[\tau_{k-1}]<[\tau_k]=[\sigma_x]<[\sigma_k]\leq[\sigma_{k-1}]\leq\cdots\leq[\sigma_1]<[\sigma_z]\,.$$

Notice that the inequalities concerning the vertices of $\widetilde{T}_1$ are "nested" inside those concerning $\widetilde{T}_2$ — that is, the inequalities $[\sigma_1]<[\sigma_z]<[\tau_1]$ relating the vertices of $\widetilde{T}_1$ are inserted (as $*$) into the inequalities $[\sigma_2]\leq *<[\tau_1]<[\tau_2]$ relating the vertices of $\widetilde{T}_2$ — the inequalities concerning the vertices of $\widetilde{T}_2$ are

nested inside those concerning $\widetilde{T}_3$, and so on. This demonstrates that the ideal triangles $\widetilde{T}_i$, $1 \le i \le k$, have disjoint interiors and proves the lemma. □

The following corollaries are used in the proofs of the fan lemma (Lemmas 6.24 and 6.22).

Corollary 6.20 *Any ideal fan in D can be split into at most three subfans: $\{\widetilde{T}_1, \dots, \widetilde{T}_n\}$, which is disjoint from* hull (G); $\{\widetilde{T}_{n+1}, \dots, \widetilde{T}_{m-1}\}$, *for which all triangles intersect* hull (G); *and $\{\widetilde{T}_m, \dots, \widetilde{T}_k\}$, which is also disjoint from* hull (G). (*See, for example, Figure 6.*)

Recall that $\mathrm{proj}_{\overline{D}} : D \to \overline{D}$ is the ρ-closest point projection.

Corollary 6.21 *Let $F = \Delta(x, y, z)$ be a fan, and let $\widetilde{F} \subset D$ be the ideal fan associated to F. If $\widetilde{F} \cap$ hull $(G) = \varnothing$, then there is an upper bound on the diameter of* $\mathrm{proj}_{\overline{D}}(\widetilde{F}) \subset \partial$ hull (G).

Proof By Lemma 6.19, $\mathrm{proj}_{\overline{D}}(\widetilde{F})$ is contained in the image of the projection of one or two ideal triangles in $\widetilde{F}$. In the former case, $\mathrm{proj}_{\overline{D}}(\widetilde{F})$ lies in the image of the projection of a single ideal triangle, which is uniformly bounded (Lemma 6.12). In the latter case these two ideal triangles share a common vertex, so an upper bound on the lengths of the projections of both triangles implies an upper bound on the diameter of $\mathrm{proj}_{\overline{D}}(\widetilde{F})$. □

6.3 The fan lemma

This section presents a statement and proof of the fan lemma (Lemma 6.22), which says that the triangle of collapsed preferred paths associated to any Euclidean fan is uniformly slim. It is used later to demonstrate uniform slimness of general triangles of collapsed preferred paths (Theorem 6.28). Slimness of triangles of collapsed preferred paths is the only remaining condition needed to satisfy all hypotheses of the guessing geodesics criterion (Proposition 5.1), which gives hyperbolicity of $\widehat{E}$ (Theorem 5.7).

The statement of the fan lemma is identical to that of [8, Lemma 4.12], but the proof requires significant reformulation in order to accommodate the circumstance in which the Veech group G is not a lattice.

Lemma 6.22 (fan lemma) *There is $\delta' > 0$ so that if $x, y, z \in \Sigma$ and the geodesic triangle $\Delta(x, y, z)$ in E_0 is a fan with $[f(x), f(y)] = \sigma_z$ and $[f(y), f(z)] = \sigma_x$, then the triangle $\Delta^{\hat{\xi}}(x, y, z)$ of collapsed preferred paths is δ'-slim. Furthermore, if y lies on a geodesic in D_y with endpoints in $\mathcal{B}(\sigma_z)$ and $\mathcal{B}(\sigma_x)$, then the collapsed preferred paths satisfy*

$$\hat{\xi}(x, y), \hat{\xi}(y, z) \subset N_{\delta'}(\hat{\xi}(x, z)) \quad \text{and} \quad \hat{\xi}(x, z) \subset N_{\delta'}(\hat{\xi}(x, y) \cup \hat{\xi}(y, z)).$$

The "furthermore" statement is necessary for getting a uniform slimness constant for general triangles in later arguments. A Euclidean fan fitting the description in the fan lemma is shown in Figure 5. The (uncollapsed) preferred paths joining the vertices $x, y, z \in \Sigma$ are described in Tables 1 and 2.

The fan lemma is proved by [8] under the additional hypothesis that G is a lattice. In their case, all of the saddle connections are parabolic and therefore have length zero after being projected to $\widehat{E}$ by the map P. This places most of the burden of the proof on showing that the horizontal pieces of the paths along the top and bottom of the fan are sufficiently close. Here G is only required to be finitely

h'_z	horizontal geodesic in D_x from x to $X_{[\sigma_z]}$
γ'_z	saddle connection σ_z traversed in the fiber over the associated horoball or horopoint, i.e., $f_{X_{[\sigma_z]}}(\sigma_z)$
$\bar{h}'_z$	horizontal geodesic in D_y from $X_{[\sigma_z]}$ to y
$\bar{h}'_x$	horizontal geodesic in D_y from y to $X_{[\sigma_x]}$
γ'_x	saddle connection σ_x traversed in the fiber over the associated horoball or horopoint, i.e., $f_{X_{[\sigma_x]}}(\sigma_x)$
h'_x	horizontal geodesic in D_z from $X_{[\sigma_x]}$ to z

Table 1: In order, the pieces forming the concatenation of the (uncollapsed) preferred paths forming the "top" of the fan referred to in Lemma 6.22. The primed notation on the paths in this table and in Table 2 serves a practical purpose: The unprimed notation is reserved for paths which play a much more active role in the proof of the fan lemma.

generated, so preferred paths may contain nonparabolic saddle connections whose lengths are arbitrarily long even after collapsing. In the proof of the fan lemma that follows, the preferred path along the bottom of the fan is replaced with a substitute path to which it is uniformly close. The substitute path is chosen specifically to allow nonparabolic saddle connections to be traversed in a select set of fibers where

h'_0	horizontal geodesic in D_x from x to $X_{[\sigma_1]}$
γ'_1	saddle connection σ_1 traversed in the fiber over the associated horoball or horopoint, i.e., $f_{X_{[\sigma_1]}}(\sigma_1)$
h'_1	horizontal geodesic in $D_{\gamma'_1+} = D_{\gamma'_2-}$ from $X_{[\sigma_1]}$ to $X_{[\sigma_2]}$
$\dots$	$\dots$
γ'_i	saddle connection σ_i traversed in the fiber over the associated horoball or horopoint, i.e., $f_{X_{[\sigma_i]}}(\sigma_i)$
h'_i	horizontal geodesic in $D_{\gamma'_i+} = D_{\gamma'_{i+1}-}$ from $X_{[\sigma_i]}$ to $X_{[\sigma_{i+1}]}$
$\dots$	$\dots$
γ'_k	saddle connection σ_k traversed in the fiber over the associated horoball or horopoint, i.e., $f_{X_{[\sigma_k]}}(\sigma_k)$
h'_k	horizontal geodesic in D_z from $X_{[\sigma_k]}$ to z

Table 2: In order, the pieces forming the (uncollapsed) preferred path forming the "bottom" of the fan referred to in Lemma 6.22. The primed notation on the paths in this table and in Table 1 serves a practical purpose: The unprimed notation is reserved for paths which play a much more active role in the proof of the fan lemma. The superscripts $+$ and $-$ denote endpoints of geodesics, as described in Section 5.1.

slimness of Euclidean fans can be employed. The horizontal pieces of the substitute path are handled mostly in the same manner as in the lattice case presented by [8], except for a few horizontal pieces possibly requiring special treatment.

The following lemma will be necessary on a few occasions.

Lemma 6.23 (tube lemma; adapted from [14, Lemma 4.5]) *Let $[\sigma], [\sigma'] \in \mathscr{D}$, and let v be a geodesic in D with endpoints in $\partial B_{[\sigma]}$ and $\partial B_{[\sigma']}$. If w is a geodesic in D with endpoints each within a bounded distance of $B_{[\sigma]}$ and $B_{[\sigma']}$, then w lies within a uniform neighborhood of $v \cup B_{[\sigma]} \cup B_{[\sigma']}$.*

To simplify the proof of the fan lemma and to demonstrate the utility of substituting preferred paths, the case in which the ideal fan is disjoint from $\operatorname{hull}(G)$ is presented separately.

Lemma 6.24 (disjoint fan lemma) *The fan lemma (Lemma 6.22) holds under the additional assumption that the ideal fan in D associated to $\Delta(x, y, z)$ does not intersect* $\operatorname{hull}(G)$.

Proof Note that objects are sometimes lifted from D to various horizontal disks. For the sake of simplicity, the lifts are referred to by the same notation wherever it does not cause confusion.

Because the ideal fan never intersects the hull, all of the saddle connections in the preferred paths are traversed in the fibers over their associated horopoints, which lie in a uniformly bounded subsegment of $\partial \operatorname{hull}(G)$ (Corollary 6.21). The general trajectory of the proof is to replace the preferred paths with uniformly close substitute paths in which all of the saddle connections are traversed in the fiber over $X_{[\sigma_z]}$, then prove that the triangle of substitute paths is slim.

The substitute path along the "top" of the fan is the concatenation of the following paths, in order:

h_z	horizontal geodesic in D_x from x to $X_{[\sigma_z]}$
γ_z	saddle connection σ_z traversed in the fiber over $X_{[\sigma_z]}$, i.e., $f_{X_{[\sigma_z]}}(\sigma_z)$
$\bar{h}_z$	horizontal geodesic in D_y from $X_{[\sigma_z]}$ to y
$\bar{h}_x$	horizontal geodesic in D_y from y to $X_{[\sigma_z]}$
γ_x	saddle connection σ_x traversed in the fiber over $X_{[\sigma_z]}$, i.e., $f_{X_{[\sigma_z]}}(\sigma_x)$
h_x	horizontal geodesic in D_z from $X_{[\sigma_z]}$ to z

This substitute path is uniformly close to the path in Table 1, which can be seen by comparing each pair of analogous pieces as follows.

- Since $h_z = h'_z$, these are uniformly close.
- Since $\gamma_z = \gamma'_z$, these are uniformly close.
- Since $\bar{h}_z = \bar{h}'_z$, these are uniformly close.
- Since $\bar{h}_x, \bar{h}'_x \subset D_y$ share an initial point and have terminal points ($X_{[\sigma_z]}$ and $X_{[\sigma_x]}$, respectively) which are a uniformly bounded distance apart by Corollary 6.21, $\bar{h}_x$ and $\bar{h}'_x$ are uniformly close.

• The saddle connections γ_x and γ'_x are saddle connections in the fibers over $X_{[\sigma_z]}$ and $X_{[\sigma_x]}$, respectively. The map $f_{X_{[\sigma_z]},X_{[\sigma_x]}}$ maps γ'_x to γ_x, and so the distance between them is exactly the distance between $X_{[\sigma_x]}$ and $X_{[\sigma_z]}$, which is uniformly bounded.

• Since $h_x, h'_x \subset D_z$ share a terminal point and have initial points ($X_{[\sigma_z]}$ and $X_{[\sigma_x]}$, respectively) a uniformly bounded distance apart, h_x and h'_x are uniformly close.

The substitute path along the "bottom" of the fan is the concatenation of the following paths, in order:

h_0	horizontal geodesic in D_x from x to $X_{[\sigma_z]}$
γ_1	saddle connection σ_1 traversed in the fiber over $X_{[\sigma_z]}$, i.e., $f_{X_{[\sigma_z]}}(\sigma_1)$
h_1	point $\gamma_1^+ = \gamma_2^-$
$\dots$	$\dots$
γ_i	saddle connection σ_i traversed in the fiber over $X_{[\sigma_z]}$, i.e., $f_{X_{[\sigma_z]}}(\sigma_i)$
h_i	point $\gamma_i^+ = \gamma_{i+1}^-$
$\dots$	$\dots$
γ_k	saddle connection σ_k traversed in the fiber over $X_{[\sigma_z]}$, i.e., $f_{X_{[\sigma_z]}}(\sigma_k)$
h_k	horizontal geodesic in D_z from $X_{[\sigma_z]}$ to z

All but the first and last horizontal paths are points because the concatenation of saddle connections γ_i in the fiber over $X_{[\sigma_z]}$ is already continuous. Again, this substitute path is uniformly close to the path in Table 2, which can be seen by comparing each pair of analogous pieces as follows.

• Since $h_0, h'_0 \subset D_x$ share an initial point and have terminal points ($X_{[\sigma_z]}$ and $X_{[\sigma_1]}$, respectively) a uniformly bounded distance apart, h_0 and h'_0 are uniformly close.

• For all i, the saddle connections γ_i and γ'_i are saddle connections in the fibers over $X_{[\sigma_z]}$ and $X_{[\sigma_i]}$, respectively. The map $f_{X_{[\sigma_z]},X_{[\sigma_i]}}$ maps γ'_i to γ_i, and so the distance between them is exactly the distance between $X_{[\sigma_i]}$ and $X_{[\sigma_z]}$, which is uniformly bounded.

• For each $1 \le i \le k-1$, h_i and h'_i have initial points which are the terminal points of γ_i and γ'_i, respectively, and terminal points which are the initial points of γ_{i+1} and γ'_{i+1}, respectively. Since for all i the pair of γ_i, γ'_i are uniformly close and h_i is contained in a hyperbolic space (with hyperbolicity constant uniform for all i), the pair h_i, h'_i must also be uniformly close.

• Since $h_k, h'_k \subset D_z$ share a terminal point and have initial points ($X_{[\sigma_z]}$ and $X_{[\sigma_k]}$, respectively) a uniformly bounded distance apart, h_k and h'_k are uniformly close.

Since h_i is degenerate for each $1 \le i \le k-1$, the substitute path along the "bottom" of the fan is simply $h_0\gamma_1\gamma_2\cdots\gamma_k h_k$. Then because $h_z = h_0$, $\bar{h}_z = \bar{h}_x$ (ignoring orientation), and $h_x = h_k$, to show that the triangle of substitute paths is slim it only remains to show that the saddle connections γ_* form a slim triangle.

Claim 6.25 *The Euclidean fan formed by the saddle connections $\gamma_x, \gamma_z, \gamma_1, \dots, \gamma_k$ in the fiber over $X_{[\sigma_z]}$ is uniformly slim.*

Proof For any $X \in D$, the fiber E_X is quasi-isometric to $(\widetilde{S}, \widetilde{X_0})$, the universal cover of the closed surface (S, X_0), which is hyperbolic. Given any X_1 and X_2 in an ϵ-neighborhood of X_0, the map $f_{X_2,X_1} : E_{X_1} \to E_{X_2}$ is e^{ϵ}-bilipschitz (Section 4.2.1), so the fibers E_{X_1} and E_{X_2} are uniformly hyperbolic. Finally, because G acts cocompactly on the truncated convex hull $\overline{D}$, it holds that all fibers E_X with $X \in \overline{D}$ are uniformly hyperbolic. Therefore the Euclidean fan formed by the saddle connections $\gamma_x, \gamma_z, \gamma_1, \dots, \gamma_k$ in the fiber over $X_{[\sigma_z]}$ is uniformly slim. □

Therefore, the original triangle of preferred paths is slim.

To prove the "furthermore" statement, note that since $\bar{h}'_z$ and $\bar{h}'_x$ are both in D_y, they form a slim triangle with the geodesic joining their endpoints, denoted h'_y. When y is assumed to lie on a geodesic with endpoints in $\mathcal{B}(\sigma_z)$ and $\mathcal{B}(\sigma_x)$, the concatenation of $\bar{h}'_z$ and $\bar{h}'_x$ lies uniformly close to h'_y (Lemma 6.23). By construction, h'_y is a uniformly bounded subsegment of $\partial \operatorname{hull}(G)$. Therefore h'_y — and the concatenation of $\bar{h}'_z$ and $\bar{h}'_x$ — is uniformly close to the shared endpoint of γ_z and γ_x, which results in the containments from the lemma. □

When the ideal fan is not disjoint from $\operatorname{hull}(G)$, the proof is more complex. First the fan is decomposed into subfans as in Corollary 6.20: One subfan consists of all of the ideal triangles that intersect $\operatorname{hull}(G)$, and the other subfans (up to two, if any exist) are disjoint from $\operatorname{hull}(G)$. Thanks to the tools in Lemma 6.13, the paths associated to the ideal subfan which intersects $\operatorname{hull}(G)$ can be handled in a manner similar to that of [8], with a few technicalities to consider for the horizontal paths at the beginning and end. A subfan which is disjoint from $\operatorname{hull}(G)$ is handled in a similar manner to the previous proof; however, some additional work is needed since slimness of the associated Euclidean subfan in any particular fiber over $\partial \operatorname{hull}(G)$ is no longer sufficient to prove slimness of the whole fan.

Proof of fan lemma Like the disjoint case, note that objects are sometimes lifted from D to other horizontal disks, and the lifts are referred to by the same notation wherever it does not cause confusion. Finally, note that some arguments are made by showing that the uncollapsed pieces (i.e., before applying P) are close, which is sufficient because P is 1-Lipschitz (Lemma 4.3).

Substituting the concatenated preferred paths along the top of the fan. Compare to the concatenated preferred paths described in Table 1. The substitute path along the top of the fan is the concatenation of the following paths, in order:

$h_z := h'_z$	horizontal geodesic in D_x from x to $X_{[\sigma_z]}$
$\gamma_z := \gamma'_z$	saddle connection σ_z traversed in the fiber over the associated horoball or horopoint, i.e., $f_{X_{[\sigma_z]}}(\sigma_z)$
h'_y	horizontal geodesic in D_y joining $\bar{h}'^{-}_z \in X_{[\sigma_z]}$ to $\bar{h}'^{+}_x \in X_{[\sigma_x]}$
$\gamma_x := \gamma'_x$	saddle connection σ_x traversed in the fiber over the associated horoball or horopoint, i.e., $f_{X_{[\sigma_x]}}(\sigma_x)$
$h_x := h'_x$	horizontal geodesic in D_z from $X_{[\sigma_x]}$ to z

Since $\bar{h}'_z$ and $\bar{h}'_x$ are both in D_y, they form a slim triangle with the geodesic joining their endpoints, denoted h'_y. In particular, h'_y lies in a uniform neighborhood of $\bar{h}'_z$ and $\bar{h}'_x$. Because the rest of the pieces of the substitute path along the top of the fan are identical to those appearing in the (uncollapsed) preferred paths, the substitute path along the top of the fan lies in a uniform neighborhood of $\varsigma(x, y) \cup \varsigma(y, z)$ (while the reverse is not necessarily true). If in addition y lies on a geodesic with endpoints in $\mathcal{B}(\sigma_z)$ and $\mathcal{B}(\sigma_x)$, then the concatenation of $\bar{h}'_z$ with $\bar{h}'_x$ is in fact *uniformly* close to $h'_y \cup B(\sigma_z) \cup B(\sigma_x)$ (Lemma 6.23); therefore the image under P of the substitute path along the top of the fan is uniformly close to $\hat{\varsigma}(x, y) \cup \hat{\varsigma}(y, z)$.

Substituting the preferred path along the bottom of the fan. Compare to the preferred path described in Table 2. There are at most two maximal subfans of the ideal fan which do not intersect hull (G), and any such subfan must be at the beginning or end of the fan (Corollary 6.20). Let n be the largest index for which $\widetilde{T}_1, \dots, \widetilde{T}_n$ do not intersect hull (G). If no such n exists, set $n = 0$. Assume that $n < k$; the case when $n = k$ (that is, when no ideal triangle in the fan intersects hull (G)) is proven in Lemma 6.24. Similarly, let m be the smallest index for which $\widetilde{T}_m, \dots, \widetilde{T}_k$ do not intersect hull (G). If no such m exists, set $m = k + 1$. Note that as a result of the structure lemma, $m - n \geq 2$ (Corollary 6.20). Define

$$\gamma_i := \begin{cases} f_{X_{[\sigma_z]}}(\sigma_i), & 1 \leq i \leq n, \\ \gamma'_i, & n < i < m, \\ f_{X_{[\sigma_x]}}(\sigma_i), & m \leq i \leq k. \end{cases}$$

The horizontal paths are chosen to make the substitute path along the bottom of the fan continuous. These are explicitly described as follows. Because the saddle connections associated to the triangles $T_{n+1}, \dots, T_{m-1}$ remain unchanged, the horizontal paths connecting them are also unchanged. That is, define

$$h_i := h'_i \quad \text{for } n + 1 \leq i \leq m - 2.$$

The paths $\gamma_1\gamma_2 \cdots \gamma_n$ and $\gamma_m\gamma_{m+1} \cdots \gamma_k$ are the concatenation of the saddle connections $\sigma_1, \dots, \sigma_n$ in the fiber over $X_{[\sigma_z]}$ and the concatenation of the saddle connections $\sigma_m, \dots, \sigma_k$ in the fiber over $X_{[\sigma_x]}$, respectively. Note in particular that these paths are continuous without the presence of horizontal paths, so the horizontal "paths" are chosen to be the endpoints of saddle connections,

$$h_i := \gamma_i^+ = \gamma_{i+1}^- \quad \text{for } 1 \leq i \leq n - 1 \text{ or } m \leq i \leq k - 1.$$

When $n \geq 1$ and $m \leq k$, define the remaining horizontal paths as follows:

$$\begin{aligned} h_0 &:= h_z && \text{horizontal geodesic in } D_x \text{ from } x \text{ to } X_{[\sigma_z]} \\ h_n & && \text{horizontal geodesic in } D_{\gamma_n^+} \text{ from } \gamma_n^+ \text{ to } \gamma_{n+1}^- \\ h_{m-1} & && \text{horizontal geodesic in } D_{\gamma_{m-1}^+} \text{ from } \gamma_{m-1}^+ \text{ to } \gamma_m^- \\ h_k &:= h_x && \text{horizontal geodesic in } D_z \text{ from } X_{[\sigma_x]} \text{ to } z \end{aligned}$$

If instead $n = 0$, define $h_0 = h_n := h'_0$, and if instead $m = k + 1$, define $h_k = h_{m-1} := h'_k$. In other words, if there is no ideal subfan $\widetilde{T}_1, \dots, \widetilde{T}_n$ (or $\widetilde{T}_m, \dots, \widetilde{T}_k$) disjoint from hull (G), then the beginning (or end) of the substitute path along the bottom of the fan is chosen to be identical to the original preferred path.

The reader will gain the most insight from this proof by assuming in addition that $n \geq 1$ and $m \leq k$, since this represents the most novel circumstance in which the ideal fan can be split into three subfans depending on their intersection with hull (G) (see Corollary 6.20). The following claim, for instance, becomes tautological otherwise.

Claim 6.26 *The substitute path along the bottom of the fan, $h_0\gamma_1 h_1 \cdots \gamma_i h_i \cdots \gamma_k h_k$, is uniformly close to the preferred path along the bottom of the fan,*

$$\varsigma(x, z) = h'_0\gamma'_1 h'_1 \cdots \gamma'_i h'_i \cdots \gamma'_k h'_k.$$

Proof of claim If $n = 0$ and $m = k + 1$ — that is, if every triangle in the ideal fan intersects hull (G) — then the preferred path and the substitute path are identical. Otherwise if there are triangles in the ideal fan which do not intersect hull (G), then only the initial and/or terminal portions of the preferred and substitute paths are distinct. This proof will demonstrate the case when $n \geq 1$; the proof for the case when $m \leq k$ is analogous, and then the case when both $n \geq 1$ and $m \leq k$ follows immediately.

For any $i > n$, $h_i = h'_i$ and $\gamma_i = \gamma'_i$. So it remains to demonstrate that

$$h'_0\gamma'_1 h'_1 \cdots \gamma'_n h'_n \quad \text{and} \quad h_0\gamma_1 h_1 \cdots \gamma_n h_n$$

are uniformly close. This comes as a consequence of Corollary 6.21, which states that the projection of the ideal subfan $\widetilde{T}_1, \dots, \widetilde{T}_n$ to hull (G) is a uniformly bounded subsegment of ∂ hull (G). Therefore the initial segments h'_0 and h_0 share an endpoint at x, and their other endpoints are uniformly close, so h'_0 and h_0 are uniformly close since they lie in a horizontal disk which is uniformly hyperbolic. Similarly, the final segments h'_n and h_n share an endpoint at γ^-_{n+1}, and their other endpoints are uniformly close, so h'_n and h_n are uniformly close. The point $\gamma_i^- = h_{i-1}^+$ is uniformly close to $\gamma_i'^- = h_{i-1}'^+$, and $\gamma_i^+ = h_i^-$ is uniformly close to $\gamma_i'^+ = h_i'^-$, so γ'_i is uniformly close to γ_i (as in the proof of Lemma 6.24). Finally, the endpoints of each h'_i, $1 \leq i \leq n-1$, must be within a uniformly bounded distance of each h_i. □

To prove the fan lemma, it now suffices to show that the substitute paths along the top and bottom of the fan are uniformly close to each other.

Horizontal pieces at x and z are close. When $n \geq 1$, $h_z = h_0$ by construction. Otherwise for $n = 0$, h_z and h_0 share an endpoint at x, and their other endpoints are uniformly close because the ideal triangle $\widetilde{T}_1$ intersects hull (G) (Lemma 6.13). So, in either case, it follows that h_z and h_0 are uniformly close.

Analogous arguments show that h_x and h_k are uniformly close.

Saddle connections γ_i for $i \leq n$ or $i \geq m$. Recall that $\gamma_1 h_1 \gamma_2 h_2 \cdots h_{n-1}\gamma_n = \gamma_1\gamma_2\cdots\gamma_n$ is a continuous path in the fiber over $X_{[\sigma_z]}$. It forms one side of a Euclidean fan whose other sides are γ_z and τ_n — that is, the fan formed by triangles $T_1, \dots, T_n$, which is slim (Claim 6.25). By construction, the saddle connection τ_n is nonparabolic. Because the ideal triangle $\widetilde{T}_{n+1}$ intersects hull (G), the saddle connection τ_n

has uniformly bounded length in the fiber over $X_{[\tau_n]}$ (Corollary 6.14). Since $X_{[\tau_n]}$ is within a uniformly bounded distance of $X_{[\sigma_z]}$ (Corollary 6.21), the saddle connection τ_n has uniformly bounded length in the fiber over $X_{[\sigma_z]}$. Therefore $\gamma_1 \cdots \gamma_n$ is uniformly close to γ_z because they are in $E_{X_{[\sigma_z]}}$ which is uniformly hyperbolic.

Analogous arguments show that for $i \geq m$ the saddle connection γ_x in the top of the fan is uniformly close to the paths $\gamma_m \gamma_{m+1} \cdots \gamma_k$ in the substitute path along the bottom of the fan.

Horizontal pieces h_i for $n+1 \leq i \leq m-2$. The ideal triangles $\widetilde{T}_{n+1}$ and $\widetilde{T}_{m-1}$ are the first and last ideal triangles intersecting the hull, respectively. Define h to be the geodesic in D between the adjusted balance points $\bar{b}_{n+1}$ and $\bar{b}_{m-1}$, and let h_y denote its lift to D_y.[2] Because the ideal triangle $\widetilde{T}_{n+1}$ intersects hull (G), there is a uniformly bounded distance from the endpoint $h_y^- = \bar{b}_{n+1}$ to $B_{[\tau_n]}$ (Lemma 6.13), which lies a uniformly bounded distance from $B_{[\sigma_z]}$ (indeed if $n = 0$, $B_{[\tau_n]} = B_{[\sigma_z]}$; otherwise this follows from Corollary 6.21). Similarly, there is a uniformly bounded distance from the endpoint $h_y^+ = \bar{b}_{m-1}$ to $B_{[\sigma_x]}$. Because $h_y'^- \in \partial B_{[\sigma_z]}$ and $h_y'^+ \in \partial B_{[\sigma_x]}$, the tube lemma (Lemma 6.23) ensures that h_y lies uniformly close to $h_y' \cup B_{[\sigma_z]} \cup B_{[\sigma_x]}$. Therefore $P(h_y)$ and $P(h_y')$ must have uniformly bounded Hausdorff distance.

The aim is to show that $P(h_y)$ and $P(h_{n+1} \cup \cdots \cup h_{m-1})$ have uniformly bounded Hausdorff distance. This is accomplished by breaking h_y into segments, each of which has uniformly bounded Hausdorff distance to its image under P.

Recall that h_i is in the disk $D_{\gamma_i^+}$ connecting the horoballs (or horopoints) $B_{[\sigma_i]}$ and $B_{[\sigma_{i+1}]}$. Define h_i'' to be the geodesic in $D_{\gamma_i^+}$ between (the lifts of) $\bar{b}_i$ and $\bar{b}_{i+1}$, the adjusted balance points of ideal triangles $\widetilde{T}_i$ and $\widetilde{T}_{i+1}$. Because $\widetilde{T}_i \cap \text{hull}\,(G) \neq \varnothing$ and $\widetilde{T}_{i+1} \cap \text{hull}\,(G) \neq \varnothing$, there is a uniform bound on the distance from $\bar{b}_i$ to $B_{[\sigma_i]}$ and from $\bar{b}_{i+1}$ to $B_{[\sigma_{i+1}]}$ (Lemma 6.13). (Note that if there are multiple saddle connections in the same direction, then h_i might be a point which is uniformly close to both endpoints of h_i''.) Then by the tube lemma (Lemma 6.23), h_i'' lies uniformly close (in $D_{\gamma_i^+}$) to $h_i \cup B_{[\sigma_i]} \cup B_{[\sigma_{i+1}]}$. Therefore $P(h_i'')$ and $P(h_i)$ have uniformly bounded Hausdorff distance.

Define g_i to be the geodesic in D_y joining $\bar{b}_i$ and $\bar{b}_{i+1}$. Since $\bar{b}_i$ and $\bar{b}_{i+1}$ are both within a uniformly bounded distance of $B_{[\tau_i]}$ (Lemma 6.13), the saddle connection τ_i has bounded length over the geodesic $[\bar{b}_i, \bar{b}_{i+1}]$ and therefore h_i'' and g_i have bounded Hausdorff distance in E (and in $\overline{E}$).

The following claim will allow h_y to be broken into segments whose images under P will be shown to be uniformly close to each of the $P(g_i)$. Recall that $p : \text{hull}\,(G) \to \widehat{D}$ is the map that collapses horoballs in $\overline{D}$ to points.

Claim 6.27 *There are points $t_{n+1}, \ldots, t_{m-1}$ appearing in order along the geodesic h whose images under $\bar{p} : D \to \widehat{D}$ respectively lie within uniformly bounded distance of the collapsed horoballs* (or *horopoints*) $\bar{p}(B_{[\sigma_{n+1}]}), \ldots, \bar{p}(B_{[\sigma_{m-1}]})$.

[2] Recall from the beginning of the proof that for the sake of simplicity, objects in D and their lifts to various horizontal disks are generally referred to by the same notation. This is the first (and only) instance when an object in D, h, is *notationally* distinguished from one of its lifts, $h_y \subset D_y$.

This statement is functionally analogous to [8, Claim 4.15]. Their proof relies on defining h between two balance points, where the associated Euclidean triangles are equilateral, and using the intermediate value theorem to identify points on h which must be close to the balance points of each of the intermediate triangles in the fan. The proof below relies instead on the structure lemma (Lemma 6.19) to force the same conclusion. Importantly, the approach presented here works in the more general case where h is defined between *adjusted* balance points, where the associated Euclidean triangles might not be equilateral. A key advantage to this construction is that all of the t_i are forced to be in hull (G), which avoids the issue of later needing to project paths back onto the bundle over the hull.

Proof of Claim 6.27 Recall that h is the geodesic joining adjusted balance points $\bar{b}_{n+1}$ and $\bar{b}_{m-1}$. Recall also that for each $i \in \{n+1, \dots, m-1\}$ the ideal triangle $\widetilde{T}_i$ intersects hull (G) by construction. Therefore each adjusted balance point $\bar{b}_i$ is within a uniformly bounded distance of all three horoballs or horopoints associated to $\widetilde{T}_i$ (Lemma 6.13). Choose $t_{n+1} := \bar{b}_{n+1}$ and $t_{m-1} := \bar{b}_{m-1}$, which are close to $B_{[\sigma_{n+1}]}$ and $B_{[\sigma_{m-1}]}$, respectively.

By the structure lemma (Lemma 6.19), h must pass through all of the ideal triangles $\widetilde{T}_i$ for $n+1 \le i \le m-1$; and by convexity of hull (G), h is entirely contained in hull (G). An example is illustrated in Figure 7. Therefore if h intersects any of the geodesics joining $\bar{b}_i$ to the horoballs or horopoints associated to $\widetilde{T}_i$, then setting t_i to be the intersection point satisfies the claim. (If h intersects more than one of these geodesics, then any choice of intersection point will suffice.) If $h \cap \widetilde{T}_i \subset B_{[\sigma_i]}$, then any choice of $t_i \in h \cap \widetilde{T}_i$ will satisfy the claim.

Finally, consider the case in which h avoids intersecting any of the three bounded length geodesics joining $\bar{b}_i$ to the horoballs or horopoints associated to $\widetilde{T}_i$ by passing through both of the horoballs $B_{[\tau_{i-1}]}$ and $B_{[\tau_i]}$. An example is illustrated in Figure 8. By Lemma 6.13 and the tube lemma (Lemma 6.23), the point $\bar{b}_i$ lies within a uniform neighborhood of $B_{[\tau_{i-1}]} \cup B_{[\tau_i]}$ and the subsegment of h between these two horoballs,

$$h_i^\sharp := h \cap \widetilde{T}_i \cap \left(B^\circ_{[\sigma_i]} \cup B^\circ_{[\tau_{i-1}]} \cup B^\circ_{[\tau_i]} \right)^C .$$

Choose t_i to be the closest point on $h_i^\sharp$ to $\bar{b}_i$, which may be in $\partial B_{[\tau_{i-1}]}$ or $\partial B_{[\tau_i]}$. Then $\bar{p}(t_i)$ is uniformly close to $\bar{b}_i$, which is uniformly close to $\bar{p}(B_{[\sigma_i]})$. □

Let s_i be the lift of t_i to D_y, and break h_y into segments

$$h_y^i = [s_i, s_{i+1}], \quad i = n+1, \dots, m-2.$$

By construction, in D_y each $P(s_i)$ is within uniformly bounded distance of $P(B_{[\sigma_i]})$, which is within uniformly bounded distance of $\bar{b}_i$ (note that $P(\bar{b}_i) = \bar{b}_i$). Recalling that g_i is the geodesic in D_y joining $\bar{b}_i$ and $\bar{b}_{i+1}$, this implies that $P(h_y^i)$ is uniformly close to $P(g_i)$. Also recalling that $P(g_i)$ and $P(h_i)$ have uniformly bounded Hausdorff distance, each $P(h_i)$ is uniformly close to each $P(h_y^i)$, so $P(h_{n+1} \cup \cdots \cup h_{m-2})$ is uniformly close to $P(h_y^{n+1} \cup \cdots \cup h_y^{m-2}) = P(h_y)$. Finally, recalling that $P(h_y)$ and $P(h'_y)$ have uniformly bounded Hausdorff distance, $P(h_{n+1} \cup \cdots \cup h_{m-1})$ and $P(h'_y)$ have

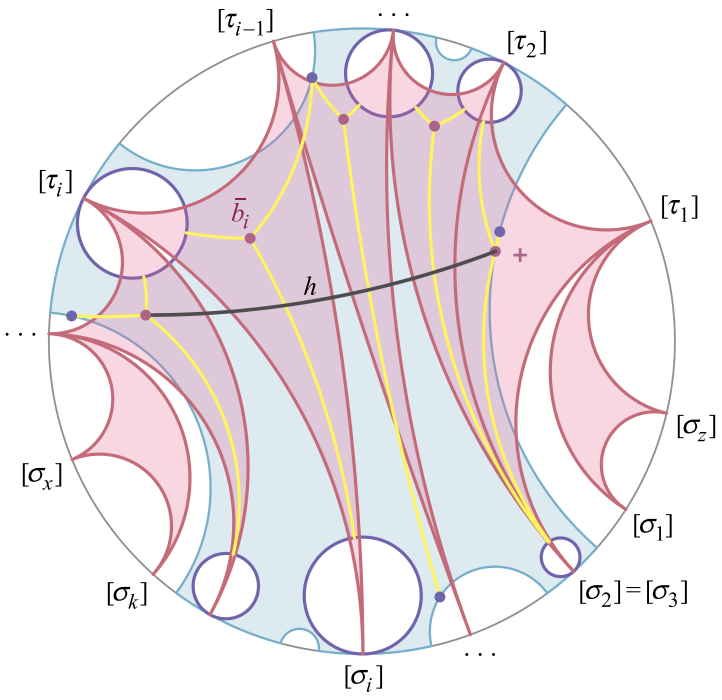


Figure 7: An example of a construction of the $\{t_i\}$ in the proof of Claim 6.27. The ideal fan from Figure 6 (red) has been superimposed on an example of an approximation to a truncated convex hull $\overline{D}$ (blue). Note that in this drawing, $n = 1$ and $m = 1$ since there is one triangle disjoint from the hull at the beginning and end of the fan, so the initial and final triangles are not considered by the claim. For the vertices of the fan at parabolic points, the dark blue circles are the boundaries of the horoballs B_*. For the vertices of the fan at nonparabolic points, the dark blue dots are the horopoints B_*. (Refer to Definition 4.1.) Then for each ideal triangle the adjusted balance point $\bar{b}_*$ is represented by a red dot; in this example, the only triangle with $\bar{b}_* \neq b_*$ is on the right, where the balance point is represented by a red $+$. Each of the adjusted balance points is a uniformly bounded distance from the horoballs and/or horopoints associated to the same triangle (Lemma 6.13), and the shortest distance to each of these is represented by a yellow curve. The geodesic h is drawn in black between adjusted balance points $\bar{b}_{n+1}$ and $\bar{b}_{m-1}$. In this example, all of the t_* are chosen to be the points of intersection of h with the yellow curves.

uniformly bounded Hausdorff distance. Since h'_y was a piece of the substitute path along the top of the fan and the h_i, $n+1 \leq i \leq m-2$ were pieces of the substitute path along the bottom of the fan, this piece of the argument is complete.

Note that in the case where $m - n = 2$ — that is, where only one ideal triangle intersects hull(G) — this argument continues to work despite that the geodesic h is in fact not a path but rather the point $\bar{b}_{n+1} = \bar{b}_{m-1}$. In fact, Claim 6.27 follows immediately from Lemma 6.13.

Horizontal pieces h_n $(n \geq 1)$ and h_{m-1} $(m \leq k)$. When $n \geq 1$, the horizontal path h_n still requires consideration. Conveniently, $P(h_n)$ must have uniformly bounded length. The endpoint γ_n^+ of h_n is joined to γ_z^+ by the saddle connection τ_n in the fiber over $X_{[\sigma_z]}$, where it has uniformly bounded length because

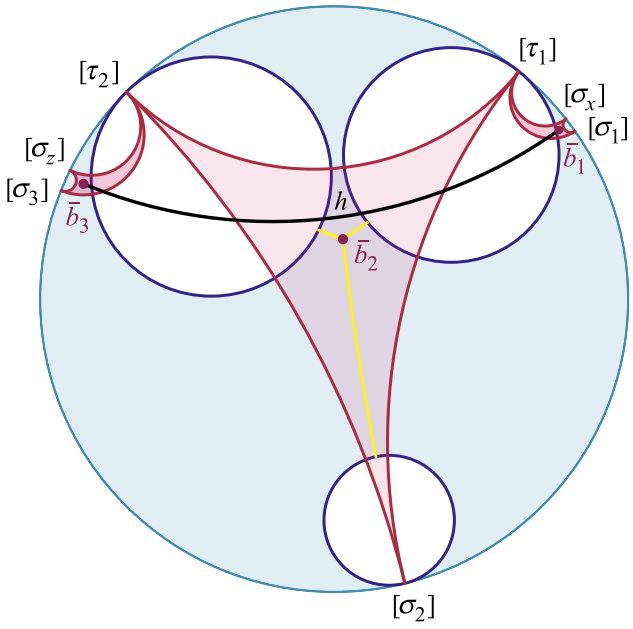


Figure 8: An example of a case in the proof of Claim 6.27. The colors and labels are analogous to those in Figure 7; for simplicity, in this example hull (G) (blue) is assumed to be all of D. Because h (black geodesic) does not intersect any of the yellow curves, t_2 will be chosen to be the point on h and in $\widetilde{T}_2$ (red) which is closest to $\bar{b}_2$; although t_2 and $\bar{b}_2$ may not be uniformly close, their images under $\bar{p}$ must be uniformly close.

the ideal triangle $\widetilde{T}_{n+1}$ intersects hull (G) and because $X_{[\tau_n]}$ is uniformly close to $X_{[\sigma_z]}$ (Corollaries 6.14 and 6.21). Since $\gamma_z^+ = h_y'^-$, the arguments in the previous part of this proof give that $P(\gamma_z^+) = P(h_y'^-)$ is within a uniformly bounded distance of $P(h_y^-) = P(\bar{b}_{n+1})$ in $P(D_y)$, which is within a uniformly bounded distance of the P-image of $B_{[\sigma_{n+1}]} \ni \gamma_{n+1}^+ = h_{n+1}^-$ (Claim 6.27). Lastly, γ_{n+1} has uniformly bounded length in the fiber over $X_{[\sigma_{n+1}]}$ — either it is a nonparabolic saddle connection associated to an ideal triangle which intersects hull (G) (Corollary 6.14), or it is a parabolic saddle connection with bounded length by construction — and $\gamma_{n+1}^- = h_n^+$, so $P(h_n^-)$ and $P(h_n^+)$ must be uniformly close.

When $m \le k$, a similar argument shows that $P(h_{m-1})$ has uniformly bounded length.

Remaining saddle connections. The remaining saddle connections must have uniformly bounded length because their endpoints — which are endpoints of the horizontal paths — have been shown to be uniformly close. However, it is insightful to see these remaining saddle connections handled explicitly.

Any remaining nonparabolic saddle connections must be associated to ideal triangles which intersect the hull, so each of these saddle connections has uniformly bounded length at its associated horopoint (Corollary 6.14). Therefore any isolated saddle connection — that is, any σ_i for which $[\sigma_{i-1}] \ne [\sigma_i] \ne [\sigma_{i+1}]$ — will not prevent slimness of the collapsed preferred paths. (Note that this also resolves any obstruction to slimness posed by the saddle connections γ_z and γ_x in the cases where $n = 0$ and $m = k + 1$, respectively.)

Suppose instead that there are consecutive indices

$$I = \{i_1, \dots, i_j\} \subset \{n+1, n+2, \dots, m-1\}$$

such that $[\sigma_{i_1}] = \cdots = [\sigma_{i_j}]$. Denote this common direction by $[\sigma_I]$. Since the ideal triangles $\widetilde{T}_{i\in I}$ intersect hull (G), all the saddle connections in $T_{i\in I}$ have uniformly bounded length at their associated (adjusted) balance points $\bar{b}_i$ (Corollary 6.10), which are all uniformly close to $X_{[\sigma_I]}$ (Lemma 6.13). Therefore all of the saddle connections associated to $T_{i\in I}$ have uniformly bounded length in the fiber over $X_{[\sigma_I]}$. In particular, because all of the $\gamma_{i\in I}$ form one side of the geodesic triangle $f_{X_{[\sigma_I]}}(T_{i_1} \cup \cdots \cup T_{i_j})$, the triangle inequality gives that

$$\sum_{i\in I} \ell(\gamma_i) \le \ell\left(f_{X_{[\sigma_I]}}(\tau_{i_1-1})\right) + \ell\left(f_{X_{[\sigma_I]}}(\tau_{i_j})\right),$$

which is uniformly bounded above. Therefore any set of consecutive nonparabolic saddle connections does not prevent slimness of the collapsed preferred paths. □

6.4 General triangles

It remains to show that any triangle of collapsed preferred paths is uniformly slim. Equipped with this new version of the fan lemma for finitely generated Veech groups, the subsequent arguments of [8] continue to hold for proving that general triangles of preferred paths are slim. As in Section 4, their results are outlined here and cited wherever their arguments apply essentially verbatim.

Theorem 6.28 (collapsed preferred paths form slim triangles) *There exists $\delta > 0$ so that collapsed preferred paths form δ-slim triangles. That is, for any $x, y, z \in \Sigma$,*

$$\hat{\varsigma}(x, y) \subset N_\delta(\hat{\varsigma}(x, z) \cup \hat{\varsigma}(y, z)).$$

Proof of Theorem 6.28 Let δ' be the constant from the fan lemma (Lemma 6.22). Any triangle in $\Delta(x, y, z)$ which has at least one side consisting of exactly one saddle connection can be decomposed into a union of fans; see Figure 9.

Slimness of each fan and the "furthermore" statement of the fan lemma (Lemma 6.22) ensure that the associated triangle of collapsed preferred paths $\Delta^{\hat{\varsigma}}(x, y, z)$ is $\delta'' = (2\delta'+2)$-slim [8, Lemma 4.16]. General triangles $\Delta(x, y, z)$ may be similarly decomposed into fans, and after a careful treatment of all cases [8, Theorem 4.2] concludes that general triangles of collapsed preferred paths $\Delta^{\hat{\varsigma}}(x, y, z)$ must be $\delta = 3\delta'' = 3(2\delta'+2)$-slim. □

Coincidentally, this gives an alternate proof for special case of a theorem by [11] (generalized by [15]) which states that a virtually free subgroup of the mapping class group is convex cocompact if and only if its extension group is hyperbolic.

Theorem 6.29 (special case of [11; 15]) *Let $G < \mathrm{MCG}(S)$ be a finitely generated Veech group with extension group Γ. If G has no parabolic elements, then Γ is Gromov hyperbolic.*

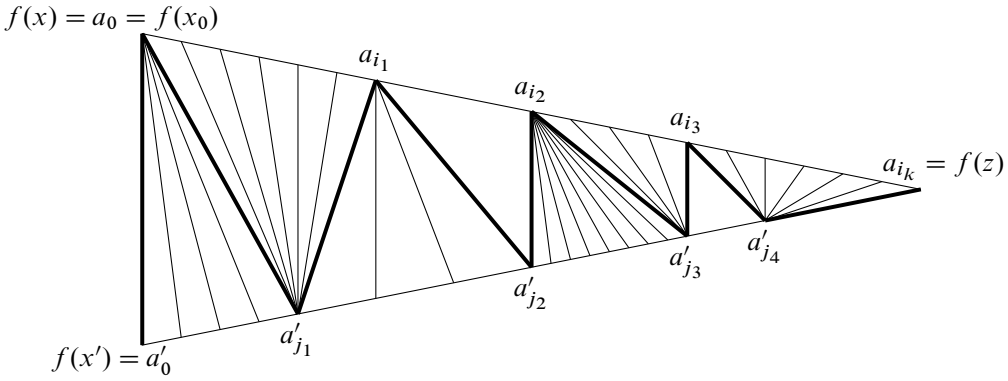


Figure 9: Illustration from [8, Figure 6] of a triangle with exactly one side being a single saddle connection, showing how it can be decomposed into fans. The labeled points are the vertices of the fans.

In the sense of [11], G is convex cocompact as a discrete subgroup of $\mathrm{Isom}(D)$ if it acts cocompactly on the convex hull of its limit set — that is, if the quotient of $\mathrm{hull}\,(G)$ by the action of G is compact. Therefore if G has no parabolic elements, it is convex cocompact as a Fuchsian group. A consequence is that the G-orbit of $X \in \mathrm{hull}\,(G)$ is quasiconvex.

Proof Let $\overline{E}$ be the space constructed in Section 4. Since there are no horoball preimages to collapse, $\overline{E} = \widehat{E}$ and Γ acts isometrically and cocompactly on $\overline{E}$ by construction. It only remains to show that $\overline{E}$ (equipped with the metric $\bar{d}$) is hyperbolic, which is accomplished by applying the guessing geodesics criterion (Proposition 5.1). Here, the sets $L(x, y)$ in the statement of the guessing geodesics criterion are precisely the preferred paths $\varsigma(x, y) = \hat{\varsigma}\,(x, y)$.

It was shown in the construction that $(\overline{E}, \bar{d})$ is a length space. Because the collection Σ of all cone points is Γ-invariant and $\overline{E}/\,\Gamma$ is compact, there exists some constant $R > 0$ so that Σ is R-dense in $\overline{E}$. For $x, y \in \Sigma$ there is a preferred path $\varsigma(x, y) \subset \overline{E}$, and the preferred paths form slim triangles (Theorem 6.28), satisfying condition (1) of the guessing geodesics criterion.

To verify condition (2) of the guessing geodesics criterion, suppose that $\bar{d}\,(x, y) \leq 3R$ for some $x \in E_X$ and $y \in E_Y$. Then $\rho(X, Y) \leq 3R$ in D. Because $f_{X,Y} = f_X|_{E_Y}$ is $e^{\rho(X,Y)}$-bilipschitz (Section 4.2.1), the length of $f_X(\varsigma(x, y))$ is bounded by $3Re^{3R}$. The geodesic in E_X joining x to $f_X(y)$ also has bounded length and in particular is a concatenation of saddle connections each with length bounded above (by $3Re^{3R}$) and below (uniformly, due to Proposition 6.3). So there are at most n saddle connections in $\varsigma(x, y)$, where n depends only on R. Each of these saddle connections σ has length bounded above in the fiber over X and bounded below in the fiber at its associated horopoint $X_{[\sigma]}$, so X and $X_{[\sigma]}$ are uniformly close for all saddle connections σ appearing in $\varsigma(x, y)$. Finally, since $\varsigma(x, y)$ consists of at most $2n + 1$ pieces (hyperbolic geodesics in the horizontal fibers and saddle connections in the vertical fibers), where n depends only on R, and each of those pieces has uniformly bounded length, it follows that $L(x, y) = \varsigma(x, y)$ has bounded length — and therefore bounded diameter, as required. □

7 Hierarchical hyperbolicity

After establishing a nice action of Γ on a hyperbolic space $\widehat{E}$, a natural next step is to prove that Γ is hierarchically hyperbolic.

Theorem 7.1 (hierarchical hyperbolicity) *Let $G < \mathrm{MCG}(S)$ be a finitely generated Veech group with extension group Γ. Then Γ is a hierarchically hyperbolic group.*

Nearly all of [9] continues to hold essentially verbatim when G is finitely generated but not necessarily a lattice: Constructing a hierarchically hyperbolic space structure once $\widehat{E}$ is shown to be hyperbolic is almost solely concerned with what happens with the Bass–Serre trees T_α for parabolic directions α for G, so the existence of nonparabolic directions does not interfere with the construction of the HHS. For this same reason, Γ is already a hierarchically hyperbolic group whenever G has no parabolic limit points; indeed, Γ is hyperbolic by [11; 15] (or by Theorem 6.29). This section is deliberately kept brief; the reader is referred to [9] for more comprehensive coverage of the HHS structure and background information.

Let $\mathcal{X}$ be the flag simplicial complex with 1-skeleton $\mathcal{X}^{(1)}$ as given in [9, Section 4.2]. Let $\mathcal{W}$ be the $\mathcal{X}$-graph whose vertices are the maximal simplices of $\mathcal{X}$ and whose edges are as given in [9, Section 4.2]. Let $\mathcal{X}^{+\mathcal{W}}$ be the $\mathcal{W}$-augmented dual graph as given in [9, Definition 4.2]. The following statement is comparable to [9, Lemma 4.17]. While this proof is nearly identical to the original, it includes consideration for when nontrivial saddle connections might appear in the case that G is not a lattice.

Lemma 7.2 (empty simplex) *The graph $\mathcal{X}^{+\mathcal{W}}$ is quasi-isometric to $\widehat{E}$.*

Proof Starting with the map $Z : \mathcal{X}^{(1)} \to \sqcup T_\alpha$ defined in [9, Section 4.2], extend this to a map $Z' : \mathcal{X}^{+\mathcal{W}} \to \widehat{E}$ as described in the proof of [9, Lemma 4.17]. As in that proof, Z is a one-sided inverse of Z'. It remains to show that Z is coarsely Lipschitz by demonstrating that for any edge $e = [x, y]$ of $\mathcal{X}^{+\mathcal{W}}$ with $v = Z(x)$ and $w = Z(y)$ in $\mathcal{V}$, there exists a path of uniformly bounded length in $\mathcal{X}^{+\mathcal{W}}$ joining v and w.

Now by Lemma 5.4, any $v \in T_\alpha$, $w \in T_\beta$ are joined by a combinatorial path of length proportional to $\hat{d}(v, w)$ which, as the proof shows, is in fact a concatenation of combinatorial subpaths of uniformly bounded length connecting points of $\mathcal{V}$. Moreover, by the lemma, the number of horizontal jumps and saddle connections is bounded by $C'\hat{d}(v, w)$. Therefore it suffices to prove the lemma in the cases that v and w are joined by a bounded length combinatorial path which is either

(a) a single horizontal jump between two horoballs, or

(b) two horizontal jumps from horoballs to a common component of the hull boundary concatenated with nonparabolic saddle connections (which are joined by horizontal jumps in lifts of the hull boundary).

Regardless of the presence of nonparabolic saddle connections, the proof may proceed as in [9] because this combinatorial path is the image of a path whose interior is in $\overline{E}$ and connects the "thickened spine neighborhoods" (see Figure 2(e)) associated to v and w. □

Proof of Theorem 7.1 The proof of [9, Theorem 4.16] holds verbatim after substituting Lemma 7.2 for [9, Lemma 4.17]. Therefore the pair $(\mathcal{X}, \mathcal{W})$ is a combinatorial HHS by the definition found in [9, Definition 4.8]. The rest follows from [9, Theorem 4.11]. □

List of symbols

(S, X, q)	flat surface determined by a closed surface S of genus at least 2, a choice of complex structure X, and its associated quadratic differential q	page 995
$[\sigma]$	direction associated to the saddle connection σ	page 1000
$\overleftrightarrow{[\sigma_x], [\sigma_y]}$	hyperbolic geodesic segment between $[\sigma_x]$ and $[\sigma_y]$	page 1013
$[x, y]$	geodesic segment between x and y	page 992
α	an element of $\mathscr{D}$ (without specific reference to an associated saddle connection)	page 1000
$\mathcal{B}_\alpha^\circ$	interior of the horoball preimage $\mathcal{B}_\alpha$	page 1003
$\mathcal{B}_\alpha$	preimage $\pi^{-1}(B_\alpha)$	page 1003
B_α°	interior of B_α	page 1003
B_α	if α is parabolic, closed horoball in D that is invariant by the maximal parabolic subgroup of G corresponding to α; if α is nonparabolic, orthogonal projection of the point $\alpha \in \partial D$ onto hull(G)	page 1001
$\bar{b}$	adjusted balance point of a Euclidean triangle T (or of its associated ideal triangle $\widetilde{T}$) — that is, the projection of the balance point b to hull(G)	page 1012
b	balance point of a Euclidean triangle T (or of its associated ideal triangle $\widetilde{T}$)	page 1012
$\overline{D}$	truncated convex hull, obtained from hull(G) by deleting the interiors of the horoballs B_α	page 1001
$\bar{d}$	metric on $\overline{E}$	page 1003
$\widehat{D}$	collapsed convex hull, quotient space obtained from $\overline{D}$ by collapsing each horoball B_α to a point	page 1001
$\hat{d}$	metric on $\widehat{E}$	page 1004
$\mathscr{D}$	set of all directions of all saddle connections on (S, X_0, q_0), considered as a subset of ∂D	page 1000
d	metric on E	page 1003
D	Teichmüller disk (determined by a flat structure (X, q), although this is often dropped from the notation)	page 996
$\Delta^\varsigma(x, y, z)$	triangle of preferred paths in E with vertices $x, y, z \in \Sigma$	page 1011
$\Delta^{\hat{\varsigma}}(x, y, z)$	triangle of collapsed preferred paths in $\widehat{E}$ with vertices $x, y, z \in \Sigma$	page 1011

$\Delta(x, y, z)$	triangle of geodesics in E_0 with vertices $f(x)$, $f(y)$, $f(z)$ for $x, y, z \in \Sigma$	page 1011
D_x	preimage $f_X^{-1}(x)$, i.e., the unique lift of hull (G) to the Teichmüller disk in $\mathcal{T}(S)$ through x that covers hull (G) via the projection π	page 1003
$\widehat{E}$	space obtained from $\overline{E}$ by collapsing horoballs to Bass–Serre trees via the map P	page 1004
E	surface bundle over hull (G) given by π; also admits a product structure $E \cong \text{hull}\,(G) \times \widetilde{S}$	page 1001
E_0	fiber of E over X_0	page 1002
E_X	fiber of E over X, i.e., $\pi^{-1}(X)$	page 1002
$f : E \to E_0$	alternative notation for f_{X_0}	page 1003
$f_X : E \to E_X$	map determined by varying $f_{X,Y}$ over all $Y \in D$	page 1003
$f_{X,Y} : E_Y \to E_X$	lift of the Teichmüller map to the universal cover	page 1002
G	finitely generated Veech group of q_0	page 1000
γ_i	i-th saddle piece of $\varsigma(x, y)$	page 1006
Γ	extension group of G	page 1001
h_i	i-th horizontal piece of $\varsigma(x, y)$	page 1006
hull (G)	convex hull of G	page 998
hull (G)	convex hull of the limit set of G	page 1001
ℓ_α	length metric on T_α	page 1008
$\ell\,()$	length of a saddle connection σ	page 995
$L(u, v)$	union of $\hat{\varsigma}(x, y)$ over all $x \in \theta^u \cap \Sigma$ and $y \in \theta^v \cap \Sigma$	page 1007
$N_\delta()$	δ-neighborhood	page 1006
$\overline{P}$	restriction of $P : E \to \widehat{E}$ to $\overline{E}$	page 1004
$\bar{p}$	map $\overline{D} \to \widehat{D}$ which collapses each B_α to a point	page 1001
p	map hull $(G) \to \widehat{D}$ which collapses the interior of each B_α to a point	page 1001
$P : E \to \widehat{E}$	quotient of E obtained by collapsing each $\mathcal{B}_\alpha$ onto T_α via the restriction $t_\alpha\vert_{\mathcal{B}_\alpha}$	page 1004
$\mathcal{P}(\sigma)$	set of saddle connections that span a triangle with σ	page 1015
$\pi : E \to \text{hull}\,(G)$	pullback bundle of the Bers fibration via the inclusion hull $(G) \subset \mathcal{T}(S)$	page 1001
$\text{proj}_{\overline{D}}$	ρ-closest-point projection of D onto $\overline{D}$	page 1001
q	quadratic differential or flat metric for a complex structure in the Teichmüller space of a closed surface of genus at least 2	page 994
q_0	flat metric for the complex structure X_0	page 1000
q_X	flat metric associated to X, i.e., terminal flat metric from the Teichmüller mapping $(S, X_0, q_0) \to (S, X, q_X)$	page 1002
$\bar{\rho}$	induced path metric on the truncated convex hull $\overline{D}$	page 1001
$\hat{\rho}$	quotient pseudometric on $\widehat{D}$	page 1001

ρ	Poincaré metric on the Teichmüller disk D	page 1000
$\widetilde{S}$	universal cover of S	page 1003
S	closed, connected, oriented surface of genus at least 2	page 1000
Σ	the set of all cone points in E, i.e., $\bigcup_{X\in\mathrm{hull}(G)}\Sigma_X$	page 1003
$\hat{\varsigma}(x,y)$	collapsed preferred path between $x, y\in\Sigma$	page 1007
$\varsigma(x,y)$	preferred path between $x, y\in\Sigma$	page 1006
Σ_X	for any $X\in\mathrm{hull}(G)$, the set of cone points of the flat structure q_X on E_X	page 1003
T_α	weighted Bass–Serre tree for a parabolic direction α, i.e., the $\mathbb{R}$-tree dual to the foliation of E_{X_α} in the direction α	page 1004
$t_\alpha : E\to T_\alpha$	composition of f_{X_α} followed by projection to T_α	page 1004
θ_X^v	v-spine in E_X	page 1007
θ^v	union of v-spines over all fibers in the horoball associated to v	page 1007
$\widetilde{T_i}$	i-th triangle in an ideal fan, often the ideal triangle associated to the Euclidean triangle T_i	page 1016
T_i	i-th triangle in a Euclidean fan	page 1016
$\mathcal{V}$	set of all vertices of all Bass–Serre trees T_α	page 1007
X_α	choice of fixed point in ∂B_α; if α is nonparabolic, identical to B_α	page 1001
X_0	fixed choice of complex structure on S	page 1000

Acknowledgements

I would like to thank my PhD advisor, Chris Leininger, for being extraordinarily generous with his time and attention as I prepared this paper. Chris provided valuable insights into existing research and ensured that the final manuscript is as sensible as possible, though neither of us could come up with a less silly term for the "horopoints" described in Definition 4.1 and appearing throughout the paper.

I would like to thank the reviewer for thoughtful comments on the preprint.

Finally, I would like to thank my friends and colleagues for their support (in every sense) throughout the final months of preparing this paper, during which I was without a home after my apartment caught fire. This paper could not have been completed in a timely manner — and, more generally, the last several months of my life would have been immeasurably more difficult — if not for their generosity. I hope to be able to pay it forward someday.

I am financially supported in part by NSF-1842494 and NSF DMS-1745670.

References

[1] **A F Beardon**, *The geometry of discrete groups*, Graduate Texts in Mathematics 91, Springer (1983) MR

[2] **L Bers**, *Fiber spaces over Teichmüller spaces*, Acta Math. 130 (1973) 89–126 MR

[3] **M Bestvina**, *Real trees in topology, geometry, and group theory*, preprint (1997) arXiv math/9712210

[4] **J S Birman**, *Braids, links, and mapping class groups*, Annals of Mathematics Studies 82, Princeton Univ. Press (1974) MR

[5] **B H Bowditch**, *Uniform hyperbolicity of the curve graphs*, Pacific J. Math. 269:2 (2014) 269–280 MR
[6] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR
[7] **R Charney**, **J Crisp**, *Relative hyperbolicity and Artin groups*, Geom. Dedicata 129 (2007) 1–13 MR
[8] **S Dowdall**, **M G Durham**, **C J Leininger**, **A Sisto**, *Extensions of Veech groups, I*: *A hyperbolic action*, J. Topol. 16:2 (2023) 757–805 MR
[9] **S Dowdall**, **M G Durham**, **C J Leininger**, **A Sisto**, *Extensions of Veech groups, II*: *Hierarchical hyperbolicity and quasi-isometric rigidity*, Comment. Math. Helv. 99:1 (2024) 149–228 MR
[10] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Mathematical Series 49, Princeton Univ. Press (2012) MR
[11] **B Farb**, **L Mosher**, *Convex cocompact subgroups of mapping class groups*, Geom. Topol. 6 (2002) 91–152 MR
[12] **F P Gardiner**, *Teichmüller theory and quadratic differentials*, Wiley, New York (1987) MR
[13] **F P Gardiner**, **N Lakic**, *Quasiconformal Teichmüller theory*, Mathematical Surveys and Monographs 76, Amer. Math. Soc., Providence, RI (2000)
[14] **D Groves**, **J F Manning**, *Dehn filling in relatively hyperbolic groups*, Israel J. Math. 168 (2008) 317–429 MR
[15] **U Hamenstaedt**, *Word hyperbolic extensions of surface groups*, preprint (2005) arXiv math/0505244
[16] **P Hubert**, **T A Schmidt**, *An introduction to Veech surfaces*, from "Handbook of Dynamical Systems" (B Hasselblatt, A Katok, editors), volume 1, Elsevier (2006) 501–526
[17] **S Katok**, *Fuchsian groups*, University of Chicago Press (1992) MR
[18] **C J Leininger**, **S Schleimer**, *Hyperbolic spaces in Teichmüller spaces*, J. Eur. Math. Soc. 16:12 (2014) 2669–2692 MR
[19] **A Margolis**, *The geometry of groups containing almost normal subgroups*, Geom. Topol. 25:5 (2021) 2405–2468 MR
[20] **H Masur**, *Geometry of Teichmüller space with the Teichmüller metric*, from "Geometry of Riemann surfaces and their moduli spaces" (L Ji, S A Wolpert, S-T Yau, editors), Surv. Differ. Geom. 14, International, Somerville, MA (2009) 295–313 MR
[21] **H Masur**, **S Schleimer**, *The geometry of the disk complex*, J. Amer. Math. Soc. 26:1 (2013) 1–62 MR
[22] **H Masur**, **S Tabachnikov**, *Rational billiards and flat structures*, from "Handbook of Dynamical Systems" (B Hasselblatt, A Katok, editors), volume 1, Elsevier (2002) 1015–1089
[23] **L Mosher**, *Problems in the geometry of surface group extensions*, from "Problems on mapping class groups and related topics" (B Farb, editor), Proc. Sympos. Pure Math. 74, Amer. Math. Soc., Providence, RI (2006) 245–256 MR
[24] **D-M Nguyen**, *Topological Veech dichotomy and tessellations of the hyperbolic plane*, Israel J. Math. 249:2 (2022) 577–616 MR
[25] **J G Ratcliffe**, *Foundations of hyperbolic manifolds*, 2nd edition, Graduate Texts in Mathematics 149, Springer (2006) MR
[26] **J Russell**, *Extensions of multicurve stabilizers are hierarchically hyperbolic*, Geom. Topol. 29:6 (2025) 3187–3240
[27] **K Strebel**, *Quadratic differentials*, Ergebnisse der Math. 5, Springer (1984) MR
[28] **R Tang**, *Affine diffeomorphism groups are undistorted*, J. Lond. Math. Soc. (2) 104:2 (2021) 747–769 MR
[29] **W P Thurston**, *The geometry and topology of three-manifolds*, lecture notes (1980) Available at http://library.msri.org/books/gt3m/
[30] **W P Thurston**, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Amer. Math. Soc. (N.S.) 19:2 (1988) 417–431 MR
[31] **J Väisälä**, *Gromov hyperbolic spaces*, Expo. Math. 23:3 (2005) 187–231 MR
[32] **W A Veech**, *Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards*, Invent. Math. 97:3 (1989) 553–583 MR
[33] **Y B Vorobets**, *Plane structures and billiards in rational polygons*: *the Veech alternative*, Uspekhi Mat. Nauk 51:5(311) (1996) 3–42 MR In Russian; translated in Russian Math. Surveys 51 (1996), 779–817

ELIOT BONGIOVANNI bongio@umich.edu
Department of Mathematics, University of Michigan, Ann Arbor, MI, United States

Received: July 9, 2024 Revised: January 15, 2025

msp

Algebraic & Geometric Topology **26**:3 (2026) 1037–1076
DOI: 10.2140/agt.2026.26.1037
Published: April 1, 2026

Primitive Feynman diagrams and the rational Goussarov–Habiro Lie algebra of string links

BRUNO DULAR

Goussarov–Habiro's theory of clasper surgeries defines a filtration of the monoid of string links $L(m)$ on m strands, in a way that geometrically realises the Feynman diagrams appearing in low-dimensional and quantum topology. Concretely, $L(m)$ is filtered by C_n-equivalence, for $n \geq 1$, which is defined via local moves that can be seen as higher-order crossing changes. The graded object associated to the Goussarov–Habiro filtration is the *Goussarov–Habiro Lie algebra of string links* $\mathcal{L}L(m)$. We give a concrete presentation, in terms of primitive Feynman (tree) diagrams and relations (1T, AS, IHX, STU^2), of the rational Goussarov–Habiro Lie algebra $\mathcal{L}L(m)_{\mathbb{Q}}$ and of the primitive Lie algebra of the Hopf algebra of Feynman diagrams. To that end, we investigate cycles in *graphs of forests*: *flip graphs* associated to forest diagrams and their STU relations. As an application, we give an alternative *diagrammatic* proof of Massuyeau's rational version of the Goussarov–Habiro conjecture for string links, which relates indistinguishability under finite type invariants of degree $< n$ and C_n-equivalence.

1 Introduction

1.1 Background

A *string link on m strands in the cylinder* is a smooth embedding of a disjoint union of m intervals into the cylinder

$$\gamma : \{1, \dots, m\} \times [0,1] \hookrightarrow \mathbb{D}^2 \times [0,1]$$

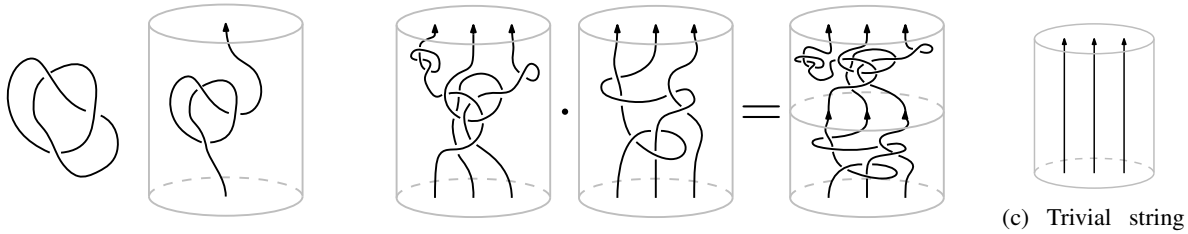
such that $\gamma(i,e) = (x_i, e)$ for all $i \in \{1,\dots,m\}$ and $e \in \{0,1\}$, where $x_1, \dots, x_m \in \mathbb{D}^2$ are fixed points (usually ordered along the x-axis), and γ is perpendicular to the boundary near those points. String links on m strands are considered up to smooth isotopy, and $L(m)$ denotes the corresponding set of isotopy classes, also called *string links* when there is no ambiguity.

The set $L(m)$ is a *monoid* with multiplication given by vertical concatenation (Figure 1(b)), with the trivial string link as unit (Figure 1(c)).

When $m = 1$, elements of $L(1)$ are called *long knots* and *closing a long knot* yields an isomorphism between $L(1)$ and the monoid of knots under connected sum (Figure 1(a)). Thus, the study and classification of string links naturally belong to knot theory, and one of the main tools for probing them is the use of *invariants*.

A string link invariant $V : L(m) \to A$ valued in an abelian group A is a *Vassiliev invariant* or *finite type invariant* of *degree n* if it satisfies specific *skein relations* (see Vassiliev [43], Birman and Lin [3] and Bar-Natan [1]), which can be read as the condition that V vanishes on the $(n+1)$-st power of the

MSC2020: primary 57K16; secondary 16T30, 17B70.



(a) A knot and a long knot

(b) Two string links and their concatenation in $L(3)$

(c) Trivial string link in $L(3)$

Figure 1: Long knots and string links.

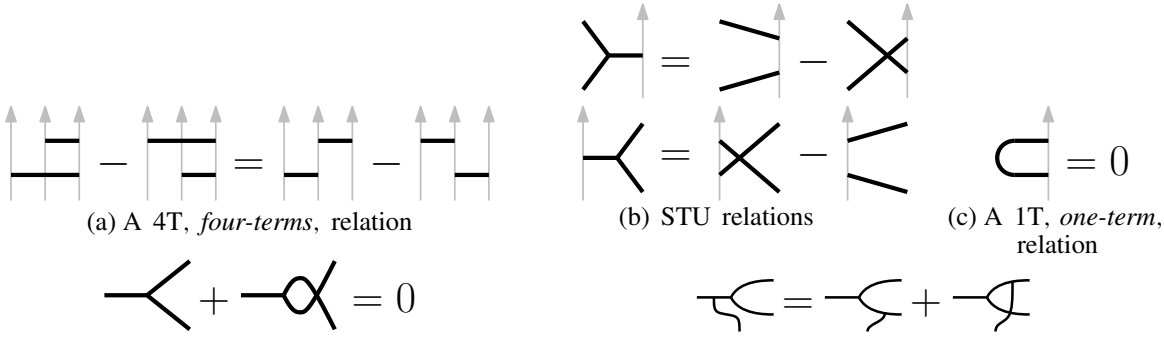
augmentation ideal in the monoid ring $\mathbb{Z}L(m)$. Vassiliev invariants dominate (dominate means that the collection of all Vassiliev invariants determines the mentioned known invariants) polynomial invariants (Conway, HOMFLY, etc.), quantum invariants (see Turaev [42]), Milnor invariants (see the survey by Meilhan [33]), etc. but it is an open question whether they separate string links (or even knots, in the case $m = 1$). One of the main results in that theory, due to Bar-Natan [1] and Kontsevich [25], is that any $\mathbb{Q}$-valued Vassiliev invariant factors through the Kontsevich integral

$$Z : L(m) \to \widehat{\mathcal{A}^{\mathrm{FI}}}(m)_{\mathbb{Q}}, \tag{1-1}$$

which is valued in the graded completion of the Hopf algebra $\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}}$ of *Feynman diagrams* on m strands, generated by uni-trivalent diagrams (with cyclic orientation at the trivalent vertices (the cyclic orientation is usually omitted, in which case it is assumed to be counterclockwise), and univalent vertices attached to m vertical strands; see Figure 4), modulo 1T and STU relations (which imply the relations 4T, AS, IHX; see Figure 2). The presence of the 1T relation translates the fact that we are working with *unframed string links*. This justifies the notation FI for *framing independence*, used throughout this paper.

For the above reason, the Kontsevich integral is called a *universal Vassiliev invariant over* $\mathbb{Q}$. The theory is much less understood over the integers.

A natural question about Vassiliev invariants is: what do they detect? In the case of knots ($m = 1$), Goussarov [17] and Habiro [22] independently answered this question by introducing C_n*-moves*, defined



(a) A 4T, *four-terms*, relation

(b) STU relations

(c) A 1T, *one-term*, relation

(d) An AS relation

(e) An IHX relation in the *Kirchhoff form* [8, Section 5.2.7]

Figure 2: The 4T, STU, 1T, AS and IHX relations.

(a) A C_1-move (b) A C_2-move

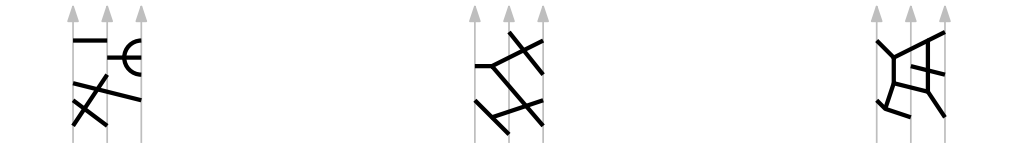
Figure 3: Diagrams, claspers realising them and C_n-moves.

using *clasper surgeries*. Those are local moves on string links that are modelled on *tree claspers*, ribbon trees with leaves attached to the strands, that geometrically realise the Feynman diagrams from above. See Figure 3 for examples of C_1- and C_2-moves. More precisely, C_n-moves generate the C_n-equivalence relation on $L(m)$, and they show that any two C_n-equivalent string links are V_n-equivalent, i.e., they cannot be distinguished by Vassiliev invariants of degree $< n$. Goussarov and Habiro conjecture that the converse holds as well: this is the *Goussarov–Habiro conjecture for string links in the cylinder* (which they prove for knots, i.e., the case $m = 1$). This is still wide open, with only progress made in small degrees by Meilhan and Yasuhara [34].

For each $n \geq 1$, the monoid $\mathcal{L}_n L(m) := L_n(m)/C_{n+1}$ of C_{n+1}-equivalence classes of C_n-trivial string links turns out to be a finitely generated abelian group. See the works of Goussarov, Polyak and Viro [18] and of Habiro [22]. Those combine into a graded Lie $\mathbb{Z}$-algebra $\mathcal{L}L(m) := \bigoplus_{n\geq 1} \mathcal{L}_n L(m)$, the *Goussarov–Habiro Lie algebra of string links on m strands*. From their work, it follows that there is a surjective $\mathbb{Z}$-linear graded *realisation map* $R : \mathbb{Z}\mathcal{D}^T(m) \twoheadrightarrow \mathcal{L}L(m)$ that sends a tree diagram to the result of surgery along a tree clasper realising it. Thus there is a presentation of the Goussarov–Habiro Lie algebra by tree diagrams on m strands modulo specific relations

$$\mathcal{L}L(m) \cong \mathbb{Z}\mathcal{D}^T(m)/\ker(R),$$

whose understanding would be a significant step towards the Goussarov–Habiro conjecture, while also being of its own interest. For example, it would provide a "*universal Goussarov–Habiro invariant*", valued in a Lie algebra of tree diagrams. This is related to the study of π_0 of the embedding calculus tower, as appearing in the works of Conant [9], Budney, Conant, Koytcheff and Sinha [5], and Kosanović [26]. The concordance analogue of this question, where C_n-equivalence is replaced by C_n*-concordance*, appears in works of Conant, Schneiderman and Teichner [11; 12], while the homotopy analogue has been settled by Habegger, Lin and Masbaum [20; 21]. Habiro and Massuyeau studied a similar question in the context of homology cylinders in [23]. Nozaki, Sato and Suzuki also studied $\ker(R)$ in [37].



(a) Chord diagram (b) Forest diagram (c) Feynman diagram

Figure 4: Examples of diagrams.

1.2 Results and strategy

The main contribution of the present paper is an explicit presentation of the *rational* Goussarov–Habiro Lie algebra $\mathcal{L}L(m)_{\mathbb{Q}} := \mathcal{L}L(m) \otimes \mathbb{Q}$, and of the primitive Lie algebra of the Hopf algebra of Feynman diagrams $\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}}$. In other words, we identify $\ker(R)_{\mathbb{Q}}$ as the subgroup generated by specific relations in $\mathbb{Q}\mathcal{D}^T(m)$ (where $\mathcal{D}^T(m)$ is the set of tree diagrams on m strands): the usual 1T, AS, IHX relations, completed with Conant's STU^2 relation; see (1-2).

Theorem 1 *The three following graded Lie $\mathbb{Q}$-algebras are isomorphic to each other*:

(i) *the Goussarov–Habiro Lie algebra of string links $\mathcal{L}L(m)_{\mathbb{Q}}$*;

(ii) *the primitive Lie algebra $\mathrm{Prim}(\mathcal{A}^{\mathrm{FI}}(m))_{\mathbb{Q}}$ of the Hopf algebra of Feynman diagrams*;

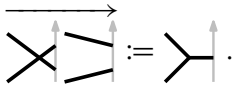
(iii) *the Lie algebra of tree diagrams $\mathcal{L}^{\mathrm{FI}}(m)_{\mathbb{Q}}$ defined as $\mathbb{Q}\mathcal{D}^T(m)/\langle 1\mathrm{T}, \mathrm{AS}, \mathrm{IHX}, \mathrm{STU}^2\rangle$*.

The isomorphism $\mathrm{Prim}(\mathcal{A}^{\mathrm{FI}}(m))_{\mathbb{Q}} \cong \mathcal{L}^{\mathrm{FI}}(m)_{\mathbb{Q}}$ is Theorem 2.4.5.1 in the text and is the focus of Section 2, which we outline now. It will be more convenient to see the algebra $\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}}$ as being generated by *forest diagrams*, i.e., Feynman diagrams without cycles, modulo 1T and STU. First, we prove that $\mathrm{Prim}(\mathcal{A}^{\mathrm{FI}}(m))_{\mathbb{Q}}$ is generated by tree diagrams. To that end, we filter $\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}}$ by the *size* of forests, and we show that this size filtration coincides with the primitive filtration (Theorem 2.2.2.3). This requires averaging over permutations and is one of the main reasons why we cannot extend Theorem 1 over $\mathbb{Z}$ with the current methods.

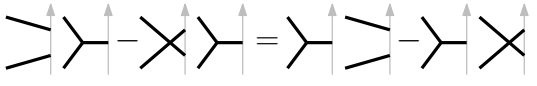
However, $\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}}$ is defined using the STU relation, which is not *size homogeneous*: it relates three forest diagrams, two of which with one more tree than the other. Here lies the main difficulty in finding a concrete presentation of $\mathrm{Prim}(\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}})$. This is resolved by introducing *graph of forests*, which are *flip graphs* with forest diagrams as vertices and with an edge for each *slide move*

$F =$  slide move $F' =$  ,

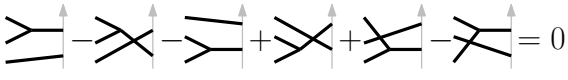
where it is understood that F and F' are identical outside of the shown part. Identifying relations in $\mathrm{Prim}(\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}})$ becomes a matter of finding cycles in such graphs of forests, and the STU relation becomes a way of assigning forests of smaller size to those cycles, via the assignment



We show that the first homology groups of graphs of forests with rational coefficients are generated by cycles of length 4 and cycles of length 6, respectively responsible for STU^2-relations

(1-2)  ,

where the four terms represent forest diagrams that are identical outside of the shown parts, and ⎔-relations

(1-3) 

In particular, the legs involved in each term of a $\hexagon$-relation must belong to two distinct trees, for otherwise another term would contain a cycle. Thus, the $\hexagon$-relations do not appear in the presentation of $\mathrm{Prim}(\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}})$, but they do appear in the presentation of the intermediate steps of the size filtration (more precisely, in size $n-1$ and degree n); see Section 2.4.

The isomorphism $\mathcal{L}^{\mathrm{FI}}(m)_{\mathbb{Q}} \cong \mathcal{L}L(m)_{\mathbb{Q}}$ follows from Theorems 3.1.3.2 and 3.2.2.1 in the text and is the focus of Section 3. In Theorem 3.1.3.2, we define a *realisation map* $\mathbb{Z}\mathcal{D}^T(m) \to \mathcal{L}L(m)$, which sends a tree diagram to the result of *clasper surgery* along a tree clasper realising it. This map is surjective and indeed satisfies the 1T, AS, IHX and STU^2 relations, by an application of clasper calculus. Similar realisation maps and clasper calculus computations appear in works of Ohtsuki, Conant, Teichner, Schneiderman and Kosanović [10; 13; 16; 26; 38], and hence we omit most details here.

By the above, there is a surjective realisation map $R : \mathcal{L}^{\mathrm{FI}}(m)_{\mathbb{Q}} \twoheadrightarrow \mathcal{L}L(m)_{\mathbb{Q}}$. To show that it is injective, we use the *Kontsevich integral* (1-1) [1; 25]. The Kontsevich integral indeed detects tree clasper surgeries, and hence it provides an inverse Z^{GH} to R, as follows from the following result.

Theorem 2 (tree preservation theorem (Theorem 3.2.2.1)) *Let $n \geq 1$ and $T \in \mathcal{D}_n^T(m)$ a degree-n tree diagram on m strands. Then*

$$Z(\sigma(C_T)) = 1 + T + O(n+1), \tag{1-4}$$

where $\sigma(C_T)$ denotes the string link obtained from the trivial one by surgery along a tree clasper C_T realising T, and $O(n+1)$ means that the equality holds modulo terms of degree $\geq n+1$.

A proof of this result is sketched by Ohtsuki in [38, Proposition E.24] but we provide a different detailed proof. The use of the Kontsevich integral is another main reason why we cannot extend Theorem 1 over $\mathbb{Z}$.

In Section 4, we apply Theorem 1 to give an alternative proof of Massuyeau's rational version of the Goussarov–Habiro conjecture for string links.

1.3 Further directions

We restricted our attention to unframed string links in order to make the presentation clearer. However, the identification of the primitive Lie algebra of $\mathcal{A}(m)_{\mathbb{Q}}$ holds with and without the 1T relation. By considering the Kontsevich integral for *framed string links*, as described by Le and Murakami [30], one should obtain a presentation of the framed version of the Goussarov–Habiro Lie algebra.

Conant's work [9] on the STU^2 relation was an important ingredient for identifying modules in the second page of the spectral sequence arising in the study of the *Goodwillie–Weiss embedding calculus tower*, as in the works of Budney, Conant, Koytcheff and Sinha [5], Kosanović [26] and Shi [41]. The present work motivates the investigation of the string link analogue of the embedding calculus tower.

Moreover, the Goodwillie–Weiss embedding calculus tower is known to be a universal Vassiliev invariant over $\mathbb{Z}_{(p)}$ in degree $\leq p+1$, by a theorem of Boavida de Brito and Horel [4, Theorem A]. Combining a *string link* version of this result with the present work could lead to a proof of the Goussarov–Habiro conjecture over $\mathbb{Z}_{(p)}$ in degree $n < p$.

1.4 Notation

Lower indices usually refer to *degrees* of diagrams and claspers, and upper indices usually refer to filtrations of diagrams by *size* (number of connected components). The notation FI, for *framing independence*, refers to the presence of the relation 1T. Sets of diagrams are denoted by $\mathcal{D}$ ornamented with various indices, Hopf algebras by $\mathcal{A}$ and Lie algebras by $\mathcal{L}$.

2 Primitive Lie algebra of Feynman diagrams

The goal of this section is to identify the primitive Lie algebra $\mathrm{Prim}(\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}})$ of the Hopf algebra $\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}}$ of Feynman diagrams on m strands.

In the case of knots, i.e., $m = 1$, tree diagrams generate the primitive Lie algebra $\mathrm{Prim}(\mathcal{A}^{\mathrm{FI}}(1)_{\mathbb{Q}})$ of $\mathcal{A}^{\mathrm{FI}}(1)_{\mathbb{Q}}$ [1], which, by [9], is isomorphic to the space of trees

$$\mathbb{Q}\mathcal{D}^{T}(1)/\langle 1\mathrm{T}, \mathrm{AS}, \mathrm{IHX}, \mathrm{STU}^2\rangle.$$

See also [26, Theorem G3] for a concise presentation of these results. Commutativity of $\mathcal{A}^{\mathrm{FI}}(1)_{\mathbb{Q}}$ implies, by the Milnor–Moore theorem, that there is an isomorphism between the primitive part $\mathrm{Prim}(\mathcal{A}^{\mathrm{FI}}(1)_{\mathbb{Q}})$ and the inseparable quotient $(\mathcal{A}^{\mathrm{FI}}(1)_{\mathbb{Q}})^{I}$ (where any diagram which is a product of nontrivial diagrams is set to zero). This fact is essential in Conant's result [9], which actually holds over $\mathbb{Z}$.

In the case of string links, tree diagrams also generate the primitive part $\mathrm{Prim}(\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}})$. However, $\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}}$ is not commutative when $m > 1$ and the argument of Conant does not translate a priori. In this section, we prove that the conclusion still holds: $\mathrm{Prim}(\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}})$ is isomorphic to the Lie algebra $\mathcal{L}^{\mathrm{FI}}(m)_{\mathbb{Q}} \coloneqq \mathbb{Q}\mathcal{D}^{T}(m)/\langle 1\mathrm{T}, \mathrm{AS}, \mathrm{IHX}, \mathrm{STU}^2\rangle$.

2.1 Algebras of diagrams

This section recalls the definition of chord and Feynman diagrams [1] on tangle skeletons, as well as their bialgebra structure.

2.1.1 Uni-trivalent graphs A *uni-trivalent graph*, also called *Feynman* or *Jacobi graph*, is an undirected graph $G = (V = U \sqcup T, E)$ consisting of univalent and trivalent vertices (forming the sets U and T, respectively). The edges are defined as pairs of distinct vertices.[1] The trivalent vertices are endowed with a cyclic orientation of their three adjacent edges. When a uni-trivalent graph is drawn, the cyclic orientations are the counterclockwise ones unless indicated otherwise. Also, four-valent vertices in drawings are not part of the graph, they are only artefacts of their planar representations.

The univalent and trivalent vertices are often called *leaves* and *nodes*, respectively. An edge adjacent to a leaf is called a *leg*.

The *degree* of a uni-trivalent graph G is $\deg(G) \coloneqq \frac{|V|}{2}$.

A uni-trivalent graph G is a *forest* if it has no cycles and a *tree* if in addition it has a single component. If G is a tree, then $\deg(G) = |U| - 1 = |T| + 1$.

[1] We could allow diagrams with edges attached to a single vertex, i.e., *tadpoles*, but those will vanish by AS.

2.1.2 Uni-trivalent forest and chord diagrams Let S be an oriented 1-manifold with boundary. In Section 3.2.1, S will be the underlying manifold of what will be referred to as a *tangle skeleton*. Since this work focuses on string links, S will usually be a disjoint union of intervals $\{1,\dots,m\}\times[0,1]$.

A *uni-trivalent* or *Feynman* or *Jacobi diagram* on S is a pair $D=(G,[u])$ where $G=(U\sqcup T,E)$ is a uni-trivalent graph and $[u]$ is the isotopy class of an embedding $u:U\hookrightarrow S-\partial S$ of the univalent vertices into the interior of S. The set of Feynman diagrams on S is denoted $\mathcal{D}^F(S)$.

The *degree* of a Feynman diagram is the degree of its underlying graph. Denote by $\mathcal{D}^F_n(S)\subset\mathcal{D}^F(S)$ the subset of degree-n Feynman diagrams. This endows $\mathcal{D}^F(S)$ with a graded set structure.

A Feynman diagram is a *forest*, resp. *tree*, diagram if its underlying graph is a forest, resp. a tree. A forest diagram is a *chord diagram* if it has no nodes. Denote by $\mathcal{D}(S)$, $\mathcal{D}^T(S)$, $\mathcal{D}^c(S)$ the corresponding graded subsets of $\mathcal{D}^F(S)$. The *size* of a forest diagram is its number of trees, i.e., the number of connected components of the underlying uni-trivalent graph. For example, the forest diagram on Figure 4(b) has size 3. For $s\geq 0$, the set of forest diagrams of size s is denoted $\mathcal{D}^s(S)$. Note that $\mathcal{D}^n_n(S)=\mathcal{D}^c_n(S)$ and $\mathcal{D}^T(S)=\mathcal{D}^1(S)$.

By [1], the inclusions $\mathcal{D}^c(S)\subset\mathcal{D}(S)\subset\mathcal{D}^F(S)$ induce isomorphisms[2] of graded $\mathbb{Z}$-modules

$$\mathcal{A}(S):=\frac{\mathbb{Z}\mathcal{D}^c(S)}{\langle 4\mathrm{T}\rangle}\cong\frac{\mathbb{Z}\mathcal{D}(S)}{\langle \mathrm{STU}\rangle}\cong\frac{\mathbb{Z}\mathbb{D}^F(S)}{\langle \mathrm{STU}\rangle} \tag{2-1}$$

and the relations AS, IHX are consequences of STU (see Figure 2). Unless stated otherwise, we use the presentation of $\mathcal{A}(S)$ using forest diagrams. The degree-n part of $\mathcal{A}(S)$ is denoted $\mathcal{A}_n(S)$. The only degree-0 diagram is the empty one, and hence $\mathcal{A}_0(S)\cong\mathbb{Z}$. In the present paper, we mostly consider those graded modules with rational coefficients, namely $\mathcal{A}(S)_{\mathbb{Q}}:=\mathcal{A}(S)\otimes\mathbb{Q}$, whose degree-$n$ part is $\mathcal{A}_n(S)_{\mathbb{Q}}=\mathcal{A}_n(S)\otimes\mathbb{Q}$.

2.1.3 Framed versus unframed There is an *unframed* or *framing independent* (FI) version of the module of Feynman diagrams

$$\mathcal{A}^{\mathrm{FI}}(S):=\mathcal{A}(S)/\langle 1\mathrm{T}\rangle, \tag{2-2}$$

where 1T is the graded submodule generated by all diagrams containing an *isolated chord*, i.e., a chord whose endpoints lie on the same strand with no other endpoint in between them.

Writing $\mathcal{A}^{(\mathrm{FI})}(S)$ indicates that the statement holds for both $\mathcal{A}(S)$ and $\mathcal{A}^{\mathrm{FI}}(S)$.

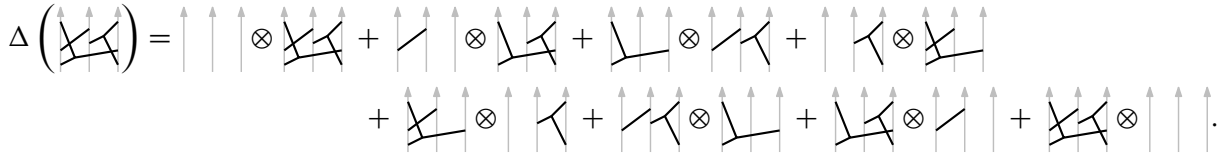
2.1.4 Coalgebra structure The graded $\mathbb{Z}$-module $\mathcal{A}^{(\mathrm{FI})}(S)$ has a structure of coalgebra, with comultiplication [1, Definition 3.7]

$$\Delta(D):=\sum_{J\subseteq\pi_0(D)} D_J\otimes D_{\pi_0(D)-J},$$

[2] In [1], it is shown that $\mathcal{D}^c(S)\hookrightarrow\mathcal{D}^F(S)$ induces an isomorphism $\mathbb{Z}\mathcal{D}^c(S)/\langle 4\mathrm{T}\rangle\cong\mathbb{Z}\mathcal{D}^F(S)/\langle \mathrm{STU}\rangle$. Thus, we get $\mathbb{Z}\mathcal{D}^c(S)/\langle 4\mathrm{T}\rangle\hookrightarrow\mathbb{Z}\mathcal{D}(S)/\langle \mathrm{STU}\rangle\twoheadrightarrow\mathbb{Z}\mathcal{D}^F(S)/\langle \mathrm{STU}\rangle$ and the first map is surjective since any forest can be written as a sum of chord diagrams by applying STU until it has no node left. Hence the first map is an isomorphism, so is the second since their composition is an isomorphism. The same holds with FI added everywhere.

where the sum runs over all the subsets J of the set $\pi_0(D)$ of connected components of D and D_J denotes the subdiagram of D consisting of those components contained in J. The counit ϵ is given by $\epsilon(D) = 1$ if D is the empty diagram, and it is zero otherwise.

For example, with $S = \{1, 2, 3\} \times [0, 1]$, one has



The comultiplication is coassociative and cocommutative.

2.1.5 Hopf algebra of diagrams on string links When $S = \{1, \dots, m\} \times [0, 1]$ is a disjoint union of intervals, we write $\mathcal{D}(m)$ for $\mathcal{D}(S)$, $\mathcal{A}(m)$ for $\mathcal{A}(S)$, $\mathcal{A}(m)_{\mathbb{Q}}$ for $\mathcal{A}(S)_{\mathbb{Q}}$ and so on. In that case, vertical stacking of diagrams endows $\mathcal{A}(m)$ and $\mathcal{A}^{\mathrm{FI}}(m)$ with a multiplication. More precisely, $D \cdot D'$ denotes the vertical concatenation of D and D', with D above D'. The unit 1 is the empty diagram.

Proposition 2.1.5.1 (Bar-Natan [1, Proposition 3.9]) *The comultiplication and multiplication defined above endow $\mathcal{A}^{(\mathrm{FI})}(m)$ with the structure of a connected cocommutative Hopf algebra. It is commutative when $m = 1$.*

2.2 Primitive filtration of forest diagrams

2.2.1 Primitive filtration of a Hopf algebra Let us recall some basic notions about primitive elements in Hopf algebras. See [6; 35; 36, Chapter 5; 40] for more details.

Fix a Hopf algebra A over $\mathbb{Q}$ with multiplication μ or $\cdot$, unit 1, comultiplication Δ and counit ϵ. An element $x \in A$ is *primitive* if $\Delta(x) = x \otimes 1 + 1 \otimes x$. A direct calculation shows that the commutator of two primitive elements is primitive as well. In particular, the set $\mathrm{Prim}(A)$ of primitive elements in A is a Lie algebra, called the *primitive Lie algebra of A*. Conversely, to any Lie algebra L one can associate its *universal enveloping algebra* $U(L)$, which is a Hopf algebra. The Milnor–Moore theorem states conditions under which the functors Prim and U are inverses of each other.

Theorem 2.2.1.1 (Milnor–Moore theorem [36, Theorem 5.6.5]) *Let A be a connected*[3] *and cocommutative Hopf algebra over a characteristic zero field. Then the inclusion $\mathrm{Prim}(A) \hookrightarrow A$ induces an isomorphism of Hopf algebras $U(\mathrm{Prim}(A)) \cong A$.*

Let us assume that A is a connected and cocommutative Hopf algebra over $\mathbb{Q}$, so that $U(\mathrm{Prim}(A)) \cong A$. The *primitive filtration* of A is the filtration

$$\mathbb{Q} \cdot 1 = P^0 A \hookrightarrow P^1 A \hookrightarrow P^2 A \hookrightarrow \cdots,$$

where $P^k A$ is the $\mathbb{Q}$-submodule of A additively generated by $\mathbb{Q} \cdot 1$ and all products of $\leq k$-many primitive elements. In particular $P^1 A = \mathbb{Q} \oplus \mathrm{Prim}(A)$ and any element $x \in A$ is in some $P^k A$, for some k, since

[3] A Hopf algebra A is *connected* if $A_0 := \mathbb{Q} \cdot 1$ is its unique simple subcoalgebra.

A is primitively generated. Under the isomorphism $U(\mathrm{Prim}(A)) \cong A$, the primitive filtration coincides with the *Lie filtration* of $U(\mathrm{Prim}(A))$ [35, Proposition 5.17].

For all $k \geq 0$, we have $P^k A = \mathbb{Q} \oplus P^k A_{\geq 1}$ where $A_{\geq 1} := A_+$ is the kernel of the counit. The filtration

$$0 = P^0 A_{\geq 1} \hookrightarrow P^1 A_{\geq 1} \hookrightarrow P^2 A_{\geq 1} \hookrightarrow \cdots$$

is called the *reduced primitive filtration* of $A_{\geq 1}$.

2.2.2 Primitive filtration of forests The focus returns to the Hopf algebra of forest diagrams on m strands, $\mathcal{A}^{(\mathrm{FI})}(m)_{\mathbb{Q}}$, which is indeed connected and cocommutative, and hence the Milnor–Moore theorem applies, i.e., $U(\mathrm{Prim}(\mathcal{A}^{(\mathrm{FI})}(m)_{\mathbb{Q}})) \cong \mathcal{A}^{(\mathrm{FI})}(m)_{\mathbb{Q}}$. It is also filtered by the primitive filtration.

In the case $m = 1$, Bar-Natan showed that $\mathrm{Prim}(\mathcal{A}(1)_{\mathbb{Q}})$ is (additively) generated by all connected diagrams [1] and his argument translates well for $m \geq 1$. Actually, this statement is also true over $\mathbb{Z}$, as shown by Lando in [28] via a beautiful counting argument. Unfortunately, the argument uses commutativity in $\mathcal{A}(1)$, and hence it does not translate directly to the general case $m \geq 1$. Instead, we give an argument that compares the primitive filtration to the size filtration defined below, in a way parallel to the ideas of the subsequent subsections in this section. First, let us introduce some notation.

Definition 2.2.2.1 The *size filtration* of $\mathcal{A}(m)$ is the filtration $\{F^k\mathcal{A}(m)\}_{k\geq 0}$ where $F^k\mathcal{A}(m)$ is the $\mathbb{Z}$-submodule generated by forest diagrams of size $\leq k$. In degree $n \geq 0$, it induces a filtration[4]

$$F^0\mathcal{A}_n(m) \hookrightarrow F^1\mathcal{A}_n(m) \hookrightarrow \cdots \hookrightarrow F^{n-1}\mathcal{A}_n(m) \hookrightarrow F^n\mathcal{A}_n(m) = \mathcal{A}_n(m)$$

of $\mathcal{A}_n(m)$. The *size filtration* of $\mathcal{A}^{\mathrm{FI}}(m)$ is defined in the same way, with FI added everywhere. Tensoring everything with $\mathbb{Q}$ yields the size filtration of $\mathcal{A}(m)_{\mathbb{Q}}$.

For $F \in \mathcal{D}(m)$ a forest diagram, define the *reduced comultiplication* $\overline{\Delta}(F)$ as

$$\overline{\Delta}(F) := \sum_{\varnothing \neq J \subsetneq \pi_0(F)} (F_J) \otimes (F_{\pi_0(F)-J}),$$

where F_J denotes the subdiagram of F that consists of the components contained in J. In other words, $\overline{\Delta}(F) = \Delta(F) - F \otimes 1 - 1 \otimes F$. Thus, $\overline{\Delta}$ satisfies AS, IHX, STU (and 1T) and descends to a well-defined map

$$\overline{\Delta} : \mathcal{A}_{\geq 1}(m) \to \mathcal{A}_{\geq 1}(m) \otimes \mathcal{A}_{\geq 1}(m).$$

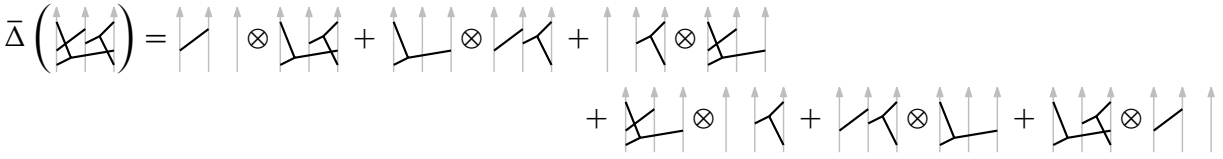
Define recursively $\overline{\Delta}^k := (\overline{\Delta}^{k-1} \otimes \mathrm{id}) \circ \overline{\Delta} : \mathcal{A}_{\geq 1}(m) \to (\mathcal{A}_{\geq 1}(m))^{\otimes k+1}$, for $k > 1$. Concretely, we have

$$\overline{\Delta}^k(F) = \sum_{\substack{J_0 \sqcup \cdots \sqcup J_k = \pi_0(F) \\ J_i \neq \varnothing}} (F_{J_0}) \otimes \cdots \otimes (F_{J_k}), \tag{2-3}$$

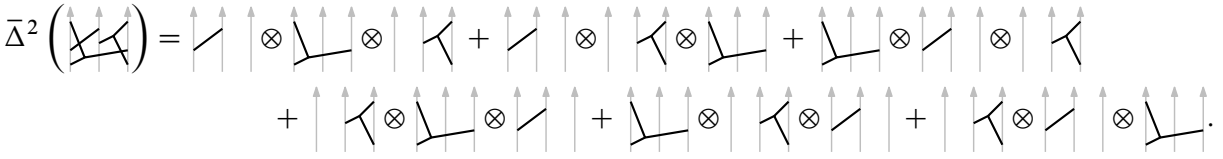
where the sum is over all possible partitions of the components of F into $k+1$ nonempty subforests.

[4]The only forest diagram of size 0 is the empty diagram 1, and hence $F^0\mathcal{A}(m) = \mathbb{Z} \cdot 1$.

For example, we have



and



Note also that the reduced primitive filtration of $\mathcal{A}_{\geq 1}(m)_{\mathbb{Q}}$ can already be defined over $\mathbb{Z}$: for $k \geq 0$, $P^k\mathcal{A}_{\geq 1}(m)$ is generated by all products of $\leq k$-many primitive elements.

Lemma 2.2.2.2 *We have the following properties*:

(i) *for all* $x, y \in \mathcal{A}_{\geq 1}(m)$,

(2-4) $$\overline{\Delta}(x \cdot y) = \overline{\Delta}(x) \cdot \overline{\Delta}(y) + \overline{\Delta}(x) \cdot (1 \otimes y + y \otimes 1) + (1 \otimes x + x \otimes 1) \cdot \overline{\Delta}(y) + x \otimes y + y \otimes x;$$

(ii) $\mathrm{Prim}(\mathcal{A}(m)) = \ker(\overline{\Delta})$;

(iii) $\overline{\Delta}(P^k\mathcal{A}_{\geq 1}(m)) \subset \sum_{i=1}^{k-1} P^i\mathcal{A}_{\geq 1}(m) \otimes P^{k-i}\mathcal{A}_{\geq 1}(m)$;

(iv) $P^k\mathcal{A}_{\geq_1}(m) \subset \ker(\overline{\Delta}^k)$;

(v) $F^k\mathcal{A}_{\geq_1}(m) \subset \ker(\overline{\Delta}^k)$, *i.e.*, $\overline{\Delta}^k$ *vanishes on forests of size* $\leq k$.

All of the above properties remain true over $\mathbb{Q}$, *and with* FI.

Proof Part (i) follows from the definition of $\overline{\Delta}$ and compatibility $\Delta(x \cdot y) = \Delta(x) \cdot \Delta(y)$. Part (ii) is immediate. Part (v) follows from the description (2-3), since for a forest F of size k the sum is empty. It remains to show (iii) and (iv).

We prove (iii) by induction on k. To lighten notation, we write P^i for $P^i\mathcal{A}_{\geq 1}(m)$. The base case, $k = 1$, is settled by (ii). Let $1 \leq k < n$ and assume that $\overline{\Delta}(P^k) \subset \sum_{i=1}^{k-1} P^i \otimes P^{k-i}$. The filtration stage P^{k+1} is generated by P^k together with all products of exactly $k+1$ primitive elements. Consider an element $z \in P^{k+1}$ of the latter sort, i.e., $z = xy$ for $x \in P^k$ and $y \in P^1 = \mathrm{Prim}(\mathcal{A}(m))$. Using $\overline{\Delta}(y) = 0$ (by (ii)), (2-4) becomes

$$\overline{\Delta}(z) = \overline{\Delta}(x) \cdot (1 \otimes y + y \otimes 1) + x \otimes y + y \otimes x.$$

Now, by induction hypothesis, we have $\overline{\Delta}(x) \in \sum_{i=1}^{k-1} P^i \otimes P^{k-i}$. Using $P^i \cdot y \subset P^{i+1}$, we finally obtain

$$\overline{\Delta}(z) \in \sum_{i=1}^{k-1} P^i \otimes P^{k+1-i} + \sum_{i=1}^{k-1} P^{i+1} \otimes P^{k-i} + P^k \otimes P^1 + P^1 \otimes P^k = \sum_{i=1}^{k} P^i \otimes P^{k+1-i},$$

which proves (iii).

The proof of (iv) also proceeds by induction on k, with the base case being (ii) again. Assume that $P^{k-1} \subset \ker(\overline{\Delta}^{k-1})$. Let $z \in P^k$. By (iii) we can write $\overline{\Delta}(z) = \sum_{i=1}^{k-1} \sum_j x_{ij} \otimes y_{ij}$ where $x_{ij} \in P^i$ and $y_{ij} \in P^{k-i}$. Then

$$\overline{\Delta}^k(z) := (\overline{\Delta}^{k-1} \otimes \mathrm{id}) \circ \overline{\Delta}(z) = \sum_{i=1}^{k-1} \sum_j \overline{\Delta}^{k-1}(x_{ij}) \otimes y_{ij} = 0$$

since all $\overline{\Delta}^{k-1}(x_{ij}) = 0$ by induction hypothesis. This concludes the proof of (iv). □

Theorem 2.2.2.3 *Over $\mathbb{Q}$, the reduced primitive filtration and the size filtration of $\mathcal{A}^{(\mathrm{FI})}(m)_{\mathbb{Q}}$ coincide. In particular,* $\mathrm{Prim}(\mathcal{A}^{(\mathrm{FI})}(m)_{\mathbb{Q}})$ *is the submodule of $\mathcal{A}^{(\mathrm{FI})}(m)_{\mathbb{Q}}$ generated by tree diagrams.*

When $m = 1$, this already holds over $\mathbb{Z}$.

Proof We omit FI in the notation since the proof for that case is the same. We write P^i for $P^i\mathcal{A}_{\geq 1}(m)$ and F^i for $F^i\mathcal{A}_{\geq 1}(m)$ to lighten notation.

In degree $n = 0$, there is nothing to prove. Fix a degree $n > 0$. First, any forest diagram of size 1 is primitive by definition of the comultiplication, i.e., $F^1\mathcal{A}_n(m)_{\mathbb{Q}} \subset P^1\mathcal{A}_n(m)_{\mathbb{Q}}$.

In fact, $F^1\mathcal{A}_n(m)_{\mathbb{Q}} \subset P^1\mathcal{A}_n(m)_{\mathbb{Q}}$ implies that $F^k\mathcal{A}_n(m)_{\mathbb{Q}} \subset P^k\mathcal{A}_n(m)_{\mathbb{Q}}$ for each $k \geq 1$. We show this by induction.[5] The base case is the above paragraph. Assume that $F^k \subset P^k$ and consider a forest diagram F of size $k+1$. Pick any tree T in F and apply a sequence $F = F_0 \rightsquigarrow \cdots \rightsquigarrow F_r$ of STU-relations to slide the legs of T above the rest of the trees. Thus,

$$F_r = T \cdot (F \setminus T) \subset P^1 \cdot F^k \subset P^1 \cdot P^k \subset P^{k+1}.$$

For each i, $F_i - F_{i+1}$ is a forest of size k by STU, i.e., $F_i - F_{i+1} \in F^k \subset P^k$. Therefore,

$$F = \sum_{i=0}^{r-1} (F_i - F_{i+1}) + F_r \subset P^{k+1}.$$

Consequently, the size filtration injects into the primitive filtration:

$$\begin{array}{ccccccccc}
F^1\mathcal{A}_n(m)_{\mathbb{Q}} & \hookrightarrow & F^2\mathcal{A}_n(m)_{\mathbb{Q}} & \hookrightarrow & \cdots & \hookrightarrow & F^{n-1}\mathcal{A}_n(m)_{\mathbb{Q}} & \hookrightarrow & F^n\mathcal{A}_n(m)_{\mathbb{Q}} \\
\downarrow & & \downarrow & & & & \downarrow & & \| \\
P^1\mathcal{A}_n(m)_{\mathbb{Q}} & \hookrightarrow & P^2\mathcal{A}_n(m)_{\mathbb{Q}} & \hookrightarrow & \cdots & \hookrightarrow & P^{n-1}\mathcal{A}_n(m)_{\mathbb{Q}} & \hookrightarrow & P^n\mathcal{A}_n(m)_{\mathbb{Q}}
\end{array}$$

We show that the vertical arrows are all equalities, inductively from right to left. Let $1 \leq k < n$ and assume that $F^{k+1} = P^{k+1}$. We seek for a section to the inclusion i_n^k as in the diagram

$$\begin{array}{ccc}
F^k\mathcal{A}_n(m)_{\mathbb{Q}} & \underset{i_n^k}{\overset{s_n^{k+1}}{\rightleftarrows}} & F^{k+1}\mathcal{A}_n(m)_{\mathbb{Q}} \\
\downarrow & & \| \\
P^k\mathcal{A}_n(m)_{\mathbb{Q}} & \hookrightarrow & P^{k+1}\mathcal{A}_n(m)_{\mathbb{Q}}
\end{array} \tag{2-5}$$

[5] This argument is similar to the proof of Proposition 6.10 in [22], which deals with forest claspers instead of forest diagrams.

Consider the linear map

$$s_n^{k+1} : F^{k+1}\mathcal{A}_n(m)_{\mathbb{Q}} \to \mathcal{A}_n(m)_{\mathbb{Q}}, \quad F \mapsto F - \frac{1}{(k+1)!}\mu^k \circ \bar{\Delta}^k(F), \tag{2-6}$$

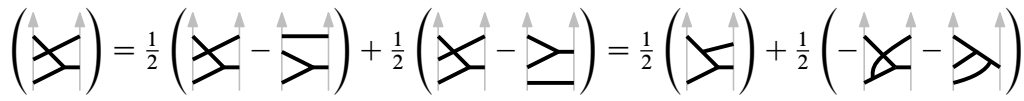
where $\mu^k : \mathcal{A}(m)^{\otimes k+1} \to \mathcal{A}(m)$ is the m-fold multiplication. We have:

• **The image of s_n^{k+1} lies in $F^k\mathcal{A}_n(m)_{\mathbb{Q}}$** Indeed, for a forest diagram F of size $k+1$, say $F = \bigcup_{i=1}^{k+1} T_i$, we have[6]

$$s_n^{k+1}(F) = \frac{1}{(k+1)!}\sum_{\sigma\in\mathfrak{S}_{k+1}} (F - (F)_\sigma), \tag{2-7}$$

where $(F)_\sigma \coloneqq T_{\sigma(1)} \dots T_{\sigma(k+1)}$ is obtained by stacking the trees of F in the order prescribed by σ. Each difference $(F - (F)_\sigma)$ can be obtained by a sequence of leg exchanges (seen later as a path in $\mathcal{F}(F)$; see Definition 2.3.3.1) from $(F)_\sigma$ to F, and hence we obtain $F - (F)_\sigma$ as a sum of forests of size k (each leg exchange yields a term where the two legs are attached together, by STU). Thus $s_n^{k+1}(F) \in F^k\mathcal{A}_n(m)_{\mathbb{Q}}$.

For example,



• **The map s_n^{k+1} is a section to the inclusion i_n^k** Indeed, for $F \in F^k\mathcal{A}_n(m)_{\mathbb{Q}}$, we have $\bar{\Delta}^k(F) = 0$ by Lemma 2.2.2.2(vi). Thus s_n^{k+1} restricts to the identity on $F^k\mathcal{A}_n(m)_{\mathbb{Q}}$.

By Lemma 2.2.2.2(v), s_n^{k+1} restricts to the identity on $P^k\mathcal{A}_n(m)_{\mathbb{Q}}$. Thus, starting with $x \in P^k\mathcal{A}_n(m)_{\mathbb{Q}}$ and going around the diagram (2-5) in the counterclockwise direction, we end up with $s_n^{k+1}(x) = x$, which shows that the vertical inclusion $F^k \hookrightarrow P^k$ is also surjective. This concludes the proof that the two filtrations coincide.

It follows that

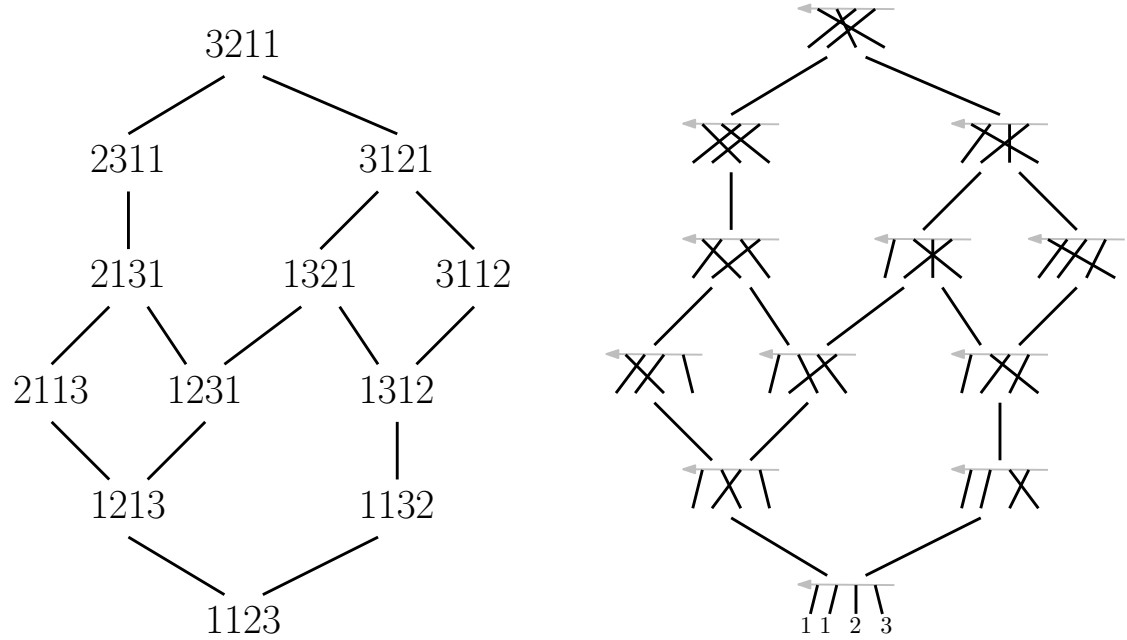
$$\mathrm{Prim}(\mathcal{A}(m)_{\mathbb{Q}}) = P^1\mathcal{A}_{\geq 1}(m)_{\mathbb{Q}} = F^1\mathcal{A}_{\geq 1}(m)_{\mathbb{Q}},$$

i.e., the primitive part of $\mathcal{A}(m)_{\mathbb{Q}}$ is additively generated by forests of size 1, that is, trees.

When $m = 1$, the bialgebra $\mathcal{A}(m)$ is commutative. Thus, if F is a forest of size $k+1$, then all $(F)_\sigma$, $\sigma \in \mathfrak{S}_{k+1}$, are equal and averaging is not needed. In particular, the section s_n^{k+1} defined in (2-6) is already defined over $\mathbb{Z}$. This yields an alternative proof of the fact that tree diagrams generate the primitive Lie algebra of $\mathcal{A}(1)$ over $\mathbb{Z}$. See [28] for a different argument. □

In order to better understand that filtration, and, in particular, the primitive Lie algebra $\mathrm{Prim}(\mathcal{A}^{(\mathrm{FI})}(m)_{\mathbb{Q}})$, a concrete description of the filtration steps, i.e., by generators and relations, is required. This is the topic of the next two sections.

[6]The choice of numbering of the trees in F is irrelevant, since the sum (2-7) is over all permutations.



(a) The permutograph $P_{(2,1,1)}$ (b) Graph of labelled legs on one strand

Figure 5: One slice of a graph of labelled forests.

2.3 Permutographs and graphs of forests

2.3.1 Permutographs For a tuple of nonnegative integers $\boldsymbol{n} := (n_1, \dots, n_s) \in \mathbb{N}^s$, consider the set $V_{\boldsymbol{n}}$ of all possible words made of n_i-many i, for each $i = 1, \dots, s$. In particular, we write

$$\underline{w}_{\boldsymbol{n}} := \underbrace{1 \dots 1}_{n_1} \dots \underbrace{s \dots s}_{n_s} \quad \text{and} \quad \overline{w}_{\boldsymbol{n}} := \underbrace{s \dots s}_{n_s} \dots \underbrace{1 \dots 1}_{n_1}.$$

The symmetric group $\mathfrak{S}_n$, where $n := \sum_i n_i$, acts on $V_{\boldsymbol{n}}$ by permutation, denoted by $\sigma \cdot w$ for $\sigma \in \mathfrak{S}_n$, $w \in V_{\boldsymbol{n}}$. The action is transitive but has many fixed points, unless all $n_i = 1$.

Definition 2.3.1.1 The ***n*-permutograph** is the undirected graph $P_{\boldsymbol{n}} := (V_{\boldsymbol{n}}, E_{\boldsymbol{n}})$, whose set of vertices is $V_{\boldsymbol{n}}$ and two distinct vertices $v, w \in V_{\boldsymbol{n}}$ are connected by an undirected edge if there is a basic transposition $\tau_i = (i, i+1) \in \mathfrak{S}_n$ such that $w = \tau \cdot v$ (hence also $v = \tau \cdot w$), i.e., $v = \dots jk \dots$ and $w = \dots kj \dots$ where $1 \le j \ne k \le s$ are the i-th and $(i+1)$-st entries of v, w.

There is at most one edge joining any pair of vertices and at most $n-1$ edges adjacent to a given vertex, since there are only $n-1$ basic transpositions $\tau_1, \dots, \tau_{n-1}$. Permutographs are connected, since the $\mathfrak{S}_{\boldsymbol{n}}$-action on $V_{\boldsymbol{n}}$ is transitive.

Remark 2.3.1.2 Permutographs are actually the 1-skeletons of permutohedra, which are rich combinatorial objects [27; 39; 44]. △

Proposition 2.3.1.3 *The first homology group of the permutograph $P_{\boldsymbol{n}}$ is generated by length 4 and 6 cycles. In other words, any 1-cycle in $P_{\boldsymbol{n}}$ is the boundary of a union of squares and hexagons.*

Proof To prove this, we scan $P_{\boldsymbol{n}}$ using the *height function*

$$h : V_{\boldsymbol{n}} \to \mathbb{N} : v \mapsto \sum_{i=1}^{n} \#\{j > i \mid v_j < v_i\},$$

which assigns to a word v the number $h(v)$ of pairs of letters $\dots v_i \dots v_j \dots$ where $v_j < v_i$ appearing in it, i.e., the number of inversions in v. For example, $h(\underline{w}_{\boldsymbol{n}}) = 0$ is minimal and $h(\overline{w}_{\boldsymbol{n}})$ is maximal in $P_{\boldsymbol{n}}$. The function h is a Morse function in the sense that, for any edge $\{v, w\}$, we have $h(w) = h(v) \pm 1$. Pictorially, when drawing $P_{\boldsymbol{n}}$ with all vertices of height i at height i, no edge is horizontal (see Figure 5(a)).

We show that any 1-cycle C in $P_{\boldsymbol{n}}$ is a sum of length 4 and 6 cycles, by induction on $\max_{v \in C} h(v)$. Up to decomposing C into smaller subcycles and removing paths going back and forth, we can assume that $C = \overrightarrow{v^0 v^1} + \cdots + \overrightarrow{v^{l-1} v^l}$ where $(v^0, \dots, v^l = v^0)$ is a sequence of adjacent vertices without repetitions (except $v^l = v^0$). If $\max_{v \in C} h(v) = 0$ then there is nothing to prove since $C = 0$. This proves the base case.

Consider a 1-cycle C with $M := \max_{v \in C} h(v) > 0$ and assume the statement to be true for all cycles of maximal height $< M$. Pick any vertex $v^M \in C$ that reaches the maximum M, then we must have $h(v^{M \pm 1}) = h(v^M) - 1$. By definition of the permutograph, $v^{M-1} = \tau_i \cdot v^M$ and $v^{M+1} = \tau_j \cdot v^M$ for some i, j such that $i \neq j$. We assume $i < j$ to simplify notation. There are two cases to consider:

(1) If $|j - i| > 1$, then τ_i and τ_j have disjoint support and the vertex $\widetilde{v}^M := \tau_i \tau_j \cdot v^M = \tau_i \cdot v^{M+1} = \tau_j \cdot v^{M-1}$ forms a square together with v^M, v^{M+1}, v^{M-1} (we use the relation $\tau_i \tau_j = \tau_j \tau_i$ in $\mathfrak{S}_n$). In particular,

$$\begin{aligned} C &= \overrightarrow{v^0 v^1} + \cdots + \overrightarrow{v^{M-1} v^M} + \overrightarrow{v^M v^{M+1}} + \cdots + \overrightarrow{v^{l-1} v^l} \\ &= \underbrace{(\overrightarrow{v^0 v^1} + \cdots + \overrightarrow{v^{M-1} \widetilde{v}^M} + \overrightarrow{\widetilde{v}^M v^{M+1}} + \cdots + \overrightarrow{v^{l-1} v^l})}_{=: C'} + c = C' + c, \end{aligned}$$

where $c := \overrightarrow{v^{M-1} v^M} + \overrightarrow{v^M v^{M+1}} - \overrightarrow{v^{M-1} \widetilde{v}^M} - \overrightarrow{\widetilde{v}^M v^{M+1}}$ is a length 4 cycle. Moreover, $\widetilde{v}^M$ has height $M - 2$ since it has exactly two inversions less than v^M.

(2) If $|j - i| = 1$, and hence $j = i + 1$, then the above does not work since the vertices $\tau_i \tau_{i+1} \cdot v^M$ and $\tau_{i+1} \tau_i \cdot v^M$ are distinct. In this case we need to dive one step lower in the permutograph and use instead the relation $\tau_i \tau_{i+1} \tau_i = \tau_{i+1} \tau_i \tau_{i+1}$ in $\mathfrak{S}_n$.

Denote the i-th, $(i+1)$-st and $(i+2)$-nd entries of v^M by c, b, a, respectively. These are the entries that are permuted by τ_i and τ_{i+1}. Given a permutation (x_1, x_2, x_3) of (c, b, a), denote by $v^M_{x_1 x_2 x_3}$ the word obtained from v^M by replacing the i-th, $(i+1)$-st and $(i+2)$-nd entries by x_1, x_2, x_3, respectively. In particular, $v^{M-1} = v^M_{bca}$, $v^M = v^M_{cba}$ and $v^{M+1} = v^M_{cab}$.

Define C' from C by replacing the vertex v^M with the vertices v^M_{bac}, v^M_{abc}, v^M_{acb}, in this order. Those new vertices have height $< M$.

Then $C = C' + c$ where c is a length 6 cycle going through the vertices v^M_{bca}, v^M_{cba}, v^M_{cab}, v^M_{acb}, v^M_{abc}, v^M_{bac}, v^M_{bca}, in this order.

See the hexagon at the bottom of Figure 5(a) for an example with $(c, b, a) = (3, 2, 1)$.

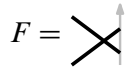
In both cases, $C = C' + c$ for a length 4 or 6 cycle c, and C' is a 1-cycle with one less vertex of height M. Repeating this process for each vertex of height M in C, we obtain $C = C'' + c_1 + \cdots + c_k$ where $c_1, \dots, c_k$ are length 4 or 6 cycles, and where C'' has no more vertex of height M. That is, C'' has maximal height $< M$, hence C'' is a sum of length 4 and 6 cycles by the induction hypothesis, so is C. □

Definition 2.3.1.4 For two graphs G_1, G_2, their *cartesian* or *box product*, denoted by $G_1 \,\square\, G_2$, is the 1-skeleton of $G_1 \times G_2$. The box product of graphs is associative.

Corollary 2.3.1.5 *The first homology of a box product of permutographs $\square_{j=1}^m P_{\boldsymbol{n}^j}$ is generated by length* 4 *and* 6 *cycles.*

Proof The exact same proof as Proposition 2.3.1.3 works, where now the two transpositions in case 1 can happen in different factors. Consequently, there are two types of length 4 cycles: product of edges from different factors and squares contained in one factor. □

2.3.2 Graph of labelled forests A *labelled forest diagram* is simply a forest diagram F in $\mathcal{D}(m)$ together with a labelling of its trees by $1, \dots, s$, where s is the size of the forest, i.e., a choice of identification $\pi_0(F) \cong \{1, \dots, s\}$. We say that two labelled forests F, F' are related by a *slide move* if F' is obtained from F by permuting adjacent legs belonging to distinct trees, that is,

$F =$  slide move  ,

where it is understood, as usual, that F and F' are identical outside of the part shown.

Definition 2.3.2.1 Given s-many tree diagrams $T_1, \dots, T_s \in \mathcal{D}^T(m)$ on m strands ($s > 0$), the *graph of labelled forests on $T_1, \dots, T_s$* is the undirected graph whose set of vertices is the set

$$\widetilde{\mathcal{F}}(T_1, \dots, T_s)$$

of all labelled forests on m strands, whose i-th tree is T_i, and where two labelled forests are connected by an edge if they are related by a slide move.

A part of a graph of labelled forests, corresponding to what happens on one strand, is shown on Figure 5(b). On this figure, the two leftmost legs (e.g., in the bottom diagram) belong to T_1, while the next ones belong to T_2 and T_3, respectively.[7] The legs from T_1 are not allowed to slide across one another, but they can slide across the two other legs. Note that any two vertices in $\widetilde{\mathcal{F}}(T_1, \dots, T_s)$ are connected by *at most* one edge. There are no edges when $s = 1$.

[7] For simplicity, we have labelled only the forest at the bottom, but all forests on the figure should be labelled accordingly.

Proposition 2.3.2.2 *Suppose that the tree T_i has n_i^j legs on the j-th strand, for $1 \le i \le s$ and $1 \le j \le m$. Then there is a graph isomorphism defined on vertices as*[8]

$$W : \widetilde{\mathcal{F}}(T_1, \dots, T_s) \xrightarrow{\cong} \square_{j=1}^{m} P_{\boldsymbol{n}^j}, \tag{2-8}$$

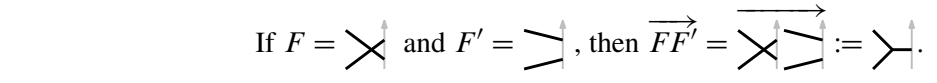
where $\boldsymbol{n}^j = (n_1^j, \dots, n_s^j)$, sending a labelled forest to $(w^j)_j$ where w^j encodes the ordering of the legs on the j-strand, from top to bottom.

Proof This is immediate, since a labelled forest is entirely determined by its trees $T_1, \dots, T_s$ and the relative positions of their legs on the strands. □

Corollary 2.3.2.3 *The first homology of the graph of labelled forests $\widetilde{\mathcal{F}}(T_1, \dots, T_s)$ is generated by length 4 and 6 cycles. In other words, any 1-cycle in it is the boundary of a union of squares and hexagons.*

Proof This follows immediately from Proposition 2.3.2.2 and Corollary 2.3.1.5. □

Definition 2.3.2.4 To a directed edge $\overrightarrow{FF'}$ in $\widetilde{\mathcal{F}}(T_1, \dots, T_s)$, $s \ge 2$, one can assign the third term appearing in the STU relation, which we denote by $\overrightarrow{FF'} \in \mathcal{D}^{s-1}(m)$:



In particular, $\overrightarrow{FF'} = F' - F$ in $\mathcal{A}(m)$.

The fact that the legs involved in a slide move belong to distinct trees is important here, otherwise $\overrightarrow{FF'}$ would have a cycle and would not be a forest diagram.

There are many paths joining forests $F, F' \in \widetilde{\mathcal{F}}(T_1, \dots, T_s)$. Hence, $\overrightarrow{FF'}$ is not well defined for nonadjacent forests. However, any two paths differ by a cycle in $\widetilde{\mathcal{F}}(T_1, \dots, T_s)$, which by Corollary 2.3.2.3 can be written as a sum of back-and-forth paths, length 4 and 6 cycles. Thus, modding out the target $\mathbb{Z}\mathcal{D}^{s-1}(m)$ by those cycles will give us a well-defined map $\overrightarrow{\cdot}$.

Definition 2.3.2.5 A $\square$ *relation* in $\mathbb{Z}\mathcal{D}^{s-1}(m)$ is a relation of the form

$$\overrightarrow{F_1F_2} + \overrightarrow{F_2F_3} + \overrightarrow{F_3F_4} + \overrightarrow{F_4F_1}, \tag{2-9}$$

where (F_1, F_2, F_3, F_4) is a length 4 cycle in $\widetilde{\mathcal{F}}(T_1, \dots, T_s)$ for some trees T_i. See Figure 6(a).

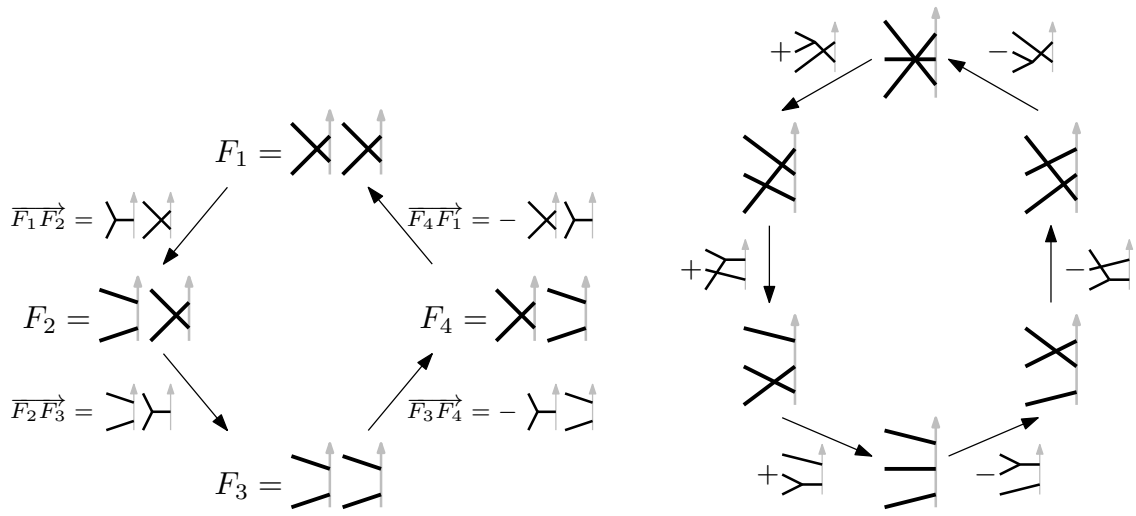
Similarly, a $\hexagon$ *relation* in $\mathbb{Z}\mathcal{D}^{s-1}(m)$ is a relation of the form

$$\overrightarrow{F_1F_2} + \overrightarrow{F_2F_3} + \overrightarrow{F_3F_4} + \overrightarrow{F_4F_5} + \overrightarrow{F_5F_6} + \overrightarrow{F_6F_1}, \tag{2-10}$$

where $(F_1, \dots, F_6)$ is a length 6 cycle in $\widetilde{\mathcal{F}}(T_1, \dots, T_s)$ for some trees T_i. See Figure 6(b).

Let $\square R$, resp. $\hexagon R$, denote the submodule of $\mathbb{Z}\mathcal{D}(m)$ generated by all $\square$, resp. $\hexagon$, relations.

Remark 2.3.2.6 A $\hexagon$ relation can be obtained in the following way: start with an IHX relation that involves a leg, and break apart that leg in the three terms. The resulting signed sum of six terms is a $\hexagon$ relation, and all $\hexagon$ relations can be obtained in this way. △



(a) A □ relation (b) A ⬡ relation

Figure 6: Relations obtained from cycles of length 4 and 6.

By the above definition and Corollary 2.3.2.3, the following definition is well defined (AS takes care of the back-and-forth paths).

Definition 2.3.2.7 For two labelled forests $F, F' \in \widetilde{\mathcal{F}}(T_1, \dots, T_s)$, $s \geq 2$, define

$$\overrightarrow{FF'} := \overrightarrow{FF_1} + \cdots + \overrightarrow{F_{l-1}F'} \in \mathbb{Z}\mathcal{D}^{s-1}(m)/\langle \mathrm{AS}, \mathrm{STU}^2, \hexagon R\rangle, \tag{2-11}$$

where $F \rightsquigarrow F_1 \rightsquigarrow \cdots \rightsquigarrow F_{l-1} \rightsquigarrow F'$ is a path from F to F' in $\widetilde{\mathcal{F}}(T_1, \dots, T_s)$.

Proposition 2.3.2.8 *Square relations and* STU^2 *relations coincide, i.e.,* $\square R = \mathrm{STU}^2$ *as submodules of* $\mathbb{Z}\mathcal{D}^s(m)$.

Proof Recall [9, Definition 3.1] that STU^2 relations among size s forests are defined starting with a *template*, i.e., a size $s-1$ forest diagram or a size s Feynman diagram with a single cycle, then choosing two legs (which must be adjacent to the cycle in the latter case) and breaking one or the other apart using STU. Breaking apart both legs yields four possible size $s+1$ forest diagrams, which form the corners of a square in a graph of labelled forests and the corresponding □ relation coincides with the STU^2 relation.

Conversely, consider a square as on Figure 6(a) and the associated □ relation. Then the template  yields a STU^2 relation which coincides with that □ relation. □

2.3.3 Graph of forests In the previous section we investigated the graph of labelled forests and the relations that are required to define differences between labelled forests that are joined by a path. However, forests diagrams in the bialgebra $\mathcal{A}(m)$ are not labelled. The *true* graph of forests is a quotient of the graph of labelled forests, where the quotient map forgets the labelling.

[8] Since there is at most one edge connecting two vertices, this determines the morphism on edges as well.

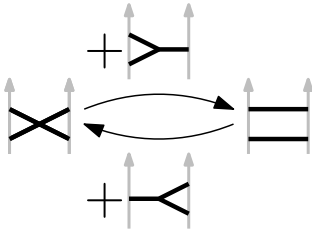


Figure 7: Two different leg moves joining two forests.

Definition 2.3.3.1 The *graph of forests* on trees $T_1, \dots, T_s$ is the graph $\mathcal{F}(T_1, \dots, T_s)$ whose vertices are all forest diagrams in $\mathcal{D}^s(m)$ consisting of those trees, and to each slide move relating two forests there is an associated edge joining the corresponding vertices.

Note that $\mathcal{F}(T_1, \dots, T_s)$ does not depend on the ordering of the trees T_i. If $F = \bigcup_{i=1}^s T_i$ is a forest diagram consisting of trees $T_1, \dots, T_s$, we write $\mathcal{F}(F) \coloneqq \mathcal{F}(T_1, \dots, T_s)$.

There is a natural map

$$\widetilde{\mathcal{F}}(T_1, \dots, T_s) \twoheadrightarrow \mathcal{F}(T_1, \dots, T_s) \tag{2-12}$$

that forgets the labelling of forests. It is surjective but not injective if there are identical trees among $T_1, \dots, T_s$.

Remark 2.3.3.2 In $\mathcal{F}(T_1, \dots, T_s)$, it can happen that two vertices are joined by multiple edges. See Figure 7 for an example: the top, resp. bottom, edge corresponds to exchanging the legs on the right, resp. left, strand.

Thus, the notation $\overrightarrow{FF'}$ is ambiguous. Given a directed edge $F \overset{e}{\rightsquigarrow} F'$ we write $\overrightarrow{e}$ to denote the third term of the corresponding STU relation, i.e., $\overrightarrow{e} = F' - F$ in $\mathcal{A}(m)$. Similarly, if P is a path consisting of directed edges $e_1, \dots, e_l$, then define $\overrightarrow{P} \coloneqq \sum_{i=1}^l \overrightarrow{e_i}$. △

Lemma 2.3.3.3 *For any labelled forest diagram F, the forgetful map $\widetilde{\mathcal{F}}(F) \twoheadrightarrow \mathcal{F}(F)$ satisfies the following* ***path-lifting property***: *For any path $P = F_0 \overset{e_0}{\rightsquigarrow} F_1 \overset{e_1}{\rightsquigarrow} \cdots \overset{e_{l-1}}{\rightsquigarrow} F_l$ in $\mathcal{F}(F)$ and any choice of labelled forest $\widetilde{F}_0$ lifting F_0, there exists a unique path $\widetilde{P} = \widetilde{F}_0 \overset{\widetilde{e}_0}{\rightsquigarrow} \widetilde{F}_1 \overset{\widetilde{e}_1}{\rightsquigarrow} \cdots \overset{\widetilde{e}_{l-1}}{\rightsquigarrow} \widetilde{F}_l$ in $\widetilde{\mathcal{F}}(F)$ that lifts P. In other words, the surjection $\widetilde{\mathcal{F}}(F) \twoheadrightarrow \mathcal{F}(F)$ is a finite covering.*

Moreover, $\overrightarrow{\widetilde{e}_i} = \overrightarrow{e_i}$ for each $i = 0, \dots, l-1$.

Proof Starting with $\widetilde{F}_0$ and applying the sequence of slide moves prescribed by P yields such a lift $\widetilde{P}$, which is uniquely determined by this process. □

Proposition 2.3.3.4 *Any cycle in $\mathcal{F}(T_1, \dots, T_s)$ has a multiple that lifts along the covering*

$$\widetilde{\mathcal{F}}(T_1, \dots, T_s) \twoheadrightarrow \mathcal{F}(T_1, \dots, T_s).$$

In particular, over $\mathbb{Q}$, for any C, the element $\overrightarrow{C}$ belongs to the submodule $\langle \mathrm{AS}, \square R, \circlearrowright R \rangle \subset \mathbb{Q}\mathcal{D}^{s-1}(m)$.

Proof Let G be the group of deck transformations of the finite covering $p : \widetilde{\mathcal{F}}(T_1,\dots,T_s) \twoheadrightarrow \mathcal{F}(T_1,\dots,T_s)$, and $|G|$ its size. Consider the transfer homomorphism $\tau : H_1(\mathcal{F}(T_1,\dots,T_s)) \to H_1(\widetilde{\mathcal{F}}(T_1,\dots,T_s))$. Then $p_* \circ \tau$ equals multiplication by $|G|$. In particular, the cokernel of p_* is $|G|$-torsion and p_* becomes surjective when passing to coefficients in $\mathbb{Q}$. This concludes.

The second statement then follows from the first and Corollary 2.3.2.3. □

With the above result in hands, we can adapt Definition 2.3.2.7 to the graph of forests, but with rational coefficients. See Remark 2.3.3.2 for the notation $\overrightarrow{e}$ and $\overrightarrow{P}$.

Definition 2.3.3.5 For two forests $F, F' \in \mathcal{F}(T_1,\dots,T_s)$, define

$$\overrightarrow{FF'} := \overrightarrow{P} \in \mathbb{Q}\mathcal{D}^{s-1}(m)/\langle \mathrm{AS}, \mathrm{STU}^2, \hexagon R\rangle, \tag{2-13}$$

where P is a path from F to F' in $\mathcal{F}(T_1,\dots,T_s)$.

2.4 Presentation of the primitive filtration of $\mathcal{A}(m)_{\mathbb{Q}}$

2.4.1 Splitting trees

Definition 2.4.1.1 For $1 \le s \le n-1$, the $\mathbb{Q}$*-module of size s forest diagrams* $\mathcal{F}^s(m)_{\mathbb{Q}}$ is defined by

$$\mathcal{F}^s(m)_{\mathbb{Q}} := \mathbb{Q}\mathcal{D}^s(m)/\langle \mathrm{AS}, \mathrm{IHX}, \mathrm{STU}^2, \hexagon R\rangle.$$

It is graded by the degree of diagrams, with degree-n graded part denoted by $\mathcal{F}^s_n(m)_{\mathbb{Q}}$. The *framing independent version* is defined as $\mathcal{F}^{\mathrm{FI},s}(m)_{\mathbb{Q}} := \mathcal{F}^s(m)_{\mathbb{Q}}/\langle \mathrm{1T}\rangle$. We use the notation (FI) as explained in Section 2.1.3.

Remark 2.4.1.2 When $s = 1$, the submodule $\hexagon R$ is actually trivial, since a $\hexagon$ relation requires at least two trees. Thus, $\mathcal{F}^1(m)_{\mathbb{Q}} = \mathcal{L}(m)_{\mathbb{Q}}$ as $\mathbb{Q}$-modules (see Theorem 2.4.5.1 for the definition of $\mathcal{L}(m)_{\mathbb{Q}}$).

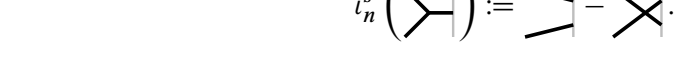
In degree n and with $s = n$, STU^2 becomes 4T while $\hexagon R$ is trivial; hence we have $\mathcal{F}^n_n(m)_{\mathbb{Q}} \cong \mathcal{A}_n(m)_{\mathbb{Q}}$.

If $s < n-1$, the $\hexagon$ relations are actually implied by STU^2 and IHX. Indeed, by Remark 2.3.2.6 any $\hexagon$ relation can be obtained by breaking apart a leg in an IHX relation. Since $s < n-1$, the terms of the $\hexagon$ relation must contain a node not involved in the relation. Apply STU^2 to break that additional node and rebuild a sum of two IHX relations. Thus $\hexagon R \subset \langle \mathrm{STU}^2, \mathrm{IHX}\rangle$ as submodules of $\mathbb{Q}\mathcal{D}^s_n(m)$ when $s < n-1$. △

For each $1 \le s < n$, there is a linear map

$$\tilde{\iota}^s_n : \mathbb{Q}\mathcal{D}^s_n(m) \to \mathcal{F}^{s+1}_n(m)_{\mathbb{Q}} \tag{2-14}$$

defined as follows: given a size s forest diagram $F \in \mathcal{D}^s_n(m)$, pick a trivalent vertex (which must exist since $s < n$) adjacent to a strand and break it apart according to STU, i.e.,

$\tilde{\iota}^s_n$  (2-15)

By STU^2, this does not depend on the choice of trivalent vertex to break apart.

Proposition 2.4.1.3 *The linear map $\tilde{\iota}_n^s := \mathbb{Q}\mathcal{D}_n^s(m) \to \mathcal{F}_n^{(\mathrm{FI}),s+1}(m)_{\mathbb{Q}}$ factors through the linear map $\iota_n^s : \mathcal{F}_n^{(\mathrm{FI}),s}(m)_{\mathbb{Q}} \to \mathcal{F}_n^{(\mathrm{FI}),s+1}(m)_{\mathbb{Q}}$.*

Proof This proof is very similar to the proof of Claim 3.4 in [9].

We need to show that $\tilde{\iota}_n^s$ vanishes on AS, IHX, STU^2, $\hexagon R$. If $s < n-2$, then forest diagrams in $\mathcal{D}_n^s(m)$ have at least three nodes, while the relations AS, IHX, STU^2, $\hexagon R$ involve at most 2 nodes. Hence it suffices to pick another node to break apart and the image under $\tilde{\iota}_n^s$ is a difference of sums that belong to the respective submodules.

For $s < n-1$, the above argument also applies to STU^2, $\hexagon R$, AS. For $s = n-2$ and IHX, observe that an IHX relation is sent to a $\hexagon$ relation.

For $s = n-1$, there are no IHX relations, while a direct calculation shows that STU^2 and AS relations are sent to zero.

This concludes the proof that $\tilde{\iota}_n^s$ factors through $\mathcal{F}_n^s(m)_{\mathbb{Q}}$. The FI case holds as well: if D has an isolated chord, then so do the two terms in $\tilde{\iota}_n^s(D)$. □

Remark 2.4.1.4 In [9, Claim 3.4], the proof restricts to the case $s \neq n-1$ because the $\hexagon$ relation is not taken into account there. △

The linear maps ι_n^s, $s = 1, \dots, n-1$, provide a factorisation of the map $\iota_n : \mathcal{F}_n^1(m)_{\mathbb{Q}} \to \mathcal{A}_n(m)_{\mathbb{Q}}$,

$$\mathcal{F}_n^1(m)_{\mathbb{Q}} \xrightarrow{\iota_n^1} \mathcal{F}_n^2(m)_{\mathbb{Q}} \xrightarrow{\iota_n^2} \cdots \xrightarrow{\iota_s^{n-1}} \mathcal{F}_n^n(m)_{\mathbb{Q}} = \mathcal{A}_n(m)_{\mathbb{Q}},$$
$$\mathcal{F}_n^1(m)_{\mathbb{Q}} \xrightarrow{\qquad\qquad \iota_n \qquad\qquad} \mathcal{F}_n^n(m)_{\mathbb{Q}} = \mathcal{A}_n(m)_{\mathbb{Q}},$$

where nodes are broken apart one after the other, instead of all at once

2.4.2 Barycenters of forests Let $T_1, \dots, T_s \in \mathcal{D}^T(m)$ be tree diagrams. Definition 2.3.3.5 defines a map

$$\overrightarrow{\cdot} : \mathcal{F}(T_1, \dots, T_s) \times \mathcal{F}(T_1, \dots, T_s) \to \mathcal{F}^{(\mathrm{FI}),s-1}(m)_{\mathbb{Q}} : (F, F') \mapsto \overrightarrow{FF'}. \tag{2-16}$$

The vector notation is justified by the fact that $\mathcal{F}(T_1, \dots, T_s)$ inherits some properties of affine spaces. In particular, we can still talk about barycenters of vertices of $\mathcal{F}(T_1, \dots, T_s)$.

Definition 2.4.2.1 The *barycenter* of forests $F_1, \dots, F_k \in \mathcal{F}(T_1, \dots, T_s)$ with coefficients $\lambda_1, \dots, \lambda_k \in \mathbb{Q}$ summing to $\sum_i \lambda_i = 1$ is the formal sum $\sum_i \lambda_i F_i$.

Denote by $\mathrm{Bar}(\mathcal{F}(T_1, \dots, T_s))$ the set of all barycenters of forests in $\mathcal{F}(T_1, \dots, T_s)$. There is a natural inclusion $\mathcal{F}(T_1, \dots, T_s) \hookrightarrow \mathrm{Bar}(\mathcal{F}(T_1, \dots, T_s))$ sending a forest F to $1 \cdot F$.

Lemma 2.4.2.2 *We have the following properties*:

(i) *If $F_0, F_1, F_2 \in \mathcal{F}(T_1, \dots, T_s)$, then $\overrightarrow{F_0F_2} = \overrightarrow{F_0F_1} + \overrightarrow{F_1F_2}$ (Chasles relation).*

(ii) *Let $F_0, F_0', F_1, \dots, F_k \in \mathcal{F}(T_1, \dots, T_s)$ and $\lambda_1, \dots, \lambda_k \in \mathbb{Q}$ such that $\sum_i \lambda_i = 0$. Then $\sum_i \lambda_i \overrightarrow{F_0F_i} = \sum_i \lambda_i \overrightarrow{F_0'F_i}$ in $\mathcal{F}^{(\mathrm{FI}),s-1}(m)_{\mathbb{Q}}$.*

(iii) *The map*

$$\overrightarrow{\cdot} : \mathrm{Bar}(\mathcal{F}(T_1,\dots,T_s)) \times \mathrm{Bar}(\mathcal{F}(T_1,\dots,T_s)) \to \mathcal{F}^{(\mathrm{FI}),s-1}(m)_{\mathbb{Q}},$$
$$\left(B = \sum_i \lambda_i F_i,\, B' = \sum_j \lambda'_j F'_j\right) \mapsto \sum_j \lambda'_j \overrightarrow{F_0 F'_j} - \sum_i \lambda_i \overrightarrow{F_0 F_i}, \tag{2-17}$$

where $F_0 \in \mathcal{F}(T_1,\dots,T_s)$ is any forest, does not depend on the choice of F_0, extends the map (2-16) *and satisfies the Chasles relation.*

Proof Part (i) follows by choosing a path from F_0 to F_2 passing through F_1. Part (ii) follows directly from (i) and implies that (2-17) does not depend on the choice of F_0. The rest is direct. □

By Lemma 2.4.2.2, any choice of barycenter $B_0 \in \mathrm{Bar}(\mathcal{F}(T_1,\dots,T_s))$ defines a map

$$\mathcal{F}(T_1,\dots,T_s) \to \mathcal{F}^{(\mathrm{FI}),s-1}(m)_{\mathbb{Q}} : F \mapsto \overrightarrow{B_0 F},$$

which sends a forest of size s to a linear combination of forests of size $s-1$. In order to obtain sections to the maps ι^s_n, we need to consistently pick barycenters in each graph of forest.

Definition 2.4.2.3 For F a forest of size s consisting of trees $T_1,\dots,T_s$, define the *average barycenter*, or *average basepoint*, associated to F as

$$(F)_{\mathrm{avg}} := \frac{1}{s!} \sum_{\sigma \in \mathfrak{S}_s} T_{\sigma(1)} \dots T_{\sigma(s)}, \tag{2-18}$$

i.e., the average of all possible ways of stacking the trees of F one above the other.

Note that $(F)_{\mathrm{avg}} = (F')_{\mathrm{avg}}$ whenever F and F' consist of the same trees, i.e., they belong to a common graph of forests.

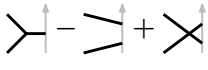
2.4.3 Merging trees Consider the linear map

$$\widetilde{\pi}^s_n : \mathbb{Q}\mathcal{D}^s_n(m) \to \mathcal{F}^{(\mathrm{FI}),s-1}_n(m)_{\mathbb{Q}} : F \mapsto \overrightarrow{(F)_{\mathrm{avg}} F}, \tag{2-19}$$

which is well defined by Lemma 2.4.2.2.

Lemma 2.4.3.1 (diagrammatic STU) *For any adjacent forests $F^=$ =  and $F^\times$ =  in a graph of forests, we have*

$$\widetilde{\pi}^s_n\left(\text{} - \text{}\right) = \text{},$$

where  *is an* STU *relation.*

Proof Since $F^=$ and $F^\times$ belong to the same graph of forests, we have $(F^=)_{\mathrm{avg}} = (F^\times)_{\mathrm{avg}}$ and

$$\widetilde{\pi}^s_n(F^= - F^\times) = \widetilde{\pi}^s_n(F^=) - \widetilde{\pi}^s_n(F^\times) = \overrightarrow{(F^=)_{\mathrm{avg}} F^=} - \overrightarrow{(F^\times)_{\mathrm{avg}} F^\times} = \overrightarrow{F^\times F^=} = F^Y$$

by the Chasles relation. □

Proposition 2.4.3.2 *The map* $\widetilde{\pi}_n^s : \mathbb{Q}\mathcal{D}_n^s(m) \to \mathcal{F}_n^{(\mathrm{FI}),s-1}(m)_{\mathbb{Q}}$ *factors through* $\pi_n^s : \mathcal{F}_n^{(\mathrm{FI}),s}(m)_{\mathbb{Q}} \to \mathcal{F}_n^{(\mathrm{FI}),s-1}(m)_{\mathbb{Q}}$.

Proof We need to show that $\widetilde{\pi}_n^s$ vanishes on (1T), AS, IHX, STU^2 and $\circlearrowright R$. The argument is similar to [9, Claim 3.6].

Start with an AS relation $F + F'$ in $\mathcal{D}_n^s(m)$. We can write $(F^{(\prime)})_{\mathrm{avg}} = \sum_\sigma \lambda_\sigma F_\sigma^{(\prime)}$ where $F_\sigma + F'_\sigma$ are AS relations (the notation $(\prime)$ indicates that we can read it with or without $\prime$). Pick a path $F_\sigma \overset{e_0}{\rightsquigarrow} \cdots \overset{e_{l-1}}{\rightsquigarrow} F$ of leg slides in $\mathcal{F}(F)$. Then we can follow the parallel path

$$F'_\sigma \overset{e'_0}{\rightsquigarrow} \cdots \overset{e'_{l-1}}{\rightsquigarrow} F'$$

in $\mathcal{F}(F')$, obtained by performing the exact same leg slides. Notice that for each j, the origin and extremities of e_j and e'_j only differ by the cyclic ordering of a node, in the same trees as for F and F'. Consequently, $\overrightarrow{e_j} + \overrightarrow{e'_j}$ is an AS relation for each j and $\overrightarrow{F_\sigma F} + \overrightarrow{F'_\sigma F'}$ is a sum of those. In particular,

$$\overrightarrow{(F)_{\mathrm{avg}} F} + \overrightarrow{(F')_{\mathrm{avg}} F'} = \sum_\sigma \lambda_\sigma (\overrightarrow{F_\sigma F} + \overrightarrow{F'_\sigma F'}) = 0.$$

For an IHX relation $F^I + F^H + F^X$, the argument is the same. As above, picking parallel paths from $F_\sigma^\bullet$ to $F^\bullet$, for $\bullet \in \{I, H, X\}$, we obtain

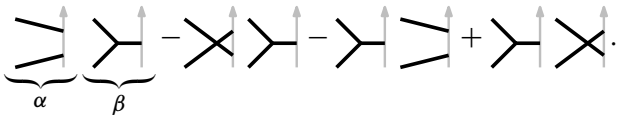
$$\overrightarrow{(F^I)_{\mathrm{avg}} F^I} + \overrightarrow{(F^H)_{\mathrm{avg}} F^H} + \overrightarrow{(F^X)_{\mathrm{avg}} F^X} = 0$$

since it is a sum of IHX relations.

Unlike AS and IHX which happen far from the strands, STU^2 happens close to the strands and we need to distinguish two cases. If the STU^2 relation $F^{Y=} - F^{Y\times} - F^{=Y} + F^{\times Y}$ involves two or three trees, then the legs forming $=$ or $\times$ must belong to distinct trees.[9] Then, two applications of the diagrammatic STU (Lemma 2.4.3.1) yield

$$\widetilde{\pi}_n^s(F^{Y=} - F^{Y\times}) = F^{YY} = \widetilde{\pi}_n^s(F^{=Y} - F^{\times Y}),$$

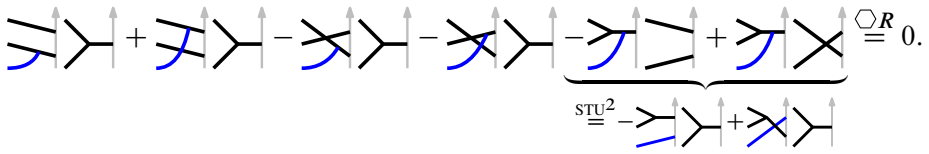
and hence the STU^2 relation is satisfied by $\widetilde{\pi}_n^s$ in this case. Let us turn to the second case, where the STU^2 relation involves a single tree, i.e., all the legs involved belong to the same tree. We argue as for the AS and IHX relations, by picking four parallel paths of slide moves from $F_\sigma^\bullet$ to $F^\bullet$, for $\bullet \in \{Y=, Y\times, =Y, \times Y\}$. More precisely, call the places involved in the STU^2 relation α and β,



A leg sliding across site α induces one or two edges in the four parallel paths, depending on whether site β contains a "Y", "=" or "×" configuration. The corresponding terms in $\sum_\bullet \overrightarrow{F_\sigma^\bullet F^\bullet}$ add up to zero

[9] Otherwise, joining them in the term with Y would form a cycle.

by STU^2 and $\hexagon R$:



When a leg slides across site β, the argument is the same. Thus $\widetilde{\pi}_n^s$ satisfies STU^2.

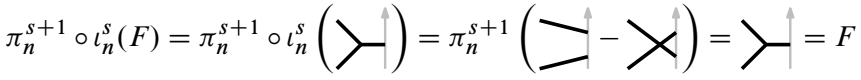
It remains to show that $\hexagon$-relations are satisfied. This immediately follows from three applications of the diagrammatic STU lemma (Lemma 2.4.3.1) and the IHX relation.

This concludes the proof that $\widetilde{\pi}_n^s$ factors through $\mathcal{F}_n^s(m)_{\mathbb{Q}}$.

In the FI case, we need to check that $\widetilde{\pi}_n^s$ vanishes on 1T. Let $F \in \mathcal{D}_n^s(m)$ be a forest diagram with an isolated chord. For $\sigma \in \mathfrak{S}_s$, the element $\overrightarrow{(F)_\sigma F}$ is a sum of diagrams in which this isolated chord remains, together with diagrams coming from sliding that isolated chord across another leg. Those terms of the latter type vanish in pairs by AS. This also follows from the commutativity property, shown in the proof of [1, Lemma 3.1]. □

Proposition 2.4.3.3 *For each $1 \le s < n$, the map π_n^{s+1} is a section of ι_n^s, i.e., $\pi_n^{s+1} \circ \iota_n^s = \mathrm{id}_{\mathcal{F}_n^{(\mathrm{FI}),s}(m)_{\mathbb{Q}}}$. In particular, ι_n^s is injective.*

Proof Given a size s and degree-n forest diagram $F =$  (represented by one of its nodes adjacent to a strand, which exists since $s < n$), we have



by the diagrammatic STU (Lemma 2.4.3.1). □

2.4.4 From $\mathcal{F}^k(m)_{\mathbb{Q}}$ to $F^k\mathcal{A}(m)_{\mathbb{Q}}$ Combining the results of Sections 2.2 and 2.4.3, we obtain the following theorem.

Theorem 2.4.4.1 *For each $1 \le k \le n$, there is a natural linear isomorphism*

$$\mathcal{F}_n^{(\mathrm{FI}),k}(m)_{\mathbb{Q}} \xrightarrow{\cong} F^k\mathcal{A}_n^{(\mathrm{FI})}(m)_{\mathbb{Q}}.$$

More precisely, we have the isomorphism of filtrations (with FI *as well)*

$$\begin{array}{ccccccccc}
\mathcal{F}_n^1(m)_{\mathbb{Q}} & \xrightarrow{\iota_n^1} & \mathcal{F}_n^2(m)_{\mathbb{Q}} & \xrightarrow{\iota_n^2} & \cdots & \xrightarrow{\iota_n^{n-2}} & \mathcal{F}_n^{n-1}(m)_{\mathbb{Q}} & \xrightarrow{\iota_n^{n-2}} & \mathcal{F}_n^n(m)_{\mathbb{Q}} = \mathcal{A}_n(m)_{\mathbb{Q}} \\
\downarrow\cong & & \downarrow\cong & & & & \downarrow\cong & & \| \\
F^1\mathcal{A}_n(m)_{\mathbb{Q}} & \xrightarrow[i_n^1]{} & F^2\mathcal{A}_n(m)_{\mathbb{Q}} & \xrightarrow[i_n^2]{} & \cdots & \xrightarrow[i_n^{n-2}]{} & F^{n-1}\mathcal{A}_n(m)_{\mathbb{Q}} & \xrightarrow[i_n^{n-1}]{} & F^n\mathcal{A}_n(m)_{\mathbb{Q}} = \mathcal{A}_n(m)_{\mathbb{Q}}
\end{array}$$

under which the sections s_n^k and π_n^k coincide.

Proof We omit FI. By definition, $F^k\mathcal{A}_n(m)_{\mathbb{Q}}$ is generated by size k forests in $\mathcal{A}_n(m)_{\mathbb{Q}}$, i.e., there is a linear surjection $\mathbb{Q}\mathcal{D}^k(m) \twoheadrightarrow F^k\mathcal{A}_n(m)_{\mathbb{Q}}$. The relations AS, IHX, STU^2, $\circlearrowleft R$ in $\mathbb{Q}\mathcal{D}^k(m)$ all hold in $\mathcal{A}_n(m)_{\mathbb{Q}}$, and hence this surjection descends to a surjection

$$\mathcal{F}_n^k(m)_{\mathbb{Q}} \twoheadrightarrow F^k\mathcal{A}_n(m)_{\mathbb{Q}}.$$

Those surjections, for $1 \leq k \leq n$, combine into the vertical maps in the commutative diagram of the statement and commute with the maps i_n^k and ι_n^k. By Proposition 2.4.3.3, the maps ι are injective, and hence so are the vertical maps.

To see that the sections s_n^k and π_n^k coincide, compare (2-7) with (2-19). □

2.4.5 The Lie algebra of trees We directly obtain, from the presentation of the primitive filtration of $\mathcal{A}_n(m)_{\mathbb{Q}}$ in Theorem 2.4.4.1, a concrete presentation of the primitive Lie algebra $\mathrm{Prim}(\mathcal{A}(m)_{\mathbb{Q}})$.

Theorem 2.4.5.1 *The primitive Lie algebra* $\mathrm{Prim}(\mathcal{A}^{(\mathrm{FI})}(m)_{\mathbb{Q}})$ *is naturally isomorphic to the* ***Lie algebra of trees***

$$\mathcal{L}^{(\mathrm{FI})}(m)_{\mathbb{Q}} := \mathcal{F}^{(\mathrm{FI}),1}(m)_{\mathbb{Q}} = \mathbb{Q}\mathcal{D}^T(m)/\langle (1\mathrm{T}), \mathrm{AS}, \mathrm{IHX}, \mathrm{STU}^2\rangle$$

endowed with the bracket

$$[\cdot,\cdot]\colon \mathcal{L}^{(\mathrm{FI})}(m)_{\mathbb{Q}} \times \mathcal{L}^{(\mathrm{FI})}(m)_{\mathbb{Q}} \to \mathcal{L}^{(\mathrm{FI})}(m)_{\mathbb{Q}}, \quad (T, T') \mapsto [T, T'] := \overrightarrow{(T'\cdot T)(T\cdot T')},$$

for any $T, T' \in \mathcal{D}^T(m)$ *and extended linearly.*

Proof Under the isomorphism of filtrations from Theorem 2.4.4.1 and by Theorem 2.2.2.3, we have an isomorphism of $\mathbb{Q}$-modules (we omit (FI))

$$\mathcal{L}(m)_{\mathbb{Q}} \cong F^1\mathcal{A}_{\geq 1}(m)_{\mathbb{Q}} \cong \mathrm{Prim}(\mathcal{A}(m)_{\mathbb{Q}})$$

and $\overrightarrow{(T'\cdot T)(T\cdot T')}$ in $\mathcal{L}(m)_{\mathbb{Q}}$ corresponds to $T\cdot T' - T'\cdot T$ in $\mathrm{Prim}(\mathcal{A}(m)_{\mathbb{Q}})$ by definition of $\overrightarrow{(\cdot)}$. Hence $\mathcal{L}(m)_{\mathbb{Q}} \cong \mathrm{Prim}(\mathcal{A}(m)_{\mathbb{Q}})$ as Lie algebras as well. □

3 The rational Goussarov–Habiro Lie algebra of string links

3.1 Clasper calculus and realisation of diagrams

In this section, we describe the *realisation map* $R\colon \mathcal{L}^{\mathrm{FI}}(m)_{\mathbb{Q}} \to \mathcal{L}L(m)_{\mathbb{Q}}$ which sends a tree diagram T to the string link obtained by clasper surgery along a tree clasper realising T. First, we need to make precise what is meant by a tree clasper realising a tree diagram. Then, we need to show that the relations 1T, AS, IHX and STU^2 are satisfied. Most are known and recalled in Proposition 3.1.3.1.

For complete definitions of claspers and clasper surgeries, we refer the reader to the original papers of Habiro [22] and Goussarov [17; 18].

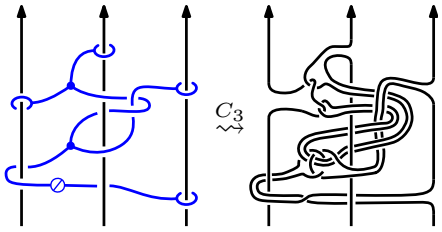


Figure 8: A degree-3 simple tree clasper and the resulting clasper surgery.

3.1.1 Clasper calculus A *simple tree clasper* C on a string link γ is a ribbon uni-trivalent tree diagram embedded in the complement of γ, with *leaves* attached to the strands. See Figure 8, on the left, where γ is the trivial string link on three strands. We use Habiro's drawing convention [22, Figure 7]: nodes (trivalent vertices) are dots, leaves (univalent vertices) are boundaries of disks, each meeting γ transversely exactly once. Ribbon edges have the blackboard framing with half-twists represented by small barred disks, where the bar indicates whether the half-twist is positive or negative (e.g., the clasper in Figure 8 has a *negative* half-twist).

A tree clasper C has a degree $\deg(C) = \#\text{leaves} - 1$, defined as for uni-trivalent tree diagrams. For example, the tree claspers in Figure 3 have degree 1 and 2, respectively, while the one in Figure 8 has degree 3.

Habiro actually defines more general claspers than just simple tree claspers, with boxes and more general leaves. Those appear when manipulating tree claspers. A clasper C determines a *framed* link L_C. Dehn surgery along L_C is called *clasper surgery along* C. A clasper surgery (along a simple clasper) does not modify the ambient manifold (here, the cylinder), but does modify the string link. The result of clasper surgery along a clasper C on γ is denoted γ^C. When γ is the trivial string link $\gamma_0 = 1$, we write $\sigma(C) := \gamma_0^C = 1^C$. See Figures 3 and 8 for examples of clasper surgeries.

Clasper calculus consists in a collection of *moves*, i.e., modifications on claspers that do not alter the resulting surgery γ^C. Those moves are pictured in [22, Proposition 2.7].

Clasper surgeries are entirely supported in a regular neighbourhood of the clasper; hence they define *local* operations on the set of string links and naturally lead to the study of equivalence classes of string links under those moves.

3.1.2 The Goussarov–Habiro Lie algebra of string links Let γ be a string link. A C_n*-move* on γ is a surgery along a degree-n tree clasper on γ. Two string links γ, γ' are said to be C_n*-equivalent*, denoted by $\gamma \overset{C_n}{\sim} \gamma'$, if one can be obtained from the other by applying a sequence of C_n-moves. The C_n-equivalence relation is symmetric by [22, Proposition 3.23]. This defines the Goussarov–Habiro filtration, of the monoid of string links $L(m)$,

$$L(m) = L_1(m) \supset L_2(m) \supset L_3(m) \supset \cdots, \tag{3-1}$$

where $L_n(m) := \{\gamma \in L(m) \mid \gamma \overset{C_n}{\sim} 1\}$ is the submonoid of C_n-trivial string links, i.e., string links that are C_n-equivalent to the trivial string link 1. The quotient monoids $L_k(m)/C_{n+1}$ are finitely generated

nilpotent groups [22, Theorem 5.4]. They are abelian groups when $k=n$, in which case they are denoted by

$$\mathcal{L}_n L(m) := \overline{L_n}(m) := L_n(m)/C_{n+1}, \tag{3-2}$$

and more generally, for any $1 \le k, k' \le n$, one has

$$[L_k(m)/C_{n+1}, L_{k'}(m)/C_{n+1}] \subset L_{k+k'}(m)/C_{n+1}. \tag{3-3}$$

Thus the abelian groups (3-2) combine into a graded Lie algebra, called the *Goussarov–Habiro Lie algebra of string links on m strands*

$$\mathcal{L}L(m) := \bigoplus_{n\ge 1} \mathcal{L}_n L(m) := \bigoplus_{n\ge 1} \overline{L_n}(m),$$

whose Lie bracket is induced from (3-3). Alternatively, the truncation $\mathcal{L}_{\le n}L(m)$ is the graded Lie algebra associated to the N-series $\{L_k(m)/C_{n+1}\}_{1\le k\le n}$ of the group $L(m)/C_{n+1}$. See [29, Theorem 2.1; 31].

3.1.3 Realising tree diagrams in the Goussarov–Habiro Lie algebra The following proposition gathers known relations between string links obtained from clasper surgeries along tree claspers "realising" the relations AS, IHX, STU^2 and STU between diagrams.

Proposition 3.1.3.1 *Let $n \ge 1$. The following relations, involving claspers on the trivial string link 1, hold in $\overline{L_n}(m)$, where the additive notation is used and the C_{n+1}-equivalence class of a string link 1^C is denoted by $[1^C]$:*

(i) *If C and C' are simple tree claspers of degree n that are* ***homotopic*** *with respect to the trivial string link (see [22, §4.1]), then $[1^C]=[1^{C'}]$.*

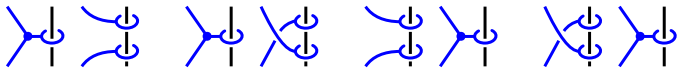
(ii) *If C and C' are simple tree claspers of degree n, that only differ at one place as in Figure 9(b), then $[1^C]+[1^{C'}]=0$ (AS relation). The same holds if C' is obtained from C by adding a half-twist to an edge.*

(iii) *If I, H and X are three simple tree claspers of degree n, that only differ at one place as in Figure 9(c), then $[1^I]-[1^H]+[1^X]=0$ (IHX relation).*

(iv) *If $Y=$, $Y\times$, $=Y$ and $=\times$ are four simple tree claspers of degree n, that only differ at the two places shown in Figure 9(a), then $[1^{Y=}]-[1^{Y\times}]=[1^{=Y}]-[1^{Y\times}]$ (STU^2 relation).*

We also have:

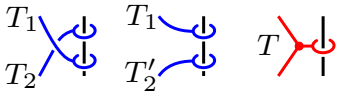
(v) *Let $F^\times = T_1 \cup T_2$, $F^= = T_1 \cup T_2'$, where T_1, T_2, T_2' are three simple tree claspers of degree n_1 (T_1) and n_2 (T_2, T_2'), be such that $F^\times$ and $F^=$ are identical outside of the part shown in Figure 9(d), which involves one leg of T_1 and another of T_2, respectively T_2' (i.e., T_2 and T_2' only differ by the position of one leg with respect to the shown leg of T_1). Then $1^{F_\times} \overset{C_{n+1}}{\sim} 1^{F^=}\cdot 1^T$ where T is the simple tree clasper (in red) of degree $n := n_1+n_2$ obtained by joining T_1 and T_2. We say that $F^=$ and $F^\times$ are related by a* ***slide move****.*



(a) From left to right, the claspers $Y=$, $Y\times$, $=Y$ and $\times Y$

(b) C and C'

(c) I, H and X



(d) $F^{\times}$, $F^{=}$ and T

Figure 9: Illustration of claspers mentioned in Proposition 3.1.3.1.

Proof See [22, Theorem 4.3] for (i), and [22, Theorem 4.7; 38, E.9] for (ii). Proofs and sketches of (iii) can be found in [10; 16, Proposition 3.4.4.1; 19, Theorem 6.6; 32, Lemma 2.9; 38, E.11]. In [26, Section 10.3], (iv) is proved for knots, but the argument applies for string links. For (v), follow the proof of [22, Proposition 4.4] then "separate" T from $F^{=}$ using a homotopy [22, Theorem 4.3] and then remove the two half-twists, which does not change the C_{n+1}-equivalence class of the result. □

We are now able to describe the realisation map from tree diagrams to string links. Similar realisation maps are mentioned by Habiro [22, §8.2] and described by Ohtsuki [38], Conant–Teichner [13; 14] and Kosanovic [26].

Theorem 3.1.3.2 *For each $n \geq 1$, there is a surjective $\mathbb{Q}$-linear morphism*

$$R_n : \mathcal{L}_n^{\mathrm{FI}}(m)_{\mathbb{Q}} = \mathbb{Q}\mathcal{D}_n^T(m)/\langle 1\mathrm{T}, \mathrm{AS}, \mathrm{IHX}, \mathrm{STU}^2\rangle \to \mathcal{L}_n L(m)_{\mathbb{Q}}$$

that sends a tree diagram T to the class of the string link obtained by performing a clasper surgery along a clasper C whose underlying diagram is T (defined in the proof). Those morphisms combine into a surjective morphism of Lie algebras, over $\mathbb{Q}$,

$$R : \mathcal{L}^{\mathrm{FI}}(m)_{\mathbb{Q}} \twoheadrightarrow \mathcal{L}L(m)_{\mathbb{Q}}$$

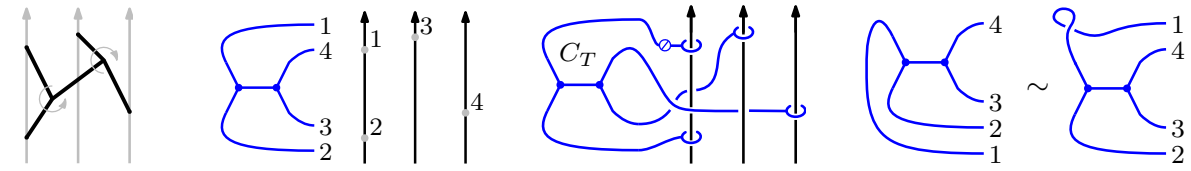
from the primitive Lie algebra of trees (Theorem 2.4.5.1) onto the rational Goussarov–Habiro Lie algebra.

Proof First, we define the realisation map

$$\widetilde{R} : \mathcal{D}_n^T(m) \to \overline{L_n}(m)$$

as follows: Start with a tree diagram $T \in \mathcal{D}_n^T(m)$ as on Figure 10(a). Since T is a tree, its underlying graph is planar and we can pick a planar projection where the cyclic orderings of the nodes are counterclockise. Place that planar projection of the tree to the left of the m vertical strands, with the univalent vertices lying on a vertical line parallel to the strands, as on Figure 10(b). Attach the leaves of the tree to their respective positions on the strand, as given by the diagram T, with an additional (negative or positive) half-twist added to a single chosen leaf[10] (see Figure 10(c)). Outside of that additional half-twist, the attachment

[10]The addition of that additional half-twist is related to the footnote on page 68 in [22]. Although it is not crucial in the case of realising trees, it will be when realising forests later on, in order for the STU relation to hold with the usual signs.



(a) Diagram T (b) Planar tree (c) Clasper C_T (d) Changing the order

Figure 10: Realising a tree diagram.

must be made without introducing half-twists, always going from left to right, and the disk-leaves must clasp to the tangle strands as on Figure 10(c). This yields a simple tree clasper C_T on m strands.

The choices made in the above construction do not alter the result, modulo C_{n+1}-equivalence:

(1) Changing the choice of ordering of the univalent vertices on the vertical line in Figure 10(b) introduces full twists, which do not matter modulo C_{n+1}-moves by (6) in the proof of [22, Theorem 4.3]. See Figure 10(d).

(2) Any two choices of embeddings of the edges joining the planar graph to the disk-leaves are related by a homotopy, which does not change the C_{n+1}-equivalence class of the result by Proposition 3.1.3.1(i).

(3) Adding the half-twist to a different leaf can be obtained by introducing two new half-twists: one to cancel out with the former one, and a new one. Since adding a half-twist corresponds to inversing modulo C_{n+1}-equivalence (Proposition 3.1.3.1(ii)) this choice does not matter modulo C_{n+1}-equivalence. For the same reason, it does not matter whether we chose a positive or negative half-twist.

For a tree clasper C_T in such a position, we say that its *underlying diagram* is $D(C_T) := T$.

Define $\widetilde{R}(T) := \gamma_0^{C_T} = \sigma(C_T)$. Since C_T is a degree-n tree clasper, the C_{n+1}-equivalence class of $\widetilde{R}(T)$ is indeed an element of $\overline{L_n}(m)$.

Since $\overline{L_n}(m)$ is an abelian group, the map $\widetilde{R}$ induces a morphism of abelian groups

$$\widetilde{R} : \mathbb{Z}\mathcal{D}_n^T(m) \to \overline{L_n}(m). \tag{3-4}$$

This morphism sends AS, IHX and STU^2 to zero by Proposition 3.1.3.1(ii), (iii) and (iv), respectively. It also sends 1T to zero; see Figure 11. Thus $\widetilde{R}$ factors through the quotient $\mathbb{Z}\mathcal{D}_n^T(m)/\langle 1\mathrm{T}, \mathrm{AS}, \mathrm{IHX}, \mathrm{STU}^2\rangle$. Moreover, this map is surjective, since $\overline{L_n}(m)$ is generated by string links of the form 1^T where T is a simple tree clasper of degree n [22, Lemma 5.5], and any such clasper can be homotoped [22, Theorem 4.3] to a clasper in the position described in the construction of $\widetilde{R}$ (up to adding and removing half-twists, that is, up to a sign).

After passing to rational coefficients, we obtain the map R_n as in the statement. Those combine into a surjective $\mathbb{Q}$-linear morphism $R : \mathcal{L}^{\mathrm{FI}}(m)_{\mathbb{Q}} \to \mathcal{L}L(m)_{\mathbb{Q}}$. It remains to show that R is compatible with the Lie brackets.

Pick two tree diagrams T, T' of respective degrees k, l, and let $n := k + l$. Denote by the same letters tree claspers realising them.

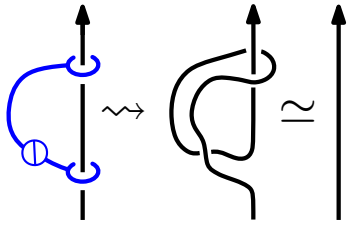


Figure 11: The morphism induced by $\widetilde{R}$ sends 1T to zero.

The commutator $[R(T), R(T')]$ is equal to the C_{n+1}-equivalence class of $\sigma(\overline{T}\,\overline{T'}\,T\,T')$ where $\overline{T}$ represents the inverse of T modulo C_{n+1}-equivalence, idem for T', the product is vertical concatenation and σ denotes surgery modulo C_{n+1}-equivalence. The strategy is to go from the clasper $\overline{T}\,\overline{T'}\,T\,T'$ to the clasper $\overline{T}\,\overline{T'}\,T'\,T$ by a sequence $\epsilon_0, \dots, \epsilon_{r-1}$ of slide moves; see Figure 12 (top). On one hand, $\overline{T}\,\overline{T'}\,T'\,T \overset{C_{n+1}}{\sim} 1$. On the other hand, each slide move creates a red tree clasper $\overrightarrow{e_i}$ (Proposition 3.1.3.1(v)). All in all, we obtain

$$[R(T), R(T')] = \sigma(\overline{T}\,\overline{T'}\,T\,T') = \sum_{i=1}^{r-1} \sigma(\overrightarrow{e_i}). \tag{3-5}$$

Let us compare (3-5) with the commutator of the original tree diagrams T, T'. We can follow the exact same path of slide moves on the diagrammatic level to get

$$[T', T] = \overrightarrow{(T\,T')(T'T)} = \sum_{i=1}^{r-1} \overrightarrow{e_i} \tag{3-6}$$

by Definition 2.3.2.7.

In (3-5) and (3-6) the roles of T and T' are reversed. This justifies the additional half-twist in the definition of the realisation map. More precisely, the tree claspers realising T, T' both have an additional

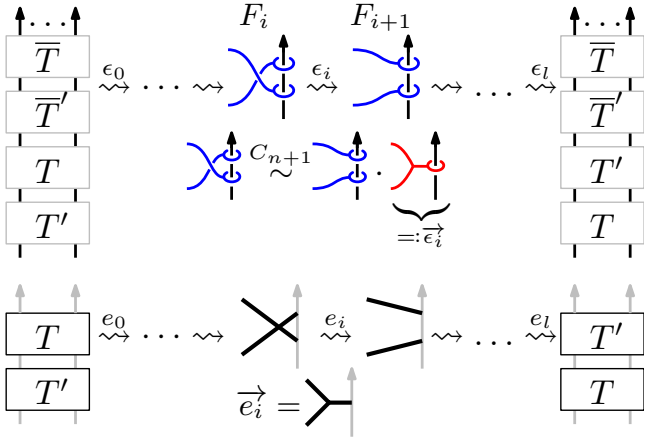


Figure 12: Commutator of string links (top) versus tree diagrams (bottom).

half-twist introduced somewhere. Thus, the red tree clasper $\overrightarrow{\epsilon_i}$ has two half-twists, or zero, modulo C_{n+1}. This means that $R(\overrightarrow{e_i}) = -\sigma(\overrightarrow{\epsilon_i})$ and therefore

$$R([T,T']) = -R([T',T]) = -\sum_i R(\overrightarrow{e_i}) = \sum_i \sigma(\overrightarrow{\epsilon_i}) = [R(T), R(T')],$$

which concludes the proof that R is a surjective morphism of Lie algebras over $\mathbb{Q}$. □

3.2 The Kontsevich integral and the tree preservation theorem

The *Kontsevich integral* is the last ingredient of our identification of the rational Goussarov–Habiro Lie algebra. Discovered by Kontsevich [25] as a universal Vassiliev invariant over $\mathbb{Q}$, it is a morphism

$$Z : L(m) \to \widehat{\mathcal{A}^{\mathrm{FI}}}(m)_{\mathbb{Q}},$$

which to a string link $\gamma \in L(m)$ associates an element $Z(\gamma)$ in the graded completion $\widehat{\mathcal{A}^{\mathrm{FI}}}(m)_{\mathbb{Q}}$ of $\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}}$. The Kontsevich integral of a string link $Z(\gamma)$ can be thought of as a *series expansion* of γ, such that degree $\leq n$ Vassiliev invariants only depend on the degree $\leq n$ part $Z_{\leq n}(\gamma)$. However, it is an open question whether Z is injective on $L(m)$, i.e., whether the Kontsevich integral classifies string links.

Since the Kontsevich integral turns string links into diagrams, it is a good candidate for inducing an inverse to the realisation map R. This is indeed the case, as we will see in this section.

3.2.1 The combinatorial Kontsevich integral The combinatorial version of the Kontsevich integral [2] is easier to work with. The combinatorial Kontsevich integral is understood as a functor whose domain is the category **PTangles** of *parenthesized tangles* up to isotopy (denoted by **PT** in [2]), i.e., whose objects are parenthesized $\updownarrow$-words (i.e., finite words in the alphabet $\{\uparrow, \downarrow\}$, together with a parenthesizing of their letters [2, §2.1]) and whose morphisms are isotopy classes of unframed tangles with the adequate $\updownarrow$-words as boundaries, also referred to as *q-tangles* [30]. Its target is the category $\widehat{\mathbf{Diag}}_{\mathbb{Q}}$ of uni-trivalent diagrams [2, Definition 3.1] on tangle skeletons, whose objects are (nonparenthesized) $\updownarrow$-words and morphisms are formal series in $\widehat{\mathcal{A}^{\mathrm{FI}}}(S)_{\mathbb{Q}}$ for tangle skeleta S with the adequate $\updownarrow$-words as boundaries. Composition in **PTangles** is simply vertical concatenation. Composition in $\widehat{\mathbf{Diag}}_{\mathbb{Q}}$ is given by concatenation of the tangle skeleta and concatenation of the diagrams (extended linearly).

Theorem 3.2.1.1 [2; 7; 15; 38] *The Kontsevich integral of unframed parenthesized tangles*

$$Z : \mathbf{PTangles} \to \widehat{\mathbf{Diag}}_{\mathbb{Q}}$$

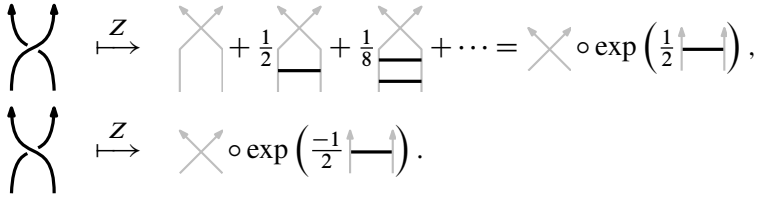
has the following properties:

(i) *It is a* ***functor***: *If $\gamma = \mathrm{id}_w$ for some parenthesized $\updownarrow$-word w, i.e., it consists of parallel vertical strands, then $Z(\gamma) = 1 \in \widehat{\mathcal{A}^{\mathrm{FI}}}(m)$ is the empty diagram. If γ_1, γ_2 are two concatenable tangles (i.e., composable in* **PTangles**), *then $Z(\gamma_1 \circ \gamma_2) = Z(\gamma_1) \circ Z(\gamma_2)$.*

(ii) *It is a* ***monoidal*** *functor, where the monoidal structure on both sides is given by juxtaposition, i.e., horizontal concatenation*:[11] *$Z(\gamma_1 \otimes \gamma_2) = Z(\gamma_1) \otimes Z(\gamma_2)$.*

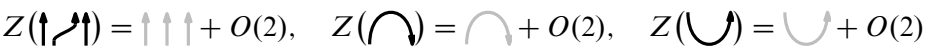
[11] In **PTangles**, the parenthesizing of the tensor product is given by $u \otimes v := ((u)(v))$.

(iii) *It takes, on the standard* ***braiding*** *tangles (which are elements of* $\mathbf{PTangles}(\uparrow\uparrow,\uparrow\uparrow)$*), the values*



(iv) *On the standard* ***associator***, ***cap*** *and* ***cup*** *parenthesized tangles,*

it is trivial modulo terms of degree ≥ 2*:*



(v) *It is compatible with the operation of* ***reversing a strand****. If* $\boldsymbol{R}_C\gamma$ *is obtained from* γ *by reversing the orientation of the component* C*, then*

$$Z(\boldsymbol{R}_C\gamma) = (\boldsymbol{R}_C)_* Z(\gamma),$$

where $(R_C)_* : \widehat{\mathcal{A}^{\mathrm{FI}}}(S)_{\mathbb{Q}} \to \widehat{\mathcal{A}^{\mathrm{FI}}}(\boldsymbol{R}_C S)_{\mathbb{Q}}$ *is defined as follows. For a diagram* D *on the tangle skeleton* $S = S(\gamma)$*,*

$$(\boldsymbol{R}_C)_*(D) = (-1)^{\#\{\text{endpoints on } C\}} D' \in \widehat{\mathcal{A}^{\mathrm{FI}}}(\boldsymbol{R}_C S)_{\mathbb{Q}},$$

where D' *is the same diagram as* D *on the tangle skeleton* $\boldsymbol{R}_C S = S(\boldsymbol{R}_C\gamma)$*.*

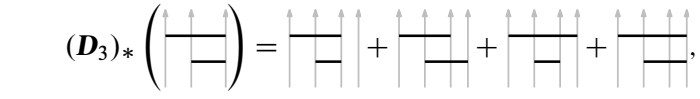
(vi) *It is compatible with the operation of* ***doubling a strand****. If* $\boldsymbol{D}_C\gamma$ *is obtained from* γ *by doubling*[12] *the strand* C*, then*

$$Z(\boldsymbol{D}_C\gamma) = (\boldsymbol{D}_C)_* Z(\gamma),$$

where $(D_C)_* : \widehat{\mathcal{A}^{\mathrm{FI}}}(S)_{\mathbb{Q}} \to \widehat{\mathcal{A}^{\mathrm{FI}}}(\boldsymbol{D}_C S)_{\mathbb{Q}}$ *is defined as follows. For a diagram* D *on the tangle skeleton* $S = S(\gamma)$*,*

$$(\boldsymbol{D}_C)_*(D) = \sum_{D' \text{ lifts } D} D' \in \widehat{\mathcal{A}^{\mathrm{FI}}}(\boldsymbol{R}_C S)$$

is the sum of all the $2^{\#\{\text{endpoints on } C\}}$ *ways of "lifting"* D *to a diagram on* $\boldsymbol{D}_C\gamma$*. For example,*



where $\boldsymbol{D}_3$ *is the operation of doubling the third strand.*

As explained in [8, Section 10.3.2], any parenthesized tangle can be decomposed into a product of tensor products of basic tangles: braidings, cups, caps, associators, or images of those under reversing

[12] If $\uparrow$ is an endpoint of C one of the end $\updownarrow$-words, then it is replaced by $(\uparrow\uparrow)$ in $\mathbf{D}_C\gamma$, and similarly for $\downarrow$.

and doubling operations. Knowing the value of Z on each of those standard tangles suffices to deduce the value of Z on any tangle.

It should be mentioned that different choices of value of Z on the associator tangle lead to different combinatorial Kontsevich integrals. The possible choices are called *Drinfeld's associators*, of which there is a deep theory (see [8] for more details). For our purpose, it only suffices to know that they are all trivial modulo terms of degree ≥ 2. This also allows us to ignore the parenthesizing of the extremities of the string links considered in the theorem below.

Denote by $Z_{\leq n}$, Z_n and $Z_{\geq n}$ the compositions proj $\circ\, Z$, where proj is the projection onto the corresponding direct summands.

3.2.2 Tree preservation theorem

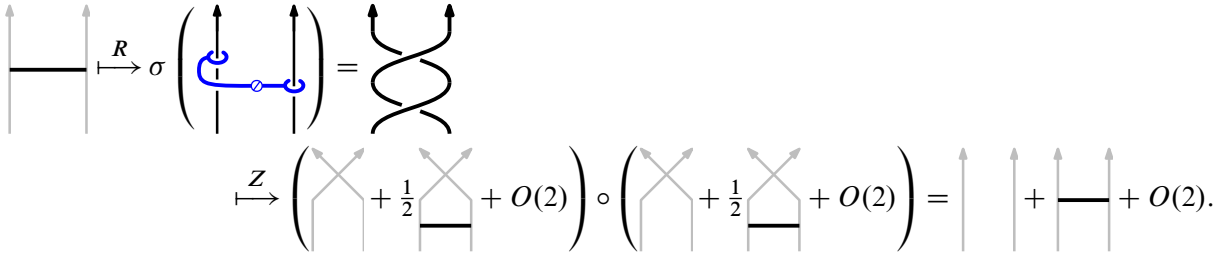
Theorem 3.2.2.1 *Let $n \geq 1$ and let $T \in \mathcal{D}_n^T(m)$ be a degree-n tree diagram on m strands. Then*

$$Z(R(T)) = 1 + T + O(n+1), \tag{3-7}$$

where $R(T) = \sigma(C_T)$ is the result of surgery along a tree clasper C_T realising T (see Theorem 3.1.3.2), and the notation $O(n+1)$ means that the equality holds modulo terms of degree $\geq n+1$. In particular, the Kontsevich integral induces an inverse to R, which then provides an isomorphism of graded Lie $\mathbb{Q}$-algebras $\mathcal{L}^{\mathrm{FI}}(m)_{\mathbb{Q}} \cong \mathcal{L}L(m)_{\mathbb{Q}}$.

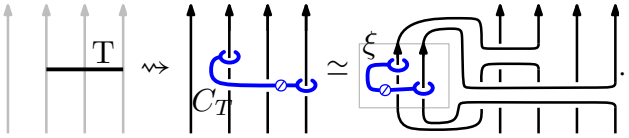
Proof We proceed by induction on n. Moreover, at each step we first make the simplifying assumption (**A**): $m = n+1$, T has one leg on each strand and (if $n > 1$) the legs attached to the rightmost two strands are adjacent to a common node. Then we show the result for the general case.

For $n = 1$ and assuming (**A**), there is only one possible diagram $T \in \mathcal{D}_1^T(2)$ with a leg on each strand. Its realisation is a full positive braiding between the two strands, and hence we obtain the desired result by Theorem 3.2.1.1(iii):

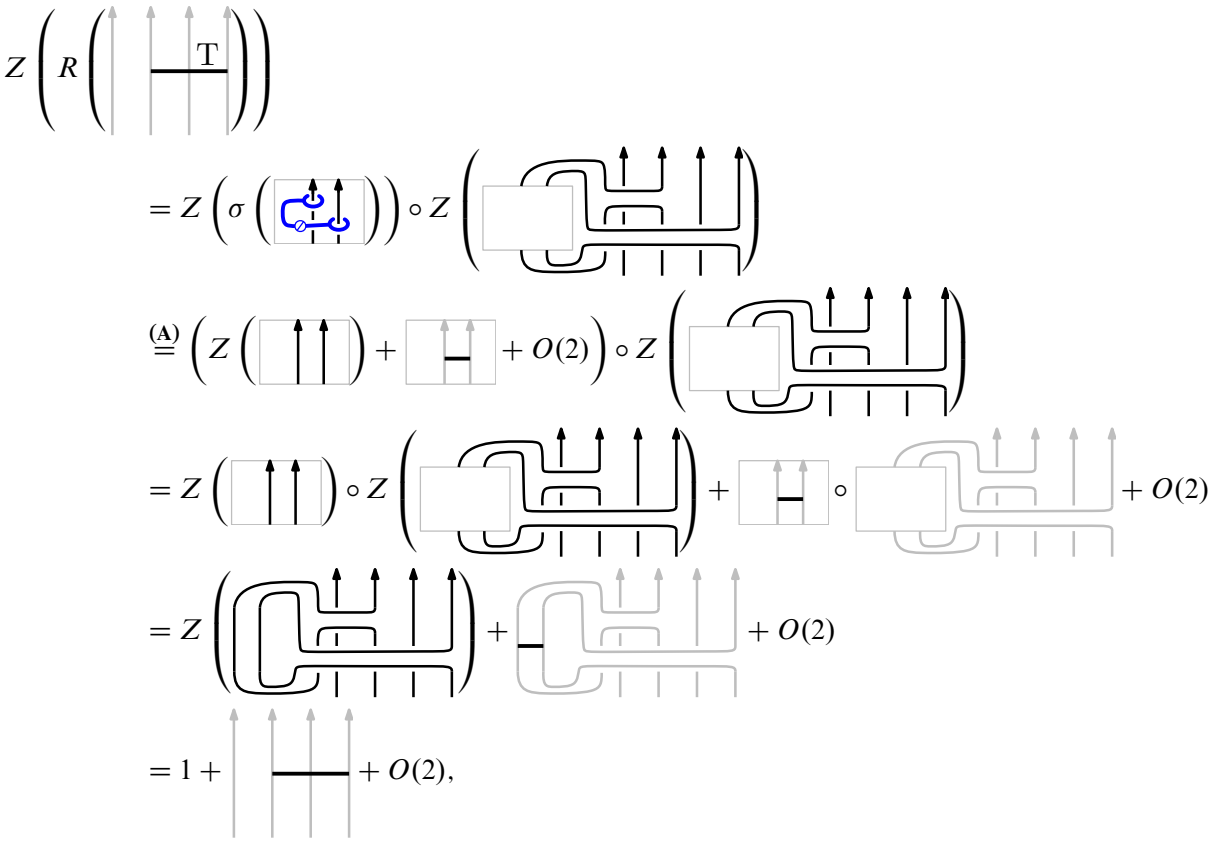


Now we remove the assumption (**A**), and hence the single chord of T can have its legs attached anywhere on the m strands. Consider a clasper C_T realising T, as in the construction of the realising map R; see Figure 10. By an isotopy, we can pull the clasper out and bring it into a small box, such that the box contains 2 parallel vertical strands and the clasper C_T is attached to them satisfying (**A**):

(3-8)



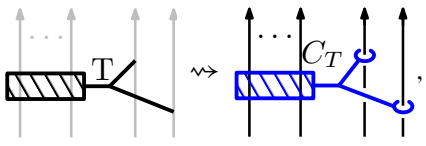
The grey box consists of a string link ξ with a degree-1 tree clasper C_T on it, satisfying (**A**). Thus we compute



where the first equality uses (3-8) and multiplicativity. Then the left factor can be computed by the above since (**A**) is satisfied in the small grey box. Then we distribute the product over the sum on the left to obtain the second line, where everything that has degree ≥ 2 is contained in $O(2)$. Use multiplicativity again to reconstruct the first term as the Kontsevich integral of a string link isotopic to the unlink, which becomes 1, while the second term gives the diagram T when attached to the outside. This concludes the proof of the case $n = 1$.

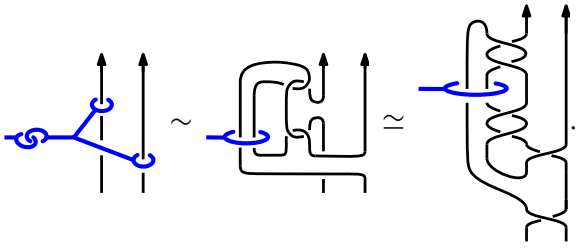
Now we show the induction step. Let $n \geq 2$ and assume that the statement holds in degree $n-1$.

First, consider a degree-n tree diagram $T \in \mathcal{D}_n^T(n+1)$ on $n+1$ strands and satisfying (**A**). It looks like the diagram on the left of



where the black dashed rectangle contains the $n-1$ other legs. Realise it as a tree clasper C_T, assuming that the additional half-twist is in the blue blob. The core of the induction step is to apply Habiro's move 2

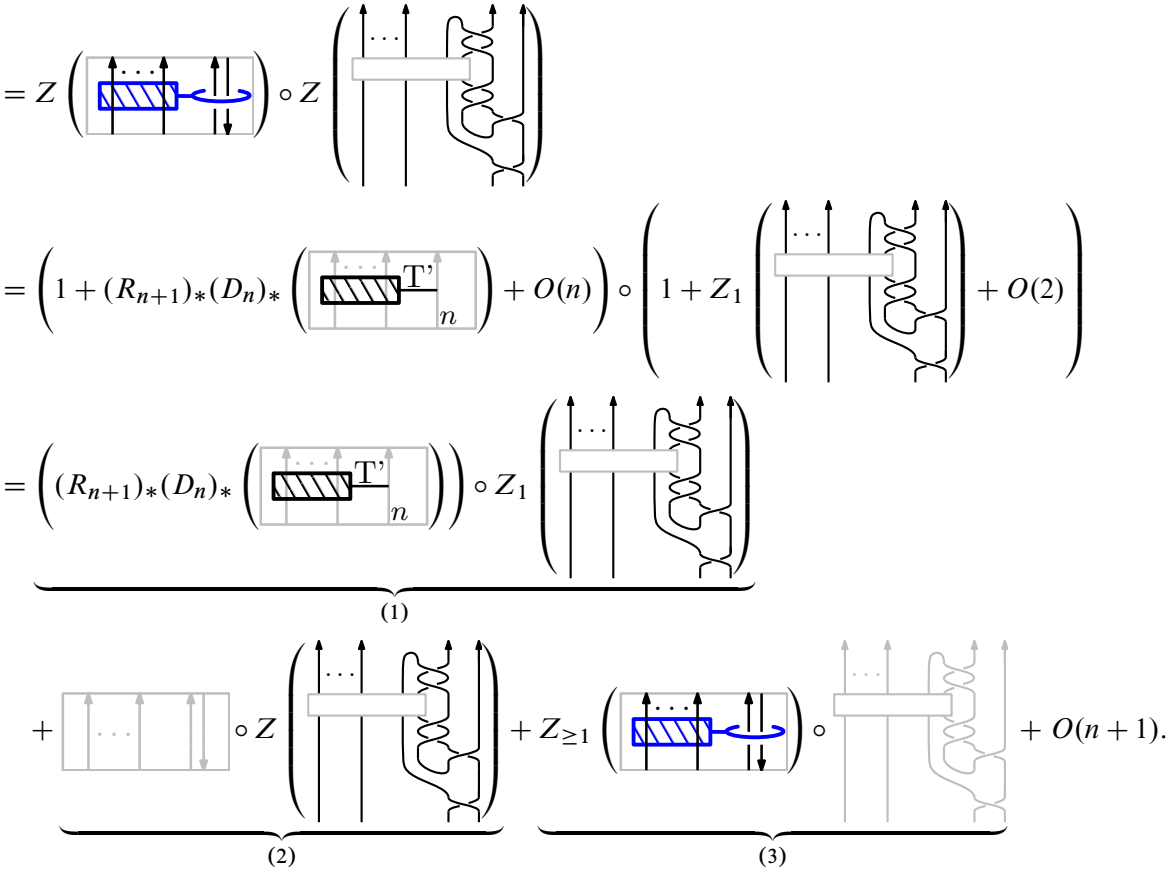
and 10 [22, pages 14, 15] to the visible tripod:



The right-hand side is obtained by isotopy. The new tree clasper C' (C_T minus the tripod) realises a degree-$(n-1)$ tree diagram T'.

Thus, to compute $R(T)$ we can first apply moves 2 and 10 as above, then perform surgery on C'. This surgery is entirely contained in the grey rectangle shown in the first line below:

$Z(R(T))$

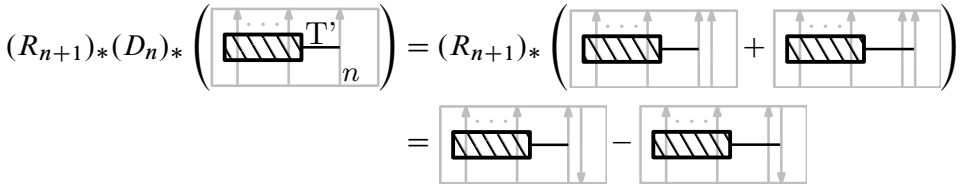


We obtain the first equality by multiplicativity of Z. From the second to the third line, the induction hypothesis

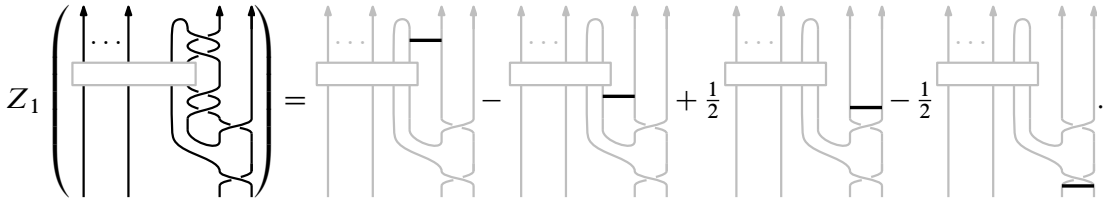
$$Z(R(T')) = 1 + T' + O(n)$$

is used, with additional strand doubling and reversing. Then the fourth line is obtained by distributing this product (Recall that 1 stands for the empty diagram on the adequate tangle skeleton), and putting all the terms of degree $\geq n+1$ inside $O(n+1)$. Let us compute the resulting three terms, denoted (1), (2) and (3), respectively:

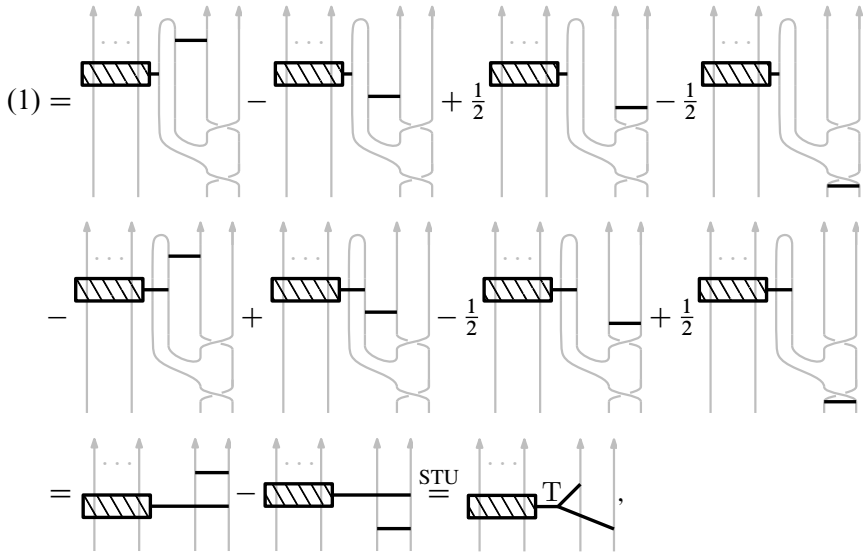
(1) The first term is a product of degree-$(n-1)$ terms with degree-1 terms. First, we compute those two factors individually. Since we are only computing the degree-1 part of Z, only the braidings matter. The first factor is obtained by doubling the n-th strand and reversing the new $(n+1)$-st strand



and the second factor is a sum of Kontsevich integrals of braidings



Combining them yields



where the last three columns vanish, since the leg attached to the blob is free to slide across the maximum. In the first column, the additional chord prevents that leg from sliding. This column gives the tree diagram T by STU.

(2) The left factor is Z applied to a trivial string link, with reversed rightmost strand. By multiplicativity of Z, we can combine the two factors into

$$Z\left(\cdots\right) = Z\left(\cdots\right) = 1.$$

(3) Since terms (1) and (2) equal T and 1, respectively, we expect (3) to vanish. Although the first factor might be very complicated, all we know is that it is obtained from a sum of diagrams by doubling the last strand, then reversing the new last strand. This fact alone actually suffices to show that (3) vanishes.

Indeed, suppose that D is any nonempty diagram on n strands, with $k \geq 1$ legs attached to the last strand. Doubling the last strand turns that diagram into a sum of all its 2^k lifts. Such a lift is indexed by a word $e_1 \dots e_k \in \{0,1\}^k$ where an $e_i = 0$, resp. $e_i = 1$, means that the i-th leg is attached to the n-th, resp. $(n+1)$-st strand. Reversing the last strand has the effect of multiplying by ± 1, depending on the number of legs attached to the last strand. The resulting sum

$$(R_{n+1})_*(D_n)_*\left(\; D \;\right) = \sum_{e_2,\dots,e_k \in \{0,1\}} (-1)^{\sum_{j\geq 2} e_j} \left(\;\; - \;\; \right)$$

is a sum of differences of the form "$e_1 = 0$" minus "$e_1 = 1$". Multiplying each such difference with the second factor of (3) gives 0 since the first leg is now free to go from one position to the other:

$$\left(\;\; - \;\; \right) \circ \; = 0.$$

The above shows that all terms in $Z(R(T'))$ that have at least one leg on the last strand vanish once we apply $(R_{n+1})_*(D_n)_*$ and multiply the result with the second factor.

In fact, all the terms in $Z(R(T'))$ must have at least 1 leg on the last strand. Indeed, by definition of the Kontsevich integral, the sum of all terms with no leg on the last strand is equal to the Kontsevich integral of $R(T')$ minus the last strand. If the last strand is removed, then T' gets a trivial leaf and the surgery is trivial. In other words, this comes from the fact that a string link obtained by surgery along a tree clasper with exactly one leg on each strand is a *Brunnian string link* [24].

This concludes the proof that (3) vanishes.

By the above computations of (1), (2) and (3), we obtain

$$Z(R(T)) = 1 + T + O(n+1)$$

as desired. This concludes the case (**A**) in degree n.

The general case is shown exactly as in degree 1, by pulling the clasper out to extract a box containing the clasper (as in (3-8)), in which assumption (**A**) is satisfied. This concludes the induction step, and the proof of (3-7).

It remains to show that Z induces an inverse to R.

By (3-7), the degree-n part of Z induces a map $Z_n : L_n(m) \to \mathcal{A}_n^{\mathrm{FI}}(m)_{\mathbb{Q}}$, which is constant on C_{n+1}-equivalence classes. Thus it factors through

$$Z_n^{GH} : \mathcal{L}_n L(m) \to \mathcal{A}_n^{\mathrm{FI}}(m)_{\mathbb{Q}},$$

whose image lies in $\mathrm{Prim}(\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}})_n$. Identifying the latter space with $\mathcal{L}_n^{\mathrm{FI}}(m)_{\mathbb{Q}}$ by Theorem 2.4.5.1, direct summing all the degrees at once and tensoring with $\mathbb{Q}$, we finally obtain

$$Z^{GH} : \mathcal{L}L(m)_{\mathbb{Q}} \to \mathcal{L}_n^{\mathrm{FI}}(m)_{\mathbb{Q}},$$

which is an inverse to R by (3-7). □

4 Application to the rational Goussarov–Habiro conjecture

4.1 The rational Goussarov–Habiro conjecture

4.1.1 Vassiliev filtration and associated graded algebra The monoid ring $\mathbb{Z}L(m)$ of string links is filtered by the *Vassiliev filtration* (4-1), with associated graded algebra $\mathcal{A}L(m)$ (4-2):

$$\mathbb{Z}L(m) \supset \partial_1(\mathbb{Z}S_1L(m)) \supset \partial_2(\mathbb{Z}S_2L(m)) \supset \cdots, \tag{4-1}$$

$$\mathcal{A}L(m) := \bigoplus_{n\geq 0} \mathcal{A}_n L(m) := \bigoplus_{n\geq 0} \frac{\partial_n(\mathbb{Z}S_nL(m))}{\partial_{n+1}(\mathbb{Z}S_{n+1}L(m))}, \tag{4-2}$$

where $S_\bullet L(m)$ denotes the graded monoid of (isotopy classes of) singular string links (where the only allowed singularities are transversal double points), graded by the number of double points. The $\mathbb{Z}$-linear maps $\partial_n : \mathbb{Z}S_nL(m) \to \mathbb{Z}L(m)$ are defined by applying the *Skein relation* [3]

(4-3)

to each double point, where it is understood that the terms are identical outside of the part shown. Thus, for $\gamma \in S_nL(m)$ a singular string link with n double points, its image $\partial_n(\gamma) \in \mathbb{Z}L(m)$ is an alternating sum of all possible resolutions of the n double points.

A string link invariant $V : L(m) \to A$ valued in an abelian group A naturally extends to a Z-linear map $V : \mathbb{Z}L(m) \to A$. We say that V is a *Vassiliev invariant* or *finite type invariant* of degree n if it vanishes on $\partial_{n+1}(\mathbb{Z}S_{n+1}L(m))$. In particular, a Vassiliev invariant of degree n induces a functional on $\mathcal{A}_nL(m)$. Two string links $\gamma, \gamma' \in L(m)$ are *V_n-equivalent* if they cannot be distinguished by Vassiliev invariants of degree $< n$ or, equivalently, if $\gamma - \gamma'$ belongs to $\partial_n(\mathbb{Z}S_nL(m))$.

There is a realisation map $R^v : \mathcal{A}^{\mathrm{FI}}(m) \twoheadrightarrow \mathcal{A}L(m)$ (v refers to *Vassiliev*), which is a morphism of $\mathbb{Z}$-algebras. Originally, R^v was defined on the presentation of $\mathcal{A}^{\mathrm{FI}}(m)$ by chord diagrams, by sending

a chord diagram to a singular string link realising it [1]. However, one can define this realisation map on forest diagrams using claspers [13, Theorem 1.1; 22, Section 6], in order to construct a surjective morphism of graded $\mathbb{Z}$-algebras

$$R^v : \mathcal{A}^{\mathrm{FI}}(m) \twoheadrightarrow \mathcal{A}L(m), \tag{4-4}$$

which in degree n sends a forest diagram $F \in \mathcal{D}_n(m)$ to

$$[\gamma_0; T_1, \dots, T_s] := \sum_{\mathcal{S}' \subset \{T_1, \dots, T_s\}} (-1)^{s-|\mathcal{S}'|} (\gamma_0^{\mathcal{S}'}) \in \mathcal{A}_n L(m),$$

where $T_1 \cup \cdots \cup T_s$ is a forest clasper realising F, γ_0 denotes the trivial string link on m strands and $\gamma_0^{\mathcal{S}'}$ denotes the result of clasper surgery on γ_0 along the trees contained in $\mathcal{S}'$.

Remark 4.1.1.1 A forest clasper that realises a forest diagram F is constructed as in the proof of Theorem 3.1.3.2, but with multiple trees at the same time and with an additional positive half-twist near a chosen leaf of each tree. Those half-twists are important in order for STU relations to be realised with correct signs. This explains the footnote on page 68 in [22]. △

Over $\mathbb{Q}$, the realisation map $R^v : \mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}} \xrightarrow{\cong} \mathcal{A}L(m)_{\mathbb{Q}}$ is an isomorphism, with inverse induced by the Kontsevich integral [1; 25].

4.1.2 The Goussarov–Habiro conjecture From the definition of the Vassiliev filtration using claspers, it follows that C_n-equivalence implies V_n-equivalence for string links in $L(m)$. Goussarov and Habiro independently conjectured the converse:

Conjecture (Goussarov–Habiro conjecture for string links in the cylinder) *For all $m \geq 1$ and $n \geq 1$, any V_n-equivalent string links on m strands are C_n-equivalent.*

Since C_n-equivalence implies V_n-equivalence, any C_n-trivial string link $\gamma \in L_n(m)$ is automatically V_n-equivalent to the trivial string link 1, and hence $\gamma - 1$ belongs to $\partial_n(\mathbb{Z}S_n L(m))$. This yields the *comparison map* [22, Section 8.2]

$$\chi : \mathcal{L}L(m) \to \mathcal{A}L(m) : [\gamma] \mapsto [\gamma - 1],$$

which is injective if and only if the Goussarov–Habiro conjecture is true. In particular, it is injective when $m = 1$ by the Goussarov–Habiro theorem for knots [22, Theorem 6.18].

Exploiting the *dimension subgroup property* in characteristic zero, Massuyeau showed the following rational version of the Goussarov–Habiro conjecture:

Theorem 4.1.2.1 (Massuyeau [31]) *Over $\mathbb{Q}$, the map $\chi : \mathcal{L}L(m)_{\mathbb{Q}} \to \mathcal{A}L(m)_{\mathbb{Q}}$ is injective.*

Let us now give an alternative proof of Massuyeau's theorem, following from the main result of this paper. Once we pass to rational coefficients, the realisation maps R, R^v and the comparison map χ

combine into the commutative square of graded (Lie or Hopf) algebras

$$\begin{array}{ccccc}
\mathcal{L}^{\mathrm{FI}}(m)_{\mathbb{Q}} & \xrightarrow[\cong]{\iota} & \mathrm{Prim}(\mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}}) & \hookrightarrow & \mathcal{A}^{\mathrm{FI}}(m)_{\mathbb{Q}} \\
R\downarrow\cong & \nearrow_{Z^{GH}} & & & R^{v}\downarrow\cong\uparrow Z^{v} \\
\mathcal{L}L(m)_{\mathbb{Q}} & & \xrightarrow{\quad\chi\quad} & & \mathcal{A}L(m)_{\mathbb{Q}}
\end{array}$$

where Z^v is the inverse to R^v by Kontsevich's theorem, ι is an isomorphism by Theorem 2.4.5.1, the left triangle commutes and R becomes an isomorphism by Theorem 3.2.2.1. Therefore, χ is injective, which concludes the alternative proof of the rational Goussarov–Habiro conjecture.

Acknowledgements

This work was done under the supervision of Peter Teichner, whom I wish to thank for his guidance. I also wish to thank James Conant for helpful discussions and comments. I also thank the anonymous referee for helpful comments and corrections. Part of the writing was supported by FNR AFR grant 18890152.

References

[1] **D Bar-Natan**, *On the Vassiliev knot invariants*, Topology 34:2 (1995) 423–472 MR

[2] **D Bar-Natan**, *Non-associative tangles*, from "Geometric topology" (Athens, GA, 1993) (W H Kazez, editor), AMS/IP Stud. Adv. Math. 2.1, Amer. Math. Soc., Providence, RI (1997) 139–183 MR

[3] **J S Birman**, **X-S Lin**, *Knot polynomials and Vassiliev's invariants*, Invent. Math. 111:2 (1993) 225–270 MR

[4] **P Boavida de Brito**, **G Horel**, *Galois symmetries of knot spaces*, Compos. Math. 157:5 (2021) 997–1021 MR

[5] **R Budney**, **J Conant**, **R Koytcheff**, **D Sinha**, *Embedding calculus knot invariants are of finite type*, Algebr. Geom. Topol. 17:3 (2017) 1701–1742 MR

[6] **P Cartier**, *A primer of Hopf algebras*, from "Frontiers in number theory, physics, and geometry, II" (P Cartier, B Julia, P Moussa, P Vanhove, editors), Springer (2007) 537–615 MR

[7] **S Chmutov**, **S Duzhin**, *The Kontsevich integral*, Acta Appl. Math. 66:2 (2001) 155–190 MR

[8] **S Chmutov**, **S Duzhin**, **J Mostovoy**, *Introduction to Vassiliev knot invariants*, Cambridge Univ. Press (2012) MR

[9] **J Conant**, *Homotopy approximations to the space of knots, Feynman diagrams, and a conjecture of Scannell and Sinha*, Amer. J. Math. 130:2 (2008) 341–357 MR

[10] **J Conant**, **R Schneiderman**, **P Teichner**, *Jacobi identities in low-dimensional topology*, Compos. Math. 143:3 (2007) 780–810 MR

[11] **J Conant**, **R Schneiderman**, **P Teichner**, *Whitney tower concordance of classical links*, Geom. Topol. 16:3 (2012) 1419–1479 MR

[12] **J Conant**, **R Schneiderman**, **P Teichner**, *Geometric filtrations of string links and homology cylinders*, Quantum Topol. 7:2 (2016) 281–328 MR

[13] **J Conant**, **P Teichner**, *Grope cobordism and Feynman diagrams*, Math. Ann. 328:1-2 (2004) 135–171 MR

[14] **J Conant**, **P Teichner**, *Grope cobordism of classical knots*, Topology 43:1 (2004) 119–156 MR

[15] **V G Drinfeld**, *On quasi-triangular quasi-Hopf algfebras and a group closely connected with* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, Leningrad Math. J. 2:4 (1991) 829–860

[16] **B Dular**, *The Lie algebra of tree diagrams and the rational Goussarov–Habiro conjecture*, Master's thesis (2023) Available at https://sites.google.com/view/brunodular/research

[17] **M N Goussarov**, *Interdependent modifications of links and invariants of finite degree*, Topology 37:3 (1998) 595–602 MR

[18] **M Goussarov**, **M Polyak**, **O Viro**, *Finite-type invariants of classical and virtual knots*, Topology 39:5 (2000) 1045–1068 MR

[19] **M N Gusarov**, *Variations of knotted graphs: the geometric technique of n-equivalence*, Algebra i Analiz 12:4 (2000) 79–125 MR In Russian; translated in St. Petersburg Math. J. 12 (2001) 569–604

[20] **N Habegger**, **X-S Lin**, *The classification of links up to link-homotopy*, J. Amer. Math. Soc. 3:2 (1990) 389–419 MR

[21] **N Habegger**, **G Masbaum**, *The Kontsevich integral and Milnor's invariants*, Topology 39:6 (2000) 1253–1289 MR

[22] **K Habiro**, *Claspers and finite type invariants of links*, Geom. Topol. 4 (2000) 1–83 MR

[23] **K Habiro**, **G Massuyeau**, *Symplectic Jacobi diagrams and the Lie algebra of homology cylinders*, J. Topol. 2:3 (2009) 527–569 MR

[24] **K Habiro**, **J-B Meilhan**, *On the Kontsevich integral of Brunnian links*, Algebr. Geom. Topol. 6 (2006) 1399–1412 MR

[25] **M Kontsevich**, *Vassiliev's knot invariants*, from "I M Gelfand Seminar, II" (S Gelfand, S Gindikin, editors), Adv. Soviet Math. 16, Amer. Math. Soc., Providence, RI (1993) 137–150 MR

[26] **D Kosanović**, *A geometric approach to the embedding calculus knot invariants*, PhD thesis, Universität Bonn (2020) Available at https://bonndoc.ulb.uni-bonn.de/xmlui/bitstream/handle/20.500.11811/8651/5979.pdf

[27] **P Lambrechts**, **V Turchin**, **I Volić**, *Associahedron, cyclohedron and permutohedron as compactifications of configuration spaces*, Bull. Belg. Math. Soc. Simon Stevin 17:2 (2010) 303–332 MR

[28] **S K Lando**, *On primitive elements in the bialgebra of chord diagrams*, from "Topics in singularity theory" (A Khovanskiĭ, A Varchenko, V Vassiliev, A B Sossinsky, editors), Amer. Math. Soc. Transl. Ser. 2 180, Amer. Math. Soc., Providence, RI (1997) 167–174 MR

[29] **M Lazard**, *Sur les groupes nilpotents et les anneaux de Lie*, Ann. Sci. École Norm. Sup. (3) 71 (1954) 101–190 MR

[30] **T Q T Le**, **J Murakami**, *The universal Vassiliev–Kontsevich invariant for framed oriented links*, Compositio Math. 102:1 (1996) 41–64 MR

[31] **G Massuyeau**, *Finite-type invariants of 3-manifolds and the dimension subgroup problem*, J. Lond. Math. Soc. (2) 75:3 (2007) 791–811 MR

[32] **J-B Meilhan**, *On surgery along Brunnian links in 3-manifolds*, Algebr. Geom. Topol. 6 (2006) 2417–2453 MR

[33] **J-B Meilhan**, *Linking number and milnor invariants*, from "Encyclopedia of knot theory" (C Adams, E Flapan, A Henrich, L H Kauffman, L D Ludwig, S Nelson, editors), Chapman and Hall/CRC (2021) 817–830

[34] **J-B Meilhan**, **A Yasuhara**, *Characterization of finite type string link invariants of degree < 5*, Math. Proc. Cambridge Philos. Soc. 148:3 (2010) 439–472 MR

[35] **J W Milnor**, **J C Moore**, *On the structure of Hopf algebras*, Ann. of Math. (2) 81 (1965) 211–264 MR

[36] **S Montgomery**, *Hopf algebras and their actions on rings*, CBMS Regional Conference Series in Mathematics 82, Amer. Math. Soc., Providence, RI (1993) MR

[37] **Y Nozaki**, **M Sato**, **M Suzuki**, *On the kernel of the surgery map restricted to the 1-loop part*, J. Topol. 15:2 (2022) 587–619 MR

[38] **T Ohtsuki**, *Quantum invariants: a study of knots, 3-manifolds, and their sets*, Series on Knots and Everything 29, World Scientific, River Edge, NJ (2002) MR

[39] **A Postnikov**, *Permutohedra, associahedra, and beyond*, Int. Math. Res. Not. 2009:6 (2009) 1026–1106 MR

[40] **M Ronco**, *Shuffle bialgebras*, Ann. Inst. Fourier (Grenoble) 61:3 (2011) 799–850 MR

[41] **Y Shi**, *Goodwillie's cosimplicial model for the space of long knots and its applications*, J. Homotopy Relat. Struct. 18:2–3 (2023) 265–312 MR

[42] **V G Turaev**, *Quantum invariants of knots and 3-manifolds*, 3rd edition, De Gruyter Studies in Mathematics 18, De Gruyter, Berlin (2016) MR

[43] **V A Vassiliev**, *Cohomology of knot spaces*, from "Theory of singularities and its applications" (V I Arnold, editor), Adv. Soviet Math. 1, Amer. Math. Soc., Providence, RI (1990) 23–69 MR

[44] **G M Ziegler**, *Lectures on polytopes*, Graduate Texts in Mathematics 152, Springer (1995) MR

BRUNO DULAR bruno.dular@uni.lu
Department of Mathematics, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Received: August 6, 2024 Revised: March 17, 2025

msp

Algebraic & Geometric Topology **26**:3 (2026) 1077–1093
DOI: 10.2140/agt.2026.26.1077
Published: April 1, 2026

Cusp-transitive 4-manifolds with every cusp section

JACOPO GUOYI CHEN AND EDOARDO RIZZI

We realize every closed flat 3-manifold as a cusp section of a complete, finite-volume hyperbolic 4-manifold whose symmetry group acts transitively on the set of cusps. Moreover, for every such 3-manifold, a dense subset of its flat metrics can be realized as cusp sections of a cusp-transitive 4-manifold. Finally, we prove that there are many 4-manifolds with pairwise isometric cusps for any given cusp type.

1 Introduction

This work is a continuation of [22], where Rizzi constructed some new cusp-transitive hyperbolic 4-manifolds. We recall in the following the general settings and results.

We say that a complete, finite-volume hyperbolic manifold is *cusp transitive* if the action of its isometry group is transitive on its cusps. This condition is trivially satisfied by manifolds with a single cusp (also called 1-*cusped* manifolds); examples are known in dimension 2, 3 (infinitely many with bounded volume, in fact) and 4: see Kolpakov and Martelli [12]. For $n > 4$, the existence of 1-cusped hyperbolic n-manifolds is unknown, and in general there are no explicit examples of cusp-transitive manifolds in the literature. However, it is easy to obtain examples of the latter in dimension up to 8 by coloring some well-known right-angled polytopes.

The *type* of a cusp of a hyperbolic manifold is the diffeomorphism class of its section, which is a closed flat hypersurface. In this article we will look at the 4-dimensional case; the possible cusp types are closed flat 3-manifolds. Following Conway and Rossetti [4], there are ten such manifolds; six of them are orientable:

- E_1, the 3-torus;
- E_2, the $\frac{1}{2}$-twist manifold;
- E_3, the $\frac{1}{3}$-twist manifold;
- E_4, the $\frac{1}{4}$-twist manifold;
- E_5, the $\frac{1}{6}$-twist manifold;
- E_6, the Hantzsche–Wendt manifold.

The other four are nonorientable: we will call them B_1, B_2, B_3 and B_4 (respectively denoted by $+a1$, $-a1$, $+a2$ and $-a2$ in [4]). These ten manifolds can be constructed as in Figures 4 and 7 by gluing facets of polyhedral fundamental domains; the latter can be obtained, along with the associated gluing maps, by inspecting [4, Table 12].

MSC2020: 57M50.

In the case of 1-cusped orientable 4-manifolds, the cusp types E_1 (Kolpakov and Martelli [12]) and E_2 (Kolpakov and Slavich [13]) can be realized, while E_3 and E_5 cannot (Long and Reid [15]). More generally, the types E_6 (Ferrari, Kolpakov and Slavich [8]) and E_4 (Rizzi [22]) can be realized by cusp-transitive 4-manifolds. The construction in the latter paper also realizes E_1, E_2 and E_6 as cusp types. As for the nonorientable case, there exist 1-cusped 4-manifolds with cusp type B_1 (Kolpakov and Slavich [14]) and B_2 (Ratcliffe and Tschantz [21]).

We improve on the technique of [22], by requiring a less explicit construction, and we prove that the remaining cusp types can also be realized; indeed, our construction actually realizes all the cusp types. Specifically we show:

Theorem 1.1 *For each closed flat 3-manifold N there exists a cusp-transitive hyperbolic 4-manifold M with cusps of type N.*

By Remark 2.4, if N is orientable, we can also take M to be orientable. The proof of this theorem is split into two parts: in Section 4 we discuss the case of the manifolds E_1, E_2, E_4, E_6, B_1, B_2, B_3 and B_4, while Section 5 is dedicated to E_3 and E_5. The 4-manifolds constructed in each section are all commensurable to each other, but they are arithmetic in the first case and nonarithmetic in the second. Hence, they fall into two commensurability classes.

Using an argument inspired by a paper of Nimershiem [19], this result is strengthened in Section 6 as follows:

Theorem 6.1 *For every closed flat 3-manifold N, the set of flat metrics on N which can be realized as cusp sections of a cusp-transitive 4-manifold is dense in the space of all flat metrics of N.*

Moreover, with the same technique, we show an analogous result in dimension 3 (where the possible cusp types are the torus and the Klein bottle).

Finally, in Section 7, a variant of our method, combined with some arguments of Burger, Gelander, Lubotzky and Mozes [2], and Belolipetsky, Gelander, Lubotzky and Shalev [1], enables the construction of a lot of manifolds with pairwise isometric cusps:

Theorem 7.1 *For every closed flat 3-manifold N, there exists a positive constant c such that, for sufficiently large $V > 0$, there exist at least V^{cV} complete hyperbolic 4-manifolds with pairwise isometric cusps of type N and volume $\leq V$.*

By [2], there are $V^{k(n)V}$ complete hyperbolic n-manifolds without boundary of volume $\leq V$, where $k(n)$ depends only on the dimension n.

The construction used in this paper is very similar to the one of Rizzi [22]. Our manifolds are built by applying Davis' *basic construction* [5] to a so-called *developable reflectofold*, in turn obtained by gluing some copies of a hyperbolic Coxeter polytope with the following two properties:

(a) it has exactly one ideal vertex;

(b) if a bounded facet and an unbounded facet intersect, then their dihedral angle is an even submultiple of π.

We have found two hyperbolic Coxeter n-polytopes with $n \geq 4$ satisfying (a) and (b): a 4-polytope P among Im Hof's polytopes associated to Napier cycles [11] (see also Felikson and Tumarkin's web page [7]), and another 4-polytope V with 8 facets, which as far as we know does not appear in the literature. Hence, we have not been able to build manifolds of dimension greater than 4 using these techniques.

The polytope V allows us to answer Questions 1.2 and 1.3 from [22] (reported here) in the affirmative:

- Does there exist a finite-volume hyperbolic Coxeter polytope of dimension $n \geq 4$ satisfying (a) and (b)? If the dimension is $n = 4$ we require that the link of the ideal vertex is neither a parallelepiped nor a prism over a $(2, 4, 4)$-triangle.
- Does there exist a cusp-transitive hyperbolic 4-manifold with cusps of type E_3 or E_5?

Indeed, the link of the ideal vertex of V is a prism K over an equilateral triangle; the fact that E_3 and E_5 are tessellated by copies of K is crucial to the construction of the corresponding cusp-transitive manifolds. The answer to the first question naturally leads to the following.

Question 1.2 Does there exist a finite-volume hyperbolic Coxeter polytope of dimension $n \geq 5$ satisfying (a) and (b)?

The paper is organized as follows. In Section 2 we describe how to obtain a cusp-transitive manifold from a 1-cusped developable reflectofold. In Section 3 we introduce the polytope P and assemble 16 copies of it to create a bigger polytope Q. In Section 4 we show how to construct reflectofolds by gluing copies of Q according to some cubical tessellations, and consequently prove the first case of Theorem 1.1. In Section 5 we introduce the polytope V and construct reflectofolds by gluing copies of it according to certain tessellations in prisms, proving the second and final case of Theorem 1.1. In Section 6, we prove Theorem 6.1 and its 3-dimensional counterpart. Lastly, in Section 7, we prove Theorem 7.1.

2 Reflectofolds

Let N be a closed flat 3-manifold. In this section we will see that, in order to show the existence of a cusp-transitive manifold with cusp type N, it suffices to prove that there exists a 1-cusped *developable reflectofold* with cusp type N. We give some definitions in the following (compare [6, Section 5.1; 22, Section 2]).

Definition 2.1 We say that a hyperbolic manifold R with cornered boundary is a *reflectofold* if R is locally a hyperbolic Coxeter polytope.

Let R be a reflectofold. Since each local model P of R is stratified into k-dimensional faces, for $k = 0, \dots, n$, there is a naturally induced stratification of R into maximal, connected, totally geodesic submanifolds with boundary, called *k-faces*. We will call *facets* and *corners* the $(n-1)$-faces and $(n-2)$-faces of R, respectively. We can define the *dihedral angle* of a corner as the dihedral angle of the corresponding ridge of a local model.

Definition 2.2 A reflectofold is *developable* if it has the following properties:

(EF) *Embedded faces*: For each corner C there are two distinct facets F and F' such that $C \subset F \cap F'$.

(AC) *Angle consistency*: If two distinct facets F and F' intersect, then the dihedral angles of all the corners in $F \cap F'$ coincide.

Proposition 2.3 *Let R be a 1-cusped developable reflectofold with complete cusp section, and built by gluing finitely many finite-volume polytopes. Then, there exists a cusp-transitive manifold M having cusps isometric to that of R.*

Proof The desired manifold M can be constructed as in [22, Section 2] by gluing finitely many copies of R following Davis' basic construction [5, Chapter 11]. Since R has finite volume, M will also have finite volume. Note that, unlike [22], we do not require reflectofolds to be complete; however, M will be complete since it is obtained by gluing polytopes, and has complete cusp sections (see [20, Theorem 11.1.6]). □

Remark 2.4 If the cusp section of the cusp-transitive manifold M thus constructed is orientable, we can also construct an orientable cusp-transitive manifold with the same cusp section, namely the orientable double cover of M, as in [22, Corollary 2.4].

3 Some polytopes

In this section we will introduce and construct the polytopes that we are going to use in Section 4.

Following Vinberg's paper [23, Introduction], we define a *hyperbolic Coxeter polytope* as a convex polytope $P \subset \mathbb{H}^n$ with finitely many facets, with dihedral angles of the form π/m for an integer $m \geq 2$. We call *facets* and *ridges* the faces of P of codimension 1 and 2, respectively.

We associate to such a polytope its *Coxeter diagram*, a labeled graph defined as follows. We refer to the facets by progressive numbers $1, \dots, k$. The graph has a vertex for each facet, and if two facets i and j intersect with dihedral angle π/m_{ij}, then the edge joining the corresponding vertices has label m_{ij}. If $m_{ij} = 2$, the edge is conventionally omitted. If the supporting hyperplanes of i and j are tangent at infinity, we label the corresponding edge with $m_{ij} = \infty$. Finally, a dashed edge is drawn between any two vertices corresponding to ultraparallel facets i and j. The edge can be labeled by $\cosh(d_{ij})$, where d_{ij} is the distance between the supporting hyperplanes.

The *Gram matrix* G of the polytope is a real symmetric $k \times k$ matrix, defined as follows:

$$G_{ij} := \begin{cases} 1 & \text{if } i = j; \\ -\cos(\pi/m_{ij}) & \text{if the facets } i \text{ and } j \text{ intersect;} \\ -1 & \text{if the facets } i \text{ and } j \text{ are parallel;} \\ -\cosh(d_{ij}) & \text{if the facets } i \text{ and } j \text{ are ultraparallel.} \end{cases} \tag{1}$$

3A The polytope P

We now introduce a Coxeter polytope from [11].

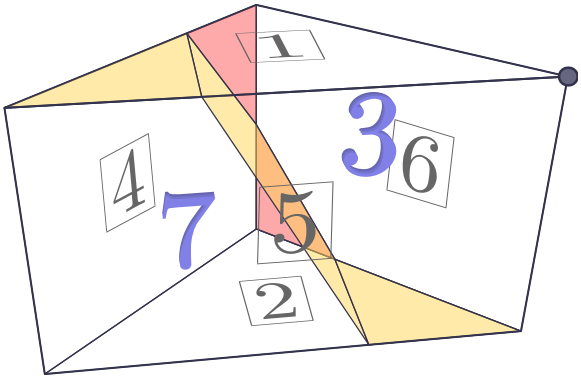
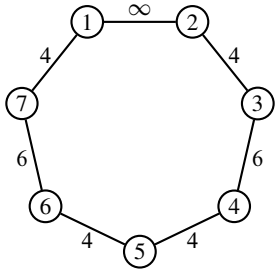


Figure 1: The projection of P onto the link L of its ideal vertex. The five faces of L and the two interior regions are labeled with the corresponding facets of P: five noncompact (**1**, **2**, **4**, **5**, **6**) and two compact (**3**, **7**).

Consider the following Coxeter diagram D_1:



This graph is the Coxeter diagram of a finite-volume arithmetic hyperbolic Coxeter 4-polytope P, introduced in [11], which satisfies (a) and (b). Indeed, by [23, Chapter 1, §3], the ideal vertices correspond to the maximal affine subdiagrams of the Coxeter diagram, and in D_1 we have exactly one of this kind (see [23, Table 2]), spanned by the vertices 1, 2, 4, 5, 6. The subdiagram spanned by these vertices is the Coxeter diagram of a Euclidean right prism over a triangle with interior angles $\pi/2$, $\pi/4$, $\pi/4$, which is the horospherical link of the ideal vertex of P. Arithmeticity can be verified using the program CoxIter [9; 10].

3B Visualizing polytopes with one ideal vertex

In order to better understand the geometry of the 4-dimensional polytope we just defined, it may prove useful to *project* the compact boundary facets onto the horospherical link of the ideal vertex as follows.

Place the polytope P in the half-space model of $\mathbb{H}^4$ with the ideal vertex at infinity, so that the noncompact facets of P become vertical, and let $\pi : \mathbb{H}^4 \to \mathbb{R}^3$ be the orthogonal projection onto a horizontal hyperplane, such as a horosphere or the ideal boundary. The image of P under π is the three-dimensional link L of the ideal vertex, and the noncompact facets are mapped to its boundary.

We can compute the face lattice of P by enumerating all spherical subdiagrams of D, and therefore determine the shape of the two regions associated to the facets **3** and **7** (Figure 1) by applying the following result.

Proposition 3.1 *Let Z be an n-polytope in $\mathbb{H}^n$ with one ideal vertex v. Let Λ be the link of v. The images of the compact facets of Z under the projection onto a horosphere centered at v partition Λ into a polyhedral complex, whose face lattice is combinatorially isomorphic to the lattice of compact faces of Z.*

Proof We consider the half-space model where v is at infinity. We say that a k-subspace of the half-space model is *vertical* if its ideal boundary contains v.

Let F be a compact k-face of Z. Then, in the half-space model, F is supported on a k-hemisphere H. Let Σ be the unique vertical $(k+1)$-space containing H. The face F is bounded by the supporting $(k-1)$-spaces of its facets $F_1, \dots, F_m$. Each F_i lies on a unique vertical k-space S_i. As such, F is the intersection of H and some hyperbolic half-$(k+1)$-spaces bounded by the S_i and contained in Σ. Let P_F be the Euclidean k-polytope obtained by intersecting the projections of these half-spaces on the horosphere; then we have $F = (P_F \times (0, +\infty)) \cap H$. Hence, the projection is a homeomorphism between F and P_F, which preserves facets. Since Z has only one ideal vertex, every compact face of Z is contained in a compact facet. Hence, by induction on the codimension of faces, the projection preserves the face lattice of the compact boundary of Z. □

Dihedral angles between facets of P are defined along ridges (codimension-2 faces), which appear in Figure 1 either as 2-faces (if the angle involves a compact facet) or as edges of L (if the angle is between two noncompact facets). We mark the former by coloring certain 2-faces of the projection, with the following rule. Every 2-face corresponds to a ridge between either two compact facets or a compact facet and a noncompact one. In the first case, we assign the dihedral angle to the 2-face, while in the second case, we assign the doubled dihedral angle. We use yellow for $\pi/2$, red for $\pi/3$, and we leave the 2-face transparent for π.

This rule keeps track of the fact that, if we glue two copies of P along a noncompact facet F, the angle along any ridge it shares with a compact facet gets doubled. For instance, the dihedral angle between facets **1** and **7** is $\pi/4$; when doubling P along **1**, the angle becomes $\pi/2$, so we color in yellow the corresponding triangle in Figure 1.

Moreover, after doubling, some ridges end up with an angle of π, meaning that the compact facet is coplanar with its mirrored copy. Leaving the 2-face transparent has the advantage of visually representing when compact facets merge together in an arbitrary gluing of copies of P.

3C The construction of the polytope Q

In this section we will take 16 copies of P and we will glue them in order to form a bigger polytope Q.

Consider the subdiagram of D spanned by the vertices $1, 5, 6$. This induces a subgroup $G \simeq \mathbb{Z}_2 \times D_4$, of cardinality 16, which is the stabilizer of a noncompact edge of P. This edge projects to the marked vertex in Figure 1.

We can define a new polytope Q by taking the orbit of P under G. Equivalently, we glue 16 copies of P along their noncompact facets **1**, **5**, **6** using the identity map. By construction, Q has exactly one ideal vertex, whose link is a right prism C with a square base, obtained by placing 16 copies of L around the marked vertex (Figure 2).

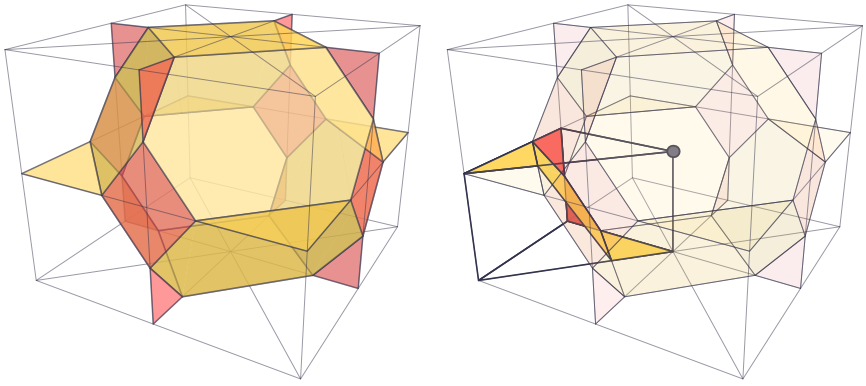


Figure 2: On the left, the projection of Q onto the link C. The truncated octahedron corresponds to a facet made of 16 copies of the facet **3** of P. The other 8 regions appear when copies of the facet **7** merge together two at a time. On the right, we emphasize one copy of L inside C.

The facets **3** merge together into a single facet of Q (which we will also call **3**), which projects to an irregular truncated octahedron in C, with 8 hexagonal and 6 quadrilateral facets. The latter correspond to ridges of Q where the facet **3** meets the noncompact facets: the dihedral angles are $\pi/6$ for the vertical quadrilaterals and $\pi/4$ for the horizontal ones.

The height, width and depth of C are in a ratio of $\cos(\pi/4) : \cos(\pi/6) : \cos(\pi/6) = \sqrt{2} : \sqrt{3} : \sqrt{3}$. Indeed, suppose that C is the cuboid $[-x_1, x_1]\times[-x_2, x_2]\times[-x_3, x_3]$. Then, in the conformal half-space model, each noncompact facet of Q is contained in the product of a facet of C and $(0, +\infty)$. The facet **3** is supported on a hemisphere, which is centered at $(0, 0, 0, 0)$ by symmetry. It is not hard to see that, if this hemisphere has radius r, then the acute angles with the supporting hyperplanes of the noncompact facets of Q are $\theta_i := \arccos(x_i/r)$. Hence, the cosines of the θ_i are in the same ratios as the x_i.

Because of this, all symmetries of C must preserve or exchange the two horizontal faces. There are 16 such symmetries, and they are generated by reflections in the facets $\mathbf{1}, \mathbf{5}, \mathbf{6}$ of L, so they extend to symmetries of Q and they form a group isomorphic to G. This group can also be defined *a priori* as the group of symmetries of a *combinatorial* cube preserving or exchanging a certain pair of opposite facets.

Remark 3.2 The polytope C naturally tessellates $\mathbb{R}^3$ by translations. This gives a tessellation into truncated octahedra in the following way: some are centered and contained in the copies of C, as in Figure 2, while the others are centered at the vertices of the tessellation; one-eighth of a truncated octahedron can be seen near each vertex of C in Figure 2. The two types of truncated octahedra are actually congruent, because of the symmetry of D that exchanges 3 and 7. Indeed, passing to the dual tessellation by copies of C exchanges the two types of truncated octahedra.

4 Layered tessellations of 3-manifolds

In this section we will prove that if a flat 3-manifold N admits a tessellation in cubes with some properties, then there is a cusp-transitive hyperbolic 4-manifold with cusp type N.

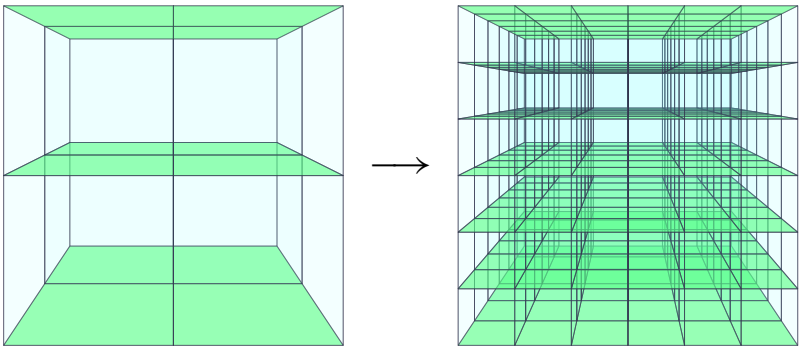


Figure 3: Subdividing a layered tessellation. Special faces are colored green.

Definition 4.1 (layered tessellation) Let T be a tessellation of a flat 3-manifold N into Euclidean cubes, with some chosen *special* 2-cells. We say that T is *layered* if each cube has two opposite special facets, and whenever two cubes C_1, C_2 share a facet F, the reflection through F sends the special facets of C_1 to those of C_2, and vice versa.

The above condition causes the special facets to fall into several embedded geodesic surfaces tessellated by squares, in a discrete analog of a foliation.

Remark 4.2 We will only make use of the combinatorial properties of layered tessellations; hence, we may define generalizations, such as layered tessellations by rectangular cuboids, in the same way.

Remark 4.3 A layered tessellation can be *subdivided* by replacing every cube with a block of $n\times n\times n$ cubes, $n \geq 1$, in which we mark as special all facets parallel to the two original special facets (see Figure 3).

Definition 4.4 We say that a layered tessellation is *proper* if

- every cube is distinct from its six neighbors, which are pairwise distinct;
- every vertex is distinct from its six neighbors, which are pairwise distinct.

Lemma 4.5 *Every layered tessellation can be subdivided into a proper one.*

Proof First, we realize each cube of the tessellation as a cube with side length 1. This gives a flat Riemannian metric on a compact 3-manifold, which has an injectivity radius $r > 0$. We then subdivide the tessellation as in Remark 4.3, in such a way that the side length of the little cubes is less than r. The resulting tessellation is proper because, for every cube, the centers of it and its neighbors are contained in an embedded ball, and a similar argument applies to the vertices. □

Theorem 4.6 *Let N be a closed 3-manifold that admits a layered tessellation. Then there exists an arithmetic cusp-transitive hyperbolic 4-manifold with cusp type N.*

Proof By Proposition 2.3, it suffices to prove that there exists a 1-cusped developable reflectofold with cusp type N, obtained by gluing copies of Q.

By Lemma 4.5, we may assume that the tessellation is proper.

We have that N is obtained by gluing some copies of C along their facets, where the gluing maps are restrictions of isometries of G: this gluing is induced by the layered tessellation, where the special facets correspond to the horizontal facets of C. Since there is a natural correspondence between facets of C and noncompact facets of Q, and every isometry in G extends to Q accordingly, we can glue copies of Q in the same pattern. This gives a 1-cusped reflectofold with cusp type N.

It remains to prove that N is developable. The facets of N are hyperbolic truncated octahedra, which arise from the merging of 16 facets **3** or **7**; we will call them *of types* **3** and **7** respectively. The former are centered at the centers of the cubes, while the latter are centered at the vertices of the tessellation.

If a facet of type **3** were adjacent to itself, it would be so along a quadrilateral corner, and so it would also occupy a neighboring cube. However, the two cubes must be distinct by properness of the tessellation. A similar argument involving vertices works for facets of type **7**.

As for angle consistency, if two facets of different type intersect, they do so in a hexagonal corner with a dihedral angle of $\pi/2$. If a facet intersects another of the same type, say **3**, it must do so at a single quadrilateral corner, since the neighbors of a given cube are pairwise distinct by properness; angle consistency follows trivially. A dual argument deals with the case of two facets of type **7**.

The 4-manifold thus constructed is arithmetic since it covers the arithmetic orbifold P. □

Corollary 4.7 *Let N be one of the manifolds E_1, E_2, E_4, E_6, B_1, B_2, B_3, B_4. Then there exists a cusp-transitive arithmetic hyperbolic* 4*-manifold with cusp type N.*

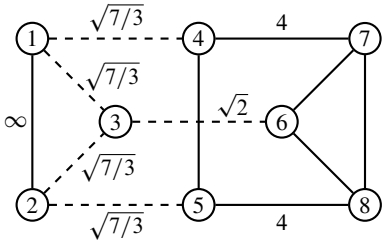
Proof As shown in Figure 4, the manifold N admits a layered tessellation, so the result follows from Theorem 4.6. □

5 The remaining manifolds

In this section we will construct cusp-transitive manifolds having the $\frac{1}{3}$-twist and $\frac{1}{6}$-twist manifolds as cusp sections.

5A The polytope V

Let D_2 be the following Coxeter diagram:



The diagram D_2 is obviously connected, and we can check that its Gram matrix has signature $(4, 1, 3)$ with nonpositive off-diagonal entries; hence, by Vinberg's theorem [23, Theorem 2.1], it defines a hyperbolic

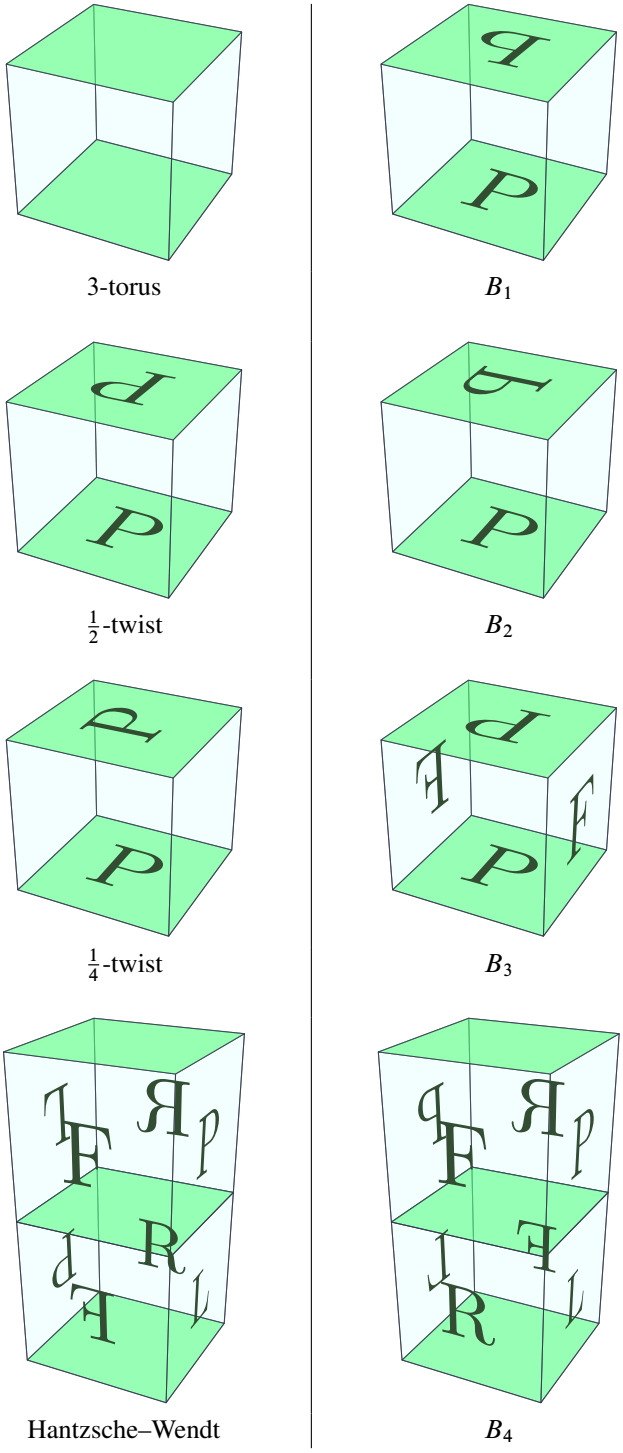


Figure 4: Layered tessellations for eight closed flat 3-manifolds. Labeled facets are glued according to their labels, while unlabeled facets are glued to the opposite facets by translation. The fundamental domains and gluing maps can be deduced from the algebraic descriptions of [4, Table 12].

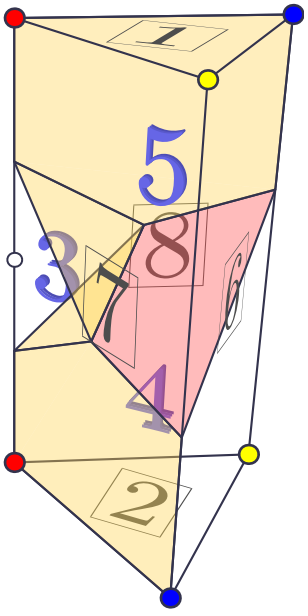


Figure 5: The projection of V onto the link K of its ideal vertex, with labels indicating the five noncompact facets (**1**, **2**, **6**, **7**, **8**) and three compact facets (**3**, **4**, **5**).

Coxeter 4-polytope V. Using CoxIter [9; 10], we find that V has finite volume and is nonarithmetic. Moreover, like the polytope P, it satisfies (a) and (b): there is exactly one ideal vertex, corresponding to the unique maximal affine subdiagram of D_2 [23, Chapter 1, §3], which is spanned by 1, 2, 6, 7, 8. Its link K is a Euclidean right prism over an equilateral triangle (Figure 5, obtained by using Proposition 3.1, and colored with the rule of Section 3B).

From the diagram D_2, we can see that the polytope V has a symmetry σ that exchanges the facets **1** and **2**, **4** and **5**, **7** and **8**. This induces a half-turn rotation of K swapping each vertex with the other one of the same color in Figure 5. This rotation is the only nontrivial color-preserving combinatorial automorphism of K.

5B Marked tessellations of 3-manifolds

Mirroring the previous sections, we will prove that if a flat 3-manifold N admits a tessellation in prisms over equilateral triangles with some properties, then there is a cusp-transitive hyperbolic 4-manifold with cusp type N.

Definition 5.1 (marked tessellation) Let T be a tessellation of a flat 3-manifold N into right prisms over equilateral triangles, such that the vertices are colored red, yellow or blue. We say that T is *marked* if the vertices of each prism are colored as in Figure 5 (up to combinatorial isomorphism), and whenever two prisms share a facet F, they are symmetrical with respect to F, including their vertex colorings.

Remark 5.2 Given a marked tessellation T and two natural numbers a, b, we can *subdivide* each prism as follows. First, note that an equilateral triangle can be subdivided into a^2 triangles (whose sides are a

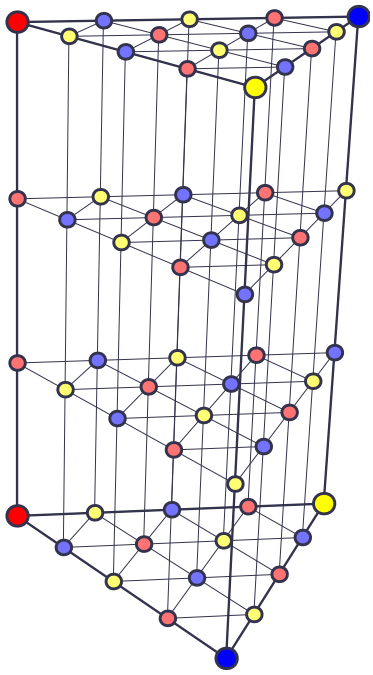


Figure 6: Subdividing a marked tessellation ($a = 4$, $b = 3$). The pattern on the bottom face of the prism works whenever $a - 1$ is a multiple of 3.

times smaller), by cutting it with three sets of equally spaced lines parallel to the sides; similarly, a prism can be subdivided into a^2 thin prisms. Each of those can then be cut, parallel to the bases, into b equal prisms. The new tessellation T' can be marked provided that b is odd and that $a - 1$ is a multiple of 3 (see Figure 6 for the case $a = 4$, $b = 3$). Indeed, for each prism t of T, we now have $b + 1$ layers of vertices of T'. It is not hard to see that, whenever $a \equiv 1 \pmod 3$, there is a unique way to 3-color the bottom layer consistently with the lower three vertices of t. Each subsequent layer will be the same as the one below it, but with yellow and blue swapped. If b is odd, then the top and bottom layers are the same with yellow and blue swapped, ensuring consistency with the coloring of t.

Remark 5.3 When gluing copies of V along noncompact facets, the compact facets merge into two kinds of new facets, which we can visualize with the help of the projection in Figure 5. The first kind arises from 6 copies of the facet **3** around the white dot, while the second comes from 12 copies of the facet **4** or **5** around the corresponding yellow dot. Both kinds of facets are bounded.

Theorem 5.4 *Let N be a closed 3-manifold that admits a marked tessellation. Then there exists a nonarithmetic cusp-transitive hyperbolic 4-manifold with cusp type N.*

Proof The proof is similar in many aspects to that of Theorem 4.6, so we will only go over the main points. Again, by Proposition 2.3, it suffices to prove that there exists a 1-cusped developable reflectofold with cusp type N, obtained by gluing copies of V.

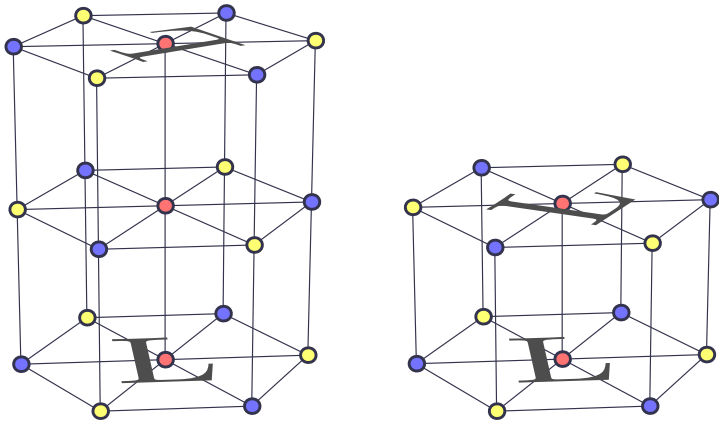


Figure 7: Marked tessellations for E_3 and E_5. Labeled facets are glued according to their labels, while unlabeled facets are glued to the opposite facets by translation. The fundamental domains and gluing maps can be deduced from the algebraic descriptions of [4, Table 12].

A marked tessellation of N induces a way to glue copies of V along their noncompact facets; a key fact is that every color-preserving combinatorial automorphism of K is induced by a symmetry of V. This gives a 1-cusped reflectofold with cusp type N.

In order to ensure (EF) and (AC), it suffices that each facet of the reflectofold, together with its adjacent facets, is inside an embedded ball. Indeed, this ensures that every facet is embedded and that the intersection of two adjacent facets is a single corner. Since the facets are bounded by Remark 5.3, this can be done by an argument involving the injectivity radius and a sufficiently fine subdivision of the marked tessellation; the reflectofold constructed from the subdivided tessellation is developable.

The 4-manifold thus constructed is nonarithmetic since it covers the nonarithmetic orbifold $V/\langle\sigma\rangle$, where σ is the order-2 isometry of V defined in Section 5A. □

Corollary 5.5 *Let N be one of the manifolds E_3, E_5. Then there exists a cusp-transitive hyperbolic 4-manifold with cusp type N.*

Proof The manifolds E_3 and E_5 have marked tessellations, as shown in Figure 7; the result follows from Theorem 5.4. □

6 Density

In this section we will strengthen the previous results by investigating the possible Euclidean structures that can be realized by the cusp sections of a cusp-transitive 4-manifold, and we show how the same techniques can be used in the 3-dimensional case. We have the following result:

Theorem 6.1 *For every closed flat 3-manifold N, the set of flat metrics on N which can be realized as cusp sections of a cusp-transitive 4-manifold is dense in the space of all flat metrics of N.*

Proof The argument is inspired by the proof of [19, Theorem 2]. We divide the proof into two cases: the manifolds E_1, E_2, E_4, E_6, B_1, B_2, B_3, B_4 (tessellated by cubes) and the manifolds E_3, E_5 (tessellated by triangular prisms).

We start with the first case. Any flat metric on N is induced by a subgroup $\Gamma < \mathrm{Isom}(\mathbb{R}^3)$. Generators for Γ are found in [19, Table 1] in the form (A, t_u), where $A \in \mathrm{O}(3)$ and t_u is a translation by the vector u, denoting the map $v \mapsto u + Av$.

The group Γ can be conjugated as in [19, pages 128–129] so that all the matrices A_i are certain fixed signed permutation matrices, which preserve the x axis, and preserve or exchange the y and z axes. In this *normalized form*, the translation vectors have some zero entries, while the k-uple of the other *free* entries can take any value in an open set of $\mathbb{R}^k$, containing $(\mathbb{R} \setminus \{0\})^k$. A dense subset of these forms can be obtained by considering only vectors of the form

$$v_i := \big(\sqrt{2}a_i, \sqrt{3}b_i, \sqrt{3}c_i\big),$$

where a_i, b_i, c_i are in $\mathbb{Q} \setminus \{0\}$ or $\{0\}$, depending on whether the corresponding entry of v_i is free or not. Let Γ' be a group with parameters in this dense subset; we shall prove that the metric resulting from Γ' can be realized by a layered tessellation by copies of C (see Remark 4.2).

The group Γ' preserves a lattice of the form

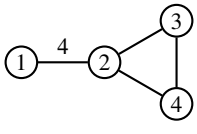
$$\left\{\frac{1}{d}\big(\sqrt{2}m, \sqrt{3}n, \sqrt{3}p\big) \mid m, n, p \in \mathbb{Z}\right\},$$

where d is a common denominator of all the a_i, b_i, c_i. Furthermore, it preserves a layered tessellation of $\mathbb{R}^3$ by copies of C — having side lengths $\sqrt{2}/d$ and $\sqrt{3}/d$ — with the special facets orthogonal to the x axis; this is because the x axis is preserved by the A_i. This tessellation descends to the quotient $\mathbb{R}^3/\Gamma'$. The result follows as in Theorem 4.6.

As for the second case, we refer to [4, Section 4]. The space of flat metrics on N has two parameters: in the hexagonal prism fundamental domains of Figure 7, they can be taken as the side length and height of the prisms. When subdividing a marked tessellation in the proof of Theorem 5.4, we can choose two parameters a and b provided that b is odd, $a-1$ is a multiple of 3, and they are sufficiently large. The effect of these parameters is to multiply the side length and height of the fundamental domains by a and b respectively. Since we can realize the tessellation with arbitrarily small copies of K, by choosing a and b in an appropriate way, we can approximate any values of the two parameters of the metric of N. The result follows as in Theorem 5.4. □

6A Dimension 3

We consider the following diagram D_3, found in [3, Appendice, Rang 4]:



The corresponding Coxeter polytope Y is a finite-volume hyperbolic tetrahedron with one ideal vertex, corresponding to the subdiagram spanned by 2, 3, 4; its link is an equilateral triangle. The polytope has the properties (a) and (b).

Theorem 6.2 *For every closed flat 2-manifold S (i.e., the torus and the Klein bottle), the set of flat metrics on S which can be realized as cusp sections of a cusp-transitive 3-manifold is dense in the space of all flat metrics of S.*

Proof We begin by gluing six copies of Y around the edge between the facets **3** and **4**. The resulting polytope Z has one ideal vertex, whose link is a regular hexagon, and one compact facet, which meets the other six facets at an angle of $\pi/4$.

If S has a tessellation by regular hexagons, such that every hexagon is embedded, then we can glue copies of Z along their noncompact facets in the same pattern. The resulting manifold with corners R is a developable reflectofold: its dihedral angles are all equal to $2\cdot\pi/4=\pi/2$, also implying (AC), and its facets are embedded hyperbolic hexagons (EF). Moreover, the section of its unique cusp is isometric to S up to global rescaling. As usual, by Proposition 2.3, we can construct a cusp-transitive manifold which covers R, with cusps isometric to S.

It remains to show that, up to arbitrarily small perturbations, every flat metric on the torus or the Klein bottle admits a tessellation by regular hexagons.

First, consider a parallelogram fundamental domain D_T for the torus, and overlay it onto a tessellation of small regular hexagons. We can perturb D_T by moving three vertices to the nearest hexagon center, and the fourth in such a way as to make a parallelogram; the fourth vertex will also fall into the center of a hexagon. This gives our desired tessellation.

As for the Klein bottle, the generic fundamental domain D_K is a rectangle, with two opposite sides s_1, s_2 identified with a twist. We overlay D_K onto a tessellation of small regular hexagons with some sides parallel to s_1, s_2. We can perturb D_K by translation or scaling along either axis, obtaining a rectangle with all vertices at the centers of hexagons. Note that the tessellation is symmetrical with respect to the line joining the midpoints of the perturbed s_1 and s_2. Hence, we have a tessellation of the Klein bottle. □

7 Lots of manifolds

In this section we prove the following theorem.

Theorem 7.1 *For every closed flat 3-manifold N, there exists a positive constant c such that, for sufficiently large $V>0$, there exist at least V^{cV} complete hyperbolic 4-manifolds with pairwise isometric cusps of type N and volume $\le V$.*

Proof Let R be a reflectofold constructed as in the proof of Theorem 4.6 or 5.4 (according to N). Let D be the Coxeter diagram with one vertex for each facet of R, where if two facets meet with dihedral angle of π/k (respectively do not intersect), the corresponding vertices are joined by an edge labeled k (respectively a dashed edge).

If R is constructed from a sufficiently fine subdivision of a tessellation, then the diagram D is connected. Indeed, two nonadjacent facets are connected by a dashed edge, while for any two adjacent facets F, F' there exists a third one not adjacent to them (in the projection onto the link, it can be found in a large embedded ball centered on, say, F). As a consequence, we can also assume that D has a dashed edge.

Let G be the Coxeter group associated to D. It is neither affine nor spherical, since D has a dashed edge (compare [23, Tables 1–2]). By [17, Corollary 2], G has a finite-index subgroup H with a quotient isomorphic to the free group F_2. Since H is a subgroup of a Coxeter group, it has a torsion-free subgroup H' such that $d := [G : H'] < +\infty$. The image of H' in F_2 is free of rank at least 2, so H' also has a quotient isomorphic to F_2. By [18, Theorem 2], there exists a constant $\alpha > 0$ such that F_2 has at least $\alpha r \cdot r!$ subgroups of index $\leq r$. Hence, by pulling back to H', we obtain at least $\alpha r \cdot r!$ torsion-free subgroups of G of index $\leq dr$. These correspond to manifold covers of R with degree $\leq dr$, which have pairwise isometric cusps of type N by construction. As in the proof of Proposition 2.3, we can show that these manifolds are complete.

Let v be the volume of R and let $V := vdr$. Then our manifolds have volume $\leq V$. Using Stirling's approximation, we have the following estimate (for some $k, k', c > 0$ and for r large):

$$\begin{aligned}
\log(\alpha r \cdot r!) &= \log\alpha + \log r + \log r! \\
&= \log\alpha + \log r + r\log r - r + O(\log r) \\
&\geq k\, r \log r \\
&= k' \frac{V}{vd} \log \frac{V}{vd} \\
&\geq c V \log V \\
&= \log(V^{cV}).
\end{aligned}$$

Hence, we have at least V^{cV} torsion-free subgroups of G. The associated manifolds (of volume $\leq V$) are not necessarily distinct; however, the same estimate holds on the number of isometry classes, with a smaller constant c. Indeed, if G is nonarithmetic, we conclude with the same argument of [2, "the lower bound"] using Margulis' theorem on the commensurator [16, Theorem 1, page 2], while if G is arithmetic, we conclude as in [1, Section 5.2] using the Kazhdan–Margulis theorem. □

Remark 7.2 Since we take subgroups of G which are not necessarily normal, the group G does not act on the associated covers. Hence, we do not necessarily get cusp-transitive manifolds.

Acknowledgements

We would like to thank our advisors Bruno Martelli and Stefano Riolo for their support.

References

[1] **M Belolipetsky**, **T Gelander**, **A Lubotzky**, **A Shalev**, *Counting arithmetic lattices and surfaces*, Ann. of Math. (2) 172:3 (2010) 2197–2221 MR

[2] **M Burger**, **T Gelander**, **A Lubotzky**, **S Mozes**, *Counting hyperbolic manifolds*, Geom. Funct. Anal. 12:6 (2002) 1161–1173 MR

[3] **M Chein**, *Recherche des graphes des matrices de Coxeter hyperboliques d'ordre* ≤ 10, Rev. Franąise Informat. Recherche Opérationnelle 3 (1969) 3–16 MR

[4] **J H Conway**, **J P Rossetti**, *Describing the platycosms*, preprint (2003) arXiv math/0311476

[5] **M W Davis**, *The geometry and topology of Coxeter groups*, London Mathematical Society Monographs Series 32, Princeton Univ. Press (2008) MR

[6] **M W Davis**, *Lectures on orbifolds and reflection groups*, from "Transformation groups and moduli spaces of curves" (L Ji, S-T Yau, editors), Adv. Lect. Math. (ALM) 16, International, Somerville, MA (2011) 63–93 MR

[7] **A Felikson**, **P Tumarkin**, *Hyperbolic Coxeter polytopes* Available at https://www.maths.dur.ac.uk/users/anna.felikson/Polytopes/polytopes.html

[8] **L Ferrari**, **A Kolpakov**, **L Slavich**, *Cusps of hyperbolic* 4*-manifolds and rational homology spheres*, Proc. Lond. Math. Soc. (3) 123:6 (2021) 636–648 MR

[9] **R Guglielmetti**, *CoxIter*, GitHub repository Available at https://rgugliel.github.io/CoxIter/index.html

[10] **R Guglielmetti**, *CoxIter—computing invariants of hyperbolic Coxeter groups*, LMS J. Comput. Math. 18:1 (2015) 754–773 MR

[11] **H-C Im Hof**, *Napier cycles and hyperbolic Coxeter groups*, Bull. Soc. Math. Belg. Sér. A 42:3 (1990) 523–545 MR

[12] **A Kolpakov**, **B Martelli**, *Hyperbolic four-manifolds with one cusp*, Geom. Funct. Anal. 23:6 (2013) 1903–1933 MR

[13] **A Kolpakov**, **L Slavich**, *Hyperbolic* 4*-manifolds, colourings and mutations*, Proc. Lond. Math. Soc. (3) 113:2 (2016) 163–184 MR

[14] **A Kolpakov**, **L Slavich**, *Symmetries of hyperbolic* 4*-manifolds*, Int. Math. Res. Not. 2016:9 (2016) 2677–2716 MR

[15] **D D Long**, **A W Reid**, *On the geometric boundaries of hyperbolic* 4*-manifolds*, Geom. Topol. 4 (2000) 171–178 MR

[16] **G A Margulis**, *Discrete subgroups of semisimple Lie groups*, Ergebnisse der Math. (3) 17, Springer (1991) MR

[17] **G A Margulis**, **E B Vinberg**, *Some linear groups virtually having a free quotient*, J. Lie Theory 10:1 (2000) 171–180 MR

[18] **M Newman**, *Asymptotic formulas related to free products of cyclic groups*, Math. Comp. 30:136 (1976) 838–846 MR

[19] **B E Nimershiem**, *All flat three-manifolds appear as cusps of hyperbolic four-manifolds*, Topology Appl. 90:1-3 (1998) 109–133 MR

[20] **J G Ratcliffe**, *Foundations of hyperbolic manifolds*, 3rd edition, Graduate Texts in Mathematics 149, Springer (2019) MR

[21] **J G Ratcliffe**, **S T Tschantz**, *Hyperbolic* 24*-cell* 4*-manifolds with one cusp*, Exp. Math. 32:2 (2023) 269–279 MR

[22] **E Rizzi**, *Some cusp-transitive hyperbolic* 4*-manifolds*, Geom. Dedicata 219:4 (2025) art. id. 61 MR

[23] **E B Vinberg**, *Hyperbolic reflection groups*, Uspekhi Mat. Nauk 40:1(241) (1985) 29–66, 255 MR In Russian; translated in Russian Math. Surveys 40 (1985) 31–75

JACOPO GUOYI CHEN jacopo.chen@sns.it
Scuola Normale Superiore, Pisa, Italy

EDOARDO RIZZI edoardo.rizzi@sns.it
Scuola Normale Superiore, Pisa, Italy

Received: August 12, 2024 Revised: December 10, 2024

msp

Algebraic & Geometric Topology **26**:3 (2026) 1095–1114
DOI: 10.2140/agt.2026.26.1095
Published: April 1, 2026

Finiteness conjecture for 3-manifolds obtained from handlebodies by attaching 2-handles

HIROAKI KARUO AND ZHIHAO WANG

We study a generalized Witten's finiteness conjecture for the skein modules of oriented compact 3-manifolds with boundary. We formulate an equivalent version of the generalized finiteness conjecture using handlebodies and 2-handles, and prove the conjecture for some classes with the handlebodies of genus 2 and 3 using the equivalent version.

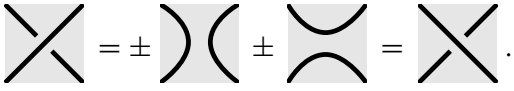
1 Introduction

Let $\mathcal{R}$ be a commutative unital ring with a distinguished invertible element q and M be an oriented 3-manifold with (possibly empty) boundary ∂M. The *skein module* of M, denoted by $\mathscr{S}_q(M)$, is introduced as the $\mathcal{R}$-module generated by isotopy classes of framed links (including the empty set) in M subject to the relations

where, in each relation, the local pictures are the intersection of framed links and an open 3-ball in M and the framed links are the same except where shown.

For an oriented surface Σ, when we replace M with $\Sigma\times[0,1]$, it allows multiplication on the skein module by stacking with respect to $[0,1]$. With this multiplication, the skein module of $\Sigma\times[0,1]$ is called the *skein algebra* of Σ, denoted by $\mathscr{S}_q(\Sigma)$.

When we substitute ±1 to q, we have



This implies skein algebras are commutative at $q=\pm1$. In addition, skein modules are commutative algebras by taking the product of two framed links as a disjoint union of them. In particular, in the case when $\mathcal{R}=\mathbb{C}$, we have $\mathscr{S}_{\pm1}(M)/\sqrt{0}$ is isomorphic to the coordinate ring of the $\mathrm{SL}_2\mathbb{C}$-character variety of $\pi_1(M)$, where $\sqrt{0}$ is the nilradical [3; 4; 19].

For any 3-manifold M, the skein algebra $\mathscr{S}_q(\partial M)$ has an obvious action on $\mathscr{S}_q(M)$. We will use "$\cdot$" to denote this action. More precisely, we identify a small closed regular neighborhood U of ∂M with $\partial M\times[0,1]$ such that $\partial M\times\{1\}$ is identified with ∂M. Then for any skeins α in $\partial M\times[0,1]$ and β in M,

MSC2020: primary 57K31; secondary 57K20, 57R65.

we first push β away from U, then define $\alpha \cdot \beta = \alpha \cup \beta$. With this action, $\mathscr{S}_q(M)$ can be regarded as a $\mathscr{S}_q(\partial M)$-module.

Skein modules have various features according to the ground ring $\mathcal{R}$. For instance, when $\mathcal{R} = \mathbb{Q}(q)$ the field of rational functions in q over $\mathbb{Q}$, it was conjectured that skein modules of closed 3-manifolds are finite dimensional; this was known as Witten's finiteness conjecture. After some partial results [6; 7; 8; 10; 11; 16], the conjecture was solved affirmatively in [9].

Theorem 1.1 (Witten's finiteness conjecture [9]) *For any oriented closed 3-manifold M, the skein module of M has finite rank over $\mathbb{Q}(q)$.*

As a generalization of this conjecture, the following conjecture was formulated by Detcherry [7].

Conjecture 1.2 (finiteness conjecture for 3-manifolds with boundaries) For any oriented compact 3-manifold M with (possibly empty) boundary ∂M, the skein module $\mathscr{S}_q(M)$ has a finite subset which spans $\mathscr{S}_q(M)$ over the $\mathbb{Q}(q)$-algebra $\mathscr{S}_q(\partial M)$, where if $\partial M = \varnothing$ then we regard $\mathscr{S}_q(\partial M) = \mathbb{Q}(q)$.

There is a stronger version of Conjecture 1.2 proposed by Detcherry [7].

Except for closed 3-manifolds, it is known that Conjecture 1.2 holds for

- $F \times [0,1]$ with a compact oriented surface F (by definition),
- Seifert fibered manifolds with orientable base and nonempty boundary [2],
- the complement of a torus knot, a 2-bridge link or a $(-2,3,2n+1)$-pretzel knot in S^3 [12; 13; 14; 15],

and their proofs are based on diagrammatic techniques.

For a general 3-manifold M with nonempty boundary, Conjecture 1.2 is more complicated than Witten's finiteness conjecture since we have to treat infinitely many skeins near the boundary. While there are some techniques to understand which skeins are near the boundary, the techniques do not apply to general cases, especially 3-manifolds having higher genus surfaces as their boundaries. This paper aims to solve an equivalent and tractable version of Conjecture 1.2 with a handlebody and 2-handles (Theorem 2.3). The idea comes from the fact that any oriented connected compact 3-manifolds are obtained from handlebodies by gluing 2-handles up to 3-balls; see Section 2.1. With the equivalent version and diagrammatic techniques, we show Conjecture 1.2 affirmatively for some 3-manifolds with boundary and with Heegaard genus 2 or 3 (Theorem 2.13). In particular, the boundary of each 3-manifold in the case of genus 3 is the closed surface of genus 2. This is the first nontrivial example for Conjecture 1.2 with higher genus boundary surfaces. This implies that the examples are not covered by the listed previous results.

2 Finiteness conjecture

2.1 Equivalent versions for the finiteness conjecture

In the following, we use $\mathbb{N}$ to denote the set of nonnegative integers and work on $\mathcal{R} = \mathbb{Q}(q)$ unless otherwise specified, where $\mathbb{Q}(q)$ is the field of rational functions in q over $\mathbb{Q}$.

Let M be an oriented connected compact 3-manifold, and let $\mathcal{C}$ be a collection of disjoint embedded closed curves in ∂M. We use $M^{\mathcal{C}}$ to denote the 3-manifold obtained from M by attaching 2-handles along all the closed curves in $\mathcal{C}$, and call the pair $(M,\mathcal{C})$ an *attaching system*. Let $\iota : M \to M^{\mathcal{C}}$ denote the natural embedding and $\iota_* : \mathscr{S}_q(M) \to \mathscr{S}_q(M^{\mathcal{C}})$ denote the induced homomorphism. Let $\mathscr{S}_q(\partial M)_{\mathcal{C}}$ be the $\mathbb{Q}(q)$-subvector space of $\mathscr{S}_q(\partial M)$ linearly spanned by all diagrams in ∂M having no intersection with $\mathcal{C}$. Clearly $\mathscr{S}_q(\partial M)_{\mathcal{C}}$ is a subalgebra of $\mathscr{S}_q(\partial M)$.

For any $\mathbb{Q}(q)$-subalgebra A of $\mathscr{S}_q(\partial M)$ and any nonempty subset X of $\mathscr{S}_q(M)$, define

$$A \cdot X = \{a_1 \cdot x_1 + \cdots + a_n \cdot x_n \mid n \in \mathbb{N},\ a_i \in A,\ x_i \in X,\ 1 \le i \le n\},$$

which is a $\mathbb{Q}(q)$-subvector space of $\mathscr{S}_q(M)$ (when $n = 0$, we define $a_1 \cdot x_1 + \cdots + a_n \cdot x_n$ to be 0).

For any nonnegative integer g, let H_g denote the handlebody of genus g.

Conjecture 2.1 For any attaching system $(M,\mathcal{C})$, there exists a finite subset Y of $\mathscr{S}_q(M)$ such that $\iota_*(\mathscr{S}_q(\partial M)_{\mathcal{C}} \cdot Y) = \mathscr{S}_q(M^{\mathcal{C}})$.

Let us focus on a special case with handlebodies.

Conjecture 2.2 For any attaching system $(H_g,\mathcal{C})$, there exists a finite subset X of $\mathscr{S}_q(H_g)$ such that $\iota_*(\mathscr{S}_q(\partial H_g)_{\mathcal{C}} \cdot X) = \mathscr{S}_q(H_g^{\mathcal{C}})$.

Let $\operatorname{rank}_\partial \mathscr{S}_q(M)$ denote the minimum number of generators of $\mathscr{S}_q(M)$ as an $\mathscr{S}_q(\partial M)$-module if the minimum number exists and ∞ otherwise. Note that if $\mathscr{S}_q(M)$ is a free $\mathscr{S}_q(\partial M)$-module then $\operatorname{rank}_\partial \mathscr{S}_q(M)$ is equal to the usual rank over $\mathscr{S}_q(\partial M)$.

Theorem 2.3 (1) *If the attaching system $(H_g,\mathcal{C})$ satisfies Conjecture 2.2 then $H_g^{\mathcal{C}}$ satisfies Conjecture 1.2 and*

$$\operatorname{rank}_\partial \mathscr{S}_q(H_g^{\mathcal{C}}) \le |X|.$$

(2) *For any attaching system $(M,\mathcal{C})$, if $M^{\mathcal{C}}$ satisfies Conjecture 1.2, then $(M,\mathcal{C})$ satisfies Conjecture 2.1.*

(3) *Conjectures 1.2, 2.1 and 2.2 are equivalent to each other.*

Proof (1) It suffices to show $\mathscr{S}_q(\partial H_g)_{\mathcal{C}}$ is generated by $\iota_*(X)$ over $\mathscr{S}_q(\partial H_g^{\mathcal{C}})$. For any diagram L in ∂H_g having no intersection with $\mathcal{C}$, and any element $x \in X$, $\iota_*(L \cdot x) = \iota_*(L \cup x) = \iota(L) \cup \iota_*(x)$, where $\iota(L) \subset \partial H_g^{\mathcal{C}}$. Thus $\iota_*(L \cdot x) \in \mathscr{S}_q(\partial H_g^{\mathcal{C}}) \cdot \iota_*(X)$. Then the definition of $\mathscr{S}_q(\partial H_g)_{\mathcal{C}}$ and the fact $\iota_*(\mathscr{S}_q(\partial H_g)_{\mathcal{C}} \cdot X) = \mathscr{S}_q(H_g^{\mathcal{C}})$ imply $\mathscr{S}_q(\partial H_g^{\mathcal{C}}) \cdot \iota_*(X) = \mathscr{S}_q(H_g^{\mathcal{C}})$. The required inequality follows from the argument.

(2) From Conjecture 1.2, we know there is a finite subset Z of $\mathscr{S}_q(M^{\mathcal{C}})$ such that $\mathscr{S}_q(\partial M^{\mathcal{C}}) \cdot Z = \mathscr{S}_q(M^{\mathcal{C}})$. Since $\iota_* : \mathscr{S}_q(M) \to \mathscr{S}_q(M^{\mathcal{C}})$ is surjective [18], there exists a finite subset Y of $\mathscr{S}_q(M)$ such that $\iota_*(Y) = Z$. Note that $\partial M^{\mathcal{C}}$ is obtained from ∂M in the following way. For a small open neighborhood $N(\mathcal{C})$ of $\mathcal{C}$, $\partial M \setminus N(\mathcal{C})$ is a surface with $2|\mathcal{C}|$ circle boundary components. Then, $\partial M^{\mathcal{C}}$ is obtained from $\partial M \setminus N(\mathcal{C})$ by gluing $2|\mathcal{C}|$ disks along all these circle boundary components. Thus any diagrams in $\partial M^{\mathcal{C}}$ can be pushed into $\partial M \setminus N(\mathcal{C})$. Then we have $\mathscr{S}_q(M^{\mathcal{C}}) = \mathscr{S}_q(\partial M^{\mathcal{C}}) \cdot Z \subset \iota_*(\mathscr{S}_q(\partial M)_{\mathcal{C}} \cdot Y)$.

(3) From (2), we know Conjecture 1.2 implies Conjecture 2.1. Obviously, Conjecture 2.1 implies Conjecture 2.2. Note that for any oriented connected compact 3-manifold M, there exists an attaching system $(H_g, \mathcal{C})$ such that M (maybe after cutting out some open 3-balls) is isomorphic to $H_g^{\mathcal{C}}$; see, e.g., [20, Theorem 3.1.10]. Then from (1), Conjecture 2.2 implies Conjecture 1.2. □

Proposition 2.4 *Suppose Conjecture 2.1 holds for an attaching system $(M, \mathcal{C})$, and φ is any diffeomorphism from M to itself. Then Conjecture 2.1 also holds for $(M, \varphi(\mathcal{C}))$.*

Proof For any subset Y of $\mathscr{S}_q(M)$, we have $\varphi_*(\mathscr{S}_q(\partial M)_{\mathcal{C}} \cdot Y) = \mathscr{S}_q(\partial M)_{\varphi(\mathcal{C})} \cdot \varphi_*(Y)$. □

Proposition 2.5 *If $\mathcal{C}$ consists of closed curves such that all the curves bound embedded disks in H_g, then Conjecture 2.2 holds.*

Proof Any skein in $H_g^{\mathcal{C}}$ is a linear sum of skeins in $H_g^{\mathcal{C}}$ having no intersection with the embedded disks bounded by closed curves in $\mathcal{C}$; see [18, Theorem 6.3]. Then we can set $X = \{\varnothing\}$, where $\varnothing$ represents the empty skein. □

Let α be an embedded closed curve in a surface Σ. We say α is *trivial* if it bounds an embedded disk, and say α is *separating* if $\Sigma \setminus \alpha$ has one more component than Σ.

Define $n(\mathcal{C})$ to be the maximum number of nontrivial and nonseparating closed curves which are mutually not isotopic in $\mathcal{C}$.

For any connected surface Σ, let $g(\Sigma)$ denote the sum of genera of all the connected components of Σ.

Corollary 2.6 *Conjecture 2.2 holds when $g(\partial H_g) = n(\mathcal{C})$ and $\mathcal{C}$ generates $H_1(H_g)$.*

Proof We know $H_g^{\mathcal{C}}$ has the same skein module as a closed 3-manifold, where $H_g^{\mathcal{C}}$ is a closed 3-manifold minus 3-balls. Then Theorems 1.1 and 2.3 complete the proof. □

Let $(M_1, \mathcal{C}_1), (M_2, \mathcal{C}_2)$ be two attaching systems. For each $i = 1, 2$, suppose D_i is an embedded disk in ∂M_i such that D_i has no intersection with $\mathcal{C}_i$. We use an orientation reversing diffeomorphism to glue D_1 and D_2. For the pair $\boldsymbol{D} = (D_1, D_2)$, denote the resulting 3-manifold as $M_1 \#_{\boldsymbol{D}} M_2$, and the gluing disk, denoted as D, is a properly embedded disk in $M_1 \#_{\boldsymbol{D}} M_2$. For each $i = 1, 2$, suppose D_i is contained in the boundary component U_i. Consider $(M_1 \#_{\boldsymbol{D}} M_2, \mathcal{C})$ with $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$ if $M_i^{\mathcal{C}_i}$ has a sphere boundary component containing D_i for $i = 1$ or 2, or $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \{\partial D\}$ otherwise.

Proposition 2.7 *If Conjecture 2.1 holds for both $(M_1, \mathcal{C}_1)$ and $(M_2, \mathcal{C}_2)$, then it holds for $(M_1 \#_{\boldsymbol{D}} M_2, \mathcal{C})$.*

Proof Suppose $M_i^{\mathcal{C}_i}$ has a sphere boundary component containing D_i for $i = 1$ or 2. Without loss of generality, we assume that D_1 lies in a sphere boundary component S^2 of $M_1^{\mathcal{C}_1}$. By applying the sphere sliding, explained in the Appendix, to any skein α in $(M_1 \#_{\boldsymbol{D}} M_2)^{\mathcal{C}}$, we get, in $\mathscr{S}_q((M_1 \#_{\boldsymbol{D}} M_2)^{\mathcal{C}})$, α is a linear sum of skeins having no intersection with D. Then the assumption implies Conjecture 2.2 holds for $(M_1 \#_{\mathbf{D}} M_2, \mathcal{C})$.

Suppose $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \{\partial D\}$. For this case, $\partial D \in \mathcal{C}$. The same argument, as above, works by doing handle sliding along the 2-handle attached to ∂D; see also [18]. □

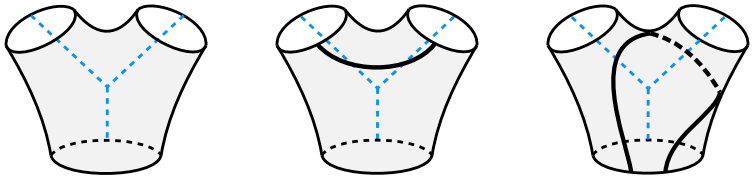


Figure 1: Left: a pair of pants with a part of the dual graph (the blue graph). Middle and right: each bold arc is an embedded arc in a pair of pants whose endpoints avoid the part of the dual graph.

2.2 Dehn–Thurston coordinates

In this subsection, we recall the definition of Dehn–Thurston coordinates; see, e.g., [17] for more details.

A *multicurve* on a surface is a disjoint union of simple closed curves on the surface. A multicurve is *essential* if it has no null-homotopic component.

Let Σ_g denote the orientable closed surface of genus g and $\{C_i\}_{i=1}^{3g-3}$ be the set of nontrivial simple closed curves on Σ_g such that any two of them are not homotopic, known as a *pants decomposition* of Σ_g. Take a small closed neighborhood $N(C_i)$ of C_i in Σ_g $(i = 1, \dots 3g-3)$. After we remove $\bigsqcup_{i=1}^{3g-3} \operatorname{Int} N(C_i)$ from Σ_g, the resulting surface is a disjoint union of pairs of pants, where a pair of pants is a surface homeomorphic to S^2 minus 3 open disks. A *dual graph* is a trivalent graph Γ on Σ_g such that C_i $(i = 1, \dots, 3g-3)$ intersects with Γ just once and the intersection of Γ and each pair of pants has just 1 trivalent vertex.

For a multicurve γ on Σ_g, let $n_i(\gamma)$ $(i = 1, 2, \dots, 3g-3)$ be the geometric intersection number of C_i and γ, i.e., the minimum of the intersection number of C_i and γ' among γ' homotopic to γ.

For an essential multicurve of Σ_g, it is in *general position* if each component of the intersection of the multicurve and the pairs of pants of $\Sigma \setminus \left(\bigsqcup_{i=1}^{3g-3} \operatorname{Int} N(C_i)\right)$ is one of the curves depicted in Figure 1. Here, any simple closed curve parallel to C_i is in $N(C_i)$. We also impose that it intersects with C_i and Γ transversely. It is known that one can isotope any essential multicurve to be in general position and an essential multicurve γ in general position realizes the geometric intersections, i.e., $n_i(\gamma) = \#\{\gamma \cap C_i\}$. We isotope an essential multicurve γ to be in general position, denoted by γ'. Let $t_i(\gamma)$ $(i = 1, \dots, 3g-3)$ be the geometric intersection number of $\gamma' \cap N(C_i)$ and $\Gamma \cap N(C_i)$ if $\gamma' \cap N(C_i)$ consists of only loops, and be the oriented intersection number of $\gamma' \cap N(C_i)$ and $\Gamma \cap N(C_i)$ defined by summing $+1$ (resp. -1) for  (resp. ) over all crossings, where the blue curve is $\Gamma \cap N(C_i)$ and the red curve is a part of $\gamma' \cap N(C_i)$ and the orientations of $\Gamma \cap N(C_i)$ and all the arcs of $\gamma' \cap N(C_i)$ are assigned to be inward/outward on the same boundary component of $N(C_i)$. It is known that $t_i(\gamma)$ is well defined. The *Dehn–Thurston coordinate* of an essential multicurve γ with respect to $\{C_i\}$ and Γ is the coordinate

$$(n_1(\gamma), \dots, n_{3g-3}(\gamma), t_1(\gamma), \dots, t_{3g-3}(\gamma)) \in \mathbb{N}^{3g-3} \times \mathbb{Z}^{3g-3}.$$

In the following, we abbreviate it to the DT coordinate and let $\mathrm{DT}(\gamma)$ denote the coordinate of γ. It is known that there is a one-to-one correspondence between the set of the DT coordinates of essential multicurves on Σ_g and the set of isotopy classes of essential multicurves on Σ_g.

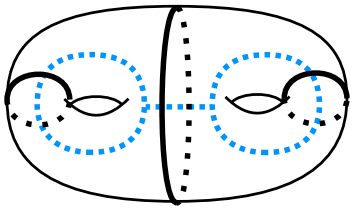


Figure 2: The bold curves on ∂H_2 are C_1, C_3, C_2 from left to right and the blue graph is a dual graph.

We will use the pants decompositions depicted in Figures 2 and 5 for ∂H_2 and ∂H_3, respectively.

2.3 Generating sets and bases for H_2 and H_3

For any nonnegative integer m, let Σ_0^m denote the surface obtained from S^2 by removing m open disks. Then H_g can be regarded as $\Sigma_0^{g+1} \times [0,1]$.

Let x, y, z be the peripheral loops of Σ_0^3 depicted as in Figure 3. We also use x, y, z to denote diagrams on $\Sigma_0^3 \times \{1\} \subset \partial(\Sigma_0^3 \times [0,1]) = \partial H_2$. So we can also regard x, y, z as elements in $\mathscr{S}_q(\partial H_2)$.

The ground ring $\mathcal{R}$ equals $\mathbb{Z}[q^{\pm 1}]$ (the Laurent polynomial ring in q with integer coefficients) in Lemmas 2.8–2.10. The following result is well known.

Lemma 2.8 *The skein algebra $\mathscr{S}_q(\Sigma_0^3)$ is isomorphic to $\mathcal{R}[x, y, z]$.*

For a polynomial $P(x,z,y) = \sum_{i,j,k} c(i,j,k) x^i z^j y^k \in \mathbb{Q}(q)[x,z,y]$, $c(i,j,k) \in \mathbb{Q}(q)$, its $(x,z;y)$-*degree* is the maximum number among $i+j+2k$ with $c(i,j,k) \neq 0$.

Let $s_1, s_2, s_3, s_{12}, s_{13}, s_{23}, s_{123}$ be the simple closed curves in Σ_0^4 as depicted in Figure 4.

For later convenience, for any $\vec{k} = (k_1, k_2, k_3, k_{12}, k_{13}, k_{23}, k_{123}) \in \mathbb{N}^7$, put

$$
\begin{aligned}
s^{\vec{k}} &:= s_1^{k_1} s_2^{k_2} s_3^{k_3} s_{12}^{k_{12}} s_{13}^{k_{13}} s_{23}^{k_{23}} s_{123}^{k_{123}} \in \mathscr{S}_q(\Sigma_0^4), \\
\mathrm{s}(\vec{k}) &:= 3(k_1 + k_2 + k_3 + k_{123}) + 4(k_{12} + k_{13} + k_{23}), \\
\mathrm{s}'(\vec{k}) &:= 2(k_1 + k_2) + k_3 + 2k_{12} + 3(k_{13} + k_{23} + k_{123}).
\end{aligned} \tag{2-1}
$$

Let Λ denote the subset of $\mathbb{N}^7$ defined by

$$\Lambda = \{(k_1, k_2, k_3, k_{12}, k_{13}, k_{23}, k_{123}) \in \mathbb{N}^7 \mid k_{12} k_{13} k_{23} = 0\}.$$

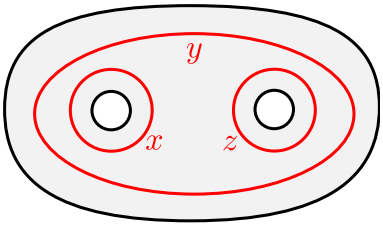


Figure 3: Σ_0^3 and algebraic generators x, y, z in Σ_0^3.

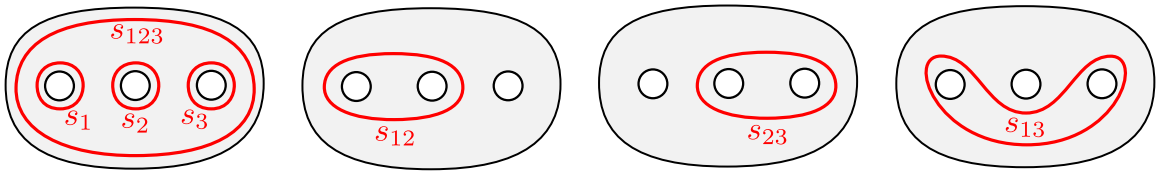


Figure 4: The algebraic generators $s_1, s_2, s_3, s_{12}, s_{13}, s_{23}, s_{123}$ of $\mathscr{S}_q(\Sigma_0^4)$.

Lemma 2.9 [5] *As a free module over $\mathcal{R}$, the skein algebra $\mathscr{S}_q(\Sigma_0^4)$ has the basis $\{s^{\vec{k}} \mid \vec{k} \in \Lambda\}$.*

Lemma 2.10 [5] *When $q = 1$, the skein algebra $\mathscr{S}_1(\Sigma_0^4)$ is a commutative algebra over $\mathbb{Z}$ generated by $s_1, s_2, s_3, s_{12}, s_{13}, s_{23}, s_{123}$ subject to*

$$s_{12}s_{13}s_{23} = s_{12}^2 + s_{13}^2 + s_{23}^2 + s_{12}(s_1s_2 + s_3s_{123}) + s_{13}(s_1s_3 + s_2s_{123}) + s_{23}(s_2s_3 + s_1s_{123}) + s_1s_2s_3s_{123} + s_1^2 + s_2^2 + s_3^2 + s_{123}^2 - 4.$$

For an essential multicurve γ in Σ_0^4, let $m_i(\gamma)$ denote the geometric intersection number of γ and the i-th edge in Figure 5. Put

$$\vec{m}(\gamma) := (m_1(\gamma), m_2(\gamma), m_3(\gamma), m_4(\gamma), m_5(\gamma), m_6(\gamma)) \in \mathbb{N}^6, \tag{2-2}$$

$$\mathrm{sum}(\gamma) := m_1(\gamma) + m_2(\gamma) + m_3(\gamma) + m_4(\gamma) + m_5(\gamma) + m_6(\gamma). \tag{2-3}$$

Then we have

$$\vec{m}(s_1) = (1,1,0,0,0,1),\ \vec{m}(s_2) = (0,1,1,0,1,0),\ \vec{m}(s_3) = (0,0,1,1,0,1),$$
$$\vec{m}(s_{12}) = (1,0,1,0,1,1),\ \vec{m}(s_{13}) = (1,1,1,1,0,0),\ \vec{m}(s_{23}) = (0,1,0,1,1,1),$$
$$\vec{m}(s_{123}) = (1,0,0,1,1,0).$$

For any $\vec{k} \in \mathbb{N}^7$, it is easy to show that $\mathrm{s}(\vec{k})$ is the sum of all the entries of

$$k_1\vec{m}(s_1) + k_2\vec{m}(s_2) + k_3\vec{m}(s_3) + k_{12}\vec{m}(s_{12}) + k_{13}\vec{m}(s_{13}) + k_{23}\vec{m}(s_{23}) + k_{123}\vec{m}(s_{123}). \tag{2-4}$$

Lemma 2.11 *For any distinct $\vec{u}, \vec{v} \in \Lambda$, the values (2-4) for $\vec{u}$ and $\vec{v}$ are also distinct.*

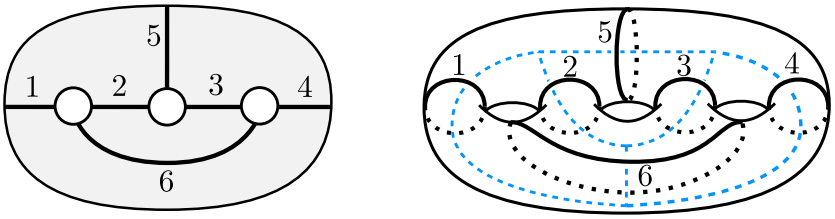


Figure 5: Left: Σ_0^4 with edges giving a triangulation of $\mathrm{Int}\,\Sigma_0^4$. The edges are numbered as in the picture. Right: a pants decomposition $\{C_i\}_{i=1}^6$ for Σ_3^0 and a dual graph (the blue graph). Each C_i corresponds with the i-th curve.

Proof One can show that, by regarding

$$\vec{m}(s_1),\quad \vec{m}(s_2),\quad \vec{m}(s_3),\quad \vec{m}(s_{12}),\quad \vec{m}(s_{13}),\quad \vec{m}(s_{23}),\quad \vec{m}(s_{123})$$

as vectors and removing one of $\vec{m}(s_{12}), \vec{m}(s_{13}), \vec{m}(s_{23})$ from them, the remaining vectors are linearly independent, and

$$\vec{m}(s_1)+\vec{m}(s_2)+\vec{m}(s_3)-\vec{m}(s_{12})-\vec{m}(s_{13})-\vec{m}(s_{23})+\vec{m}(s_{123})=0. \tag{2-5}$$

Suppose

$$k_1\vec{m}(s_1)+k_2\vec{m}(s_2)+k_3\vec{m}(s_3)+k_{12}\vec{m}(s_{12})+k_{13}\vec{m}(s_{13})+k_{23}\vec{m}(s_{23})+k_{123}\vec{m}(s_{123})=0$$

for $(k_1,k_2,k_3,k_{12},k_{13},k_{23},k_{123})\in\mathbb{Q}^7$. The above shows the solution space of (2-5) is 1-dimensional, i.e.,

$$(k_1,k_2,k_3,k_{12},k_{13},k_{23},k_{123})=c(1,1,1,-1,-1,-1,1),$$

where $c\in\mathbb{Q}$. Thus

$$k_{12},k_{13},k_{23}\le 0 \text{ or } k_{12},k_{13},k_{23}\ge 0. \tag{2-6}$$

Assume there exist distinct $\vec{u},\vec{v}\in\Lambda$ such that the values (2-4) of $\vec{u}$ and $\vec{v}$ are the same. Then we have

$$\begin{aligned}(u_1-v_1)\vec{m}(s_1)+(u_2-v_2)\vec{m}(s_2)+(u_3-v_3)\vec{m}(s_3)+(u_{12}-v_{12})k_{12}\vec{m}(s_{12})\\ +(u_{13}-k_{13})\vec{m}(s_{13})+(u_{23}-v_{23})\vec{m}(s_{23})+(u_{123}-v_{123})\vec{m}(s_{123})=0.\end{aligned} \tag{2-7}$$

If $(u_{12}-v_{12})(u_{13}-v_{13})(u_{23}-v_{23})=0$, then linear independence discussed above shows $\vec{u}=\vec{v}$. This contradicts with the assumption that $\vec{u}\neq\vec{v}$. Suppose $(u_{12}-v_{12})(u_{13}-k_{13})(u_{23}-v_{23})\le 0$. Since $\vec{u}\in\Lambda$, i.e., $u_{12}u_{13}u_{23}=0$, we suppose $u_{12}=0$ without loss of generality. Since $u_{12}-v_{12}\neq 0$, we have $v_{12}\neq 0$ and $u_{12}-v_{12}<0$. Then $v_{13}=0$ or $v_{23}=0$ from $\vec{v}\in\Lambda$. We suppose $v_{13}=0$. Since $u_{13}-v_{13}\neq 0$, we have $u_{13}\neq 0$ and $u_{13}-v_{13}>0$. By combining (2-7), $u_{12}-v_{12}<0$ and $u_{13}-v_{13}>0$, the assumption contradicts with (2-6). □

Remark 2.12 Lemma 2.11 implies the independence of the basis elements in Lemma 2.9.

2.4 Proof of the finiteness conjecture for a family of 3-manifolds

In this subsection, the goal is to prove the following theorem.

Theorem 2.13 *Let $\gamma\subset\partial H_g$ be a simple closed curve and H_g^γ denote the resulting 3-manifold obtained from H_g by attaching a 2-handle along γ. If γ satisfies one of the following conditions, the finiteness conjecture holds for H_g^γ, where the DT coordinates are with respect to the pants decompositions and the dual graphs defined in Figures 2 and 5.*

Case 1 *$g=2$ with one of the following*:

- $\mathrm{DT}(\gamma)=(n,n,2n,t_1,t_1,t_2)$ *where* $t_1t_2\ge 0$ *and* $n=|2t_2+t_1|$ *(symmetric)*,
- $\mathrm{DT}(\gamma)=(n+m,n,2n,t,0,0)$ *or* $(n,n+m,2n,t,0,0)$ *where* $t\in\{n,-n\}$,
- $\mathrm{DT}(\gamma)=(n_1,n_2,2,\pm1,\pm1,0)$,

where $n\in\mathbb{Z}_{\ge 1}$, $m,n_1,n_2\in\mathbb{N}$.

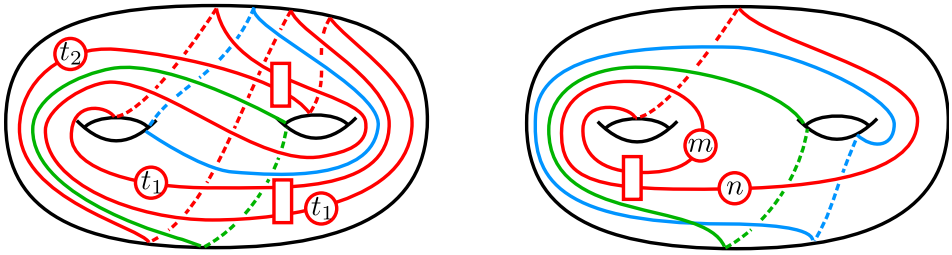


Figure 6: Each circle with nonnegative integer k represents k parallel copies of the corresponding arcs. Each coupon in the picture represents certain number of horizontal lines connecting the endpoints on the left side of the coupon and those on the right side of the coupon. Left: the red curve is the gluing curve γ whose DT coordinate is $(n, n, 2n, t_1, t_1, t_2)$. Right: the red curve is the gluing curve γ whose DT coordinate is $(n+m, n, 2n, n, 0, 0)$.

Case 2 *$g = 3$ with one of the following*:

- $\mathrm{DT}(\gamma) = (1, n, n+m, m+1, n+1, m, 0, 0, 0, 1, 1, 0)$ *or* $(1, n, n+m, m+1, n+1, m, 0, 0, 0, -1, -1, 0)$,
- $\mathrm{DT}(\gamma) = (n+m, m, 0, n, n, m+2n, t, 0, 0, 0, 0, 0)$,

where $n, m \in \mathbb{Z}_{\geq 1}$. In particular, $\operatorname{rank}_\partial(H_2^\gamma) = 1$ *in the second case of Case* 1 *and* $\operatorname{rank}_\partial(H_3^\gamma) = 1$ *for both two cases in Case* 2.

For convenience, we will use two different framings for curves on ∂H_g in the proof. The framings of all gluing curves (drawn in red) on the boundary in Theorem 2.13 are the usual blackboard framings. On the other hand, other boundary curves in Theorem 2.13 are equipped with the blackboard framings with respect to the projection $H_g = \Sigma_0^{g+1} \times [0,1] \to \Sigma_0^{g+1}$.

Proof **The first case in Case 1:** We can assume both t_1 and t_2 are nonnegative integers since the proof in the case that they are both nonpositive integers is similar. The red curve in the left of Figure 6 is the gluing curve γ. Note that the green curve α_1 and the blue curve α_2, respectively, correspond to x and z in $\mathscr{S}_q(\Sigma_0^3)(=\mathscr{S}_q(H_2))$ depicted in the left of Figure 6. Since the green curve and the blue curve do not intersect with the red curve, $x, z \in \mathscr{S}_q(\partial H_g)_\gamma$. Obviously we have $\gamma \in \mathscr{S}_q(\partial H_2)_\gamma$. Set $X = \{1, y, \cdots, y^{n-1}\} \subset \mathscr{S}_q(H_2)$. Then it suffices to show $\mathscr{S}_q(\partial H_2)_\gamma \cdot X = \mathscr{S}_q(\Sigma_0^3)(=\mathscr{S}_q(H_2))$ as $\mathbb{Q}(q)$-vector spaces. As we mentioned, $\alpha_1 = x$, $\alpha_2 = z$, $y \in \mathscr{S}_q(\partial H_2)_\gamma \cdot X$. Fix $2 \leq k \leq 2n-1$ and assume that $x^{k_1} z^{k_2} y^{k_3} \in \mathscr{S}_q(\partial H_2)_\gamma \cdot X$ for any solutions of $k_1 + k_2 + 2k_3 < k$ $(k_1, k_2, k_3 \in \mathbb{N})$. For $k_1, k_2, k_3 \in \mathbb{N}$ satisfying $k = k_1 + k_2 + 2k_3$, consider $\alpha_1^{k_1} \alpha_2^{k_2} \cdot y^{k_3}$. From the geometric intersection numbers in the DT coordinates of α_1, α_2 and y,

$$\alpha_1^{k_1} \alpha_2^{k_2} \cdot y^{k_3} = \sum_{k=i+j+2l} c(k_1, k_2, k_3; i, j, l) x^i z^j y^l + (\text{lower } (x, z; y)\text{-degree terms}), \tag{2-8}$$

where $c(k_1, k_2, k_3; i, j, l) \in \mathbb{Z}[q^{\pm 1}] \subset \mathbb{Q}(q)$. When we substitute 1 to q, $\mathscr{S}_1(H_2)$ is a commutative algebra and we have $\alpha_1^{k_1} \alpha_2^{k_2} \cdot y^{k_3} = x^{k_1} z^{k_2} y^{k_3} \in \mathscr{S}_1(H_2)$. With the lexicographic order on (k_1, k_2, k_3), the coefficient matrix $(c(k_1, k_2, k_3; i, j, l))$ obtained from the highest terms in (2-8) is full rank since it

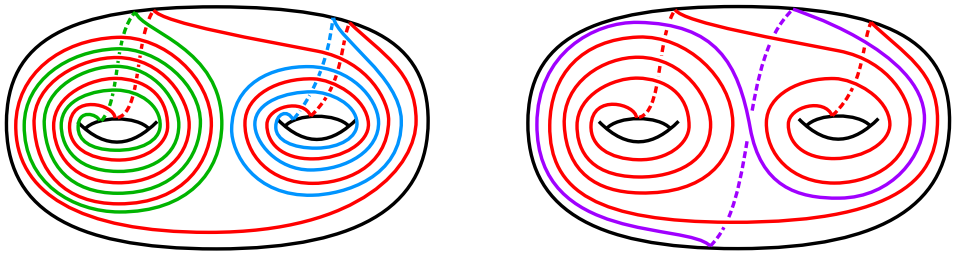


Figure 7: The red curve is the gluing curve γ with $\mathrm{DT}(\gamma)=(4,3,2,1,1,0)$.

will be the identity matrix by substituting 1 to q, where $(c(k_1,k_2,k_3;i,j,l))$ is a square matrix whose rows and columns run over all solutions of $k=k_1+k_2+2k_3$ and $k=i+j+2l$, respectively. This implies that $x^{k_1}z^{k_2}y^{k_3}\in\mathscr{S}_q(\partial H_2)_\gamma\cdot X$.

Fix $k\geq 2n$. Assume that $x^{k_1}z^{k_2}y^{k_3}\in\mathscr{S}_q(\partial H_2)_\gamma\cdot X$ for any $k_1,k_2,k_3\in\mathbb{N}$ satisfying $k_1+k_2+2k_3<k$. Consider $\alpha_1^{k_1}\alpha_2^{k_2}\cdot y^{k_3}$ for $k_1,k_2,k_3\in\mathbb{N}$ satisfying $k=k_1+k_2+2k_3$ and $k_3<n$. Also consider $\gamma^m\alpha_1^{k_1}\alpha_2^{k_2}\cdot y^{k_3-mn}$ for $k_1,k_2,k_3,m\in\mathbb{N}$ satisfying $k=k_1+k_2+2k_3$, $k_3\geq n$ and $0\leq k_3-mn<n$. When we substitute 1 to q then we take the absolute value of coefficients, we have the (absolute valued) coefficient matrix $\left(\begin{smallmatrix} I & O \\ * & T\end{smallmatrix}\right)$, where I is the identity matrix and T is obtained from a triangular matrix whose diagonal entries are 1 by interchanging some rows. Since the matrix is full rank, $x^{k_1}z^{k_2}y^{k_3}\in\mathscr{S}_q(\partial H_2)_\gamma\cdot X$.

The second case in Case 1: The red curve in the right of Figure 6 is the gluing curve with the DT coordinate $(n+m,n,2n,n,0,0)$. We only prove this case since a similar discussion works for other parallel cases. The green curve α in the right of Figure 6 shows $x\in\mathscr{S}_q(\partial H_2)_\gamma$. The blue curve β in the right of Figure 6 shows $y\in\mathscr{S}_q(\partial H_2)_\gamma$. Obviously we also have $z\in\mathscr{S}_q(\partial H_2)_\gamma$. We set $X=\{\varnothing\}$. For any $k_1,k_2,k_3\in\mathbb{N}$, we have

$$z^{k_2}\beta^{k_3}\alpha^{k_1}\cdot\varnothing=x^{k_1}z^{k_2}y^{k_3}\in\mathscr{S}_q(\partial H_2)_\gamma\cdot X.$$

In particular, $\operatorname{rank}_\partial(H_2^\gamma)=1$ in this case.

The third case in Case 1: We only prove $\mathrm{DT}(\gamma)=(n_1,n_2,2,1,1,0)$ since a similar argument works for other parallel cases. For the reader's convenience, the red curve in Figure 7 is the gluing curve with $n_1=4$, $n_2=3$. If $n_1=0$ or $n_2=0$, it is covered by Corollary 2.6. Suppose $n_1n_2\neq 0$. Let α_1,α_2 and α_3 denote simple closed curves on ∂H_2 whose DT coordinates are $(n_1,0,0,1,0,0)$, $(0,n_2,0,0,1,0)$, and $(1,1,2,0,0,1)$, respectively. Note that α_i does not intersect with γ, i.e., $\alpha_i\in\mathscr{S}_q(\partial H_2)_\gamma$ for $i=1,2,3$. In $\mathscr{S}_q(H_2)$, we have

$$\alpha_1=u_{n_1}x^{n_1}+\cdots+u_1x+u_0,\quad \alpha_2=v_{n_2}z^{n_2}+\cdots+v_1z+v_0,\quad \alpha_3=qy+q^{-1}xz,$$

where u_{n_1} and v_{n_2} are nonzero elements in $\mathbb{Q}(q)$. Set

$$X=\{x^{k_1}z^{k_2}\mid k_i\in\mathbb{N},\ k_i\leq n_i-1 \text{ for } i=1,2\}. \tag{2-9}$$

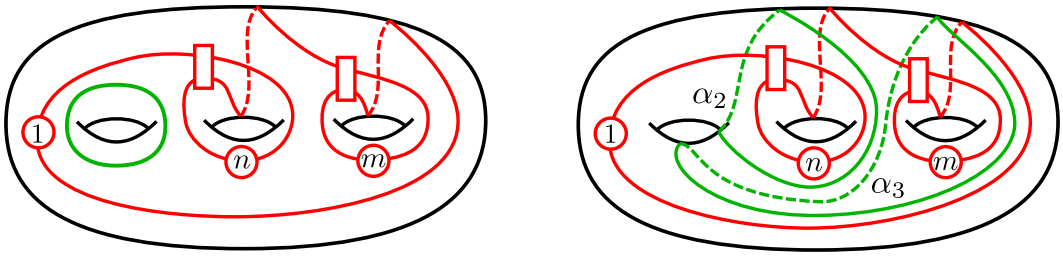


Figure 8: The red curve is the gluing curve γ. The green curve in the left picture is α_1 and the green curves in the right picture are α_2 and α_3.

To show

$$\{x^{k_1}z^{k_2} \mid k_1, k_2 \in \mathbb{N}\} \subset \mathscr{S}_q(\partial H_2)_\gamma \cdot X, \tag{2-10}$$

assume that, for $k \geq 1$, $x^{k_1}z^{k_2} \in \mathscr{S}_q(\partial H_2)_\gamma \cdot X$ for any $k_1, k_2 \in \mathbb{N}$ with $k_1 + k_2 < k$. For $k_1, k_2 \in \mathbb{N}$ with $k_1 + k_2 = k$, consider $\alpha_1^{m_1}\alpha_2^{m_2}x^{k_1-m_1n_1}z^{k_2-m_2n_2}$, where $m_i \in \mathbb{N}$ satisfies $0 \leq k_i - m_i n_i < n_i$ $(i = 1, 2)$. Then, we have

$$\alpha_1^{m_1}\alpha_2^{m_2} \cdot x^{k_1-m_1n_1}z^{k_2-m_2n_2} = q^c x^{k_1}z^{k_2} + (\text{lower } (x, z; y)\text{-degree terms}),$$

where $c \in \mathbb{Z}$ and y does not appear on the right-hand side. Hence, $x^{k_1}z^{k_2} \in \mathscr{S}_q(\partial H_2)_\gamma \cdot X$, i.e., (2-10) holds.

Fix $k \geq 1$ and suppose that $x^{k_1}z^{k_2}y^{k_3} \in \mathscr{S}_q(\partial H_2)_\gamma \cdot X$ for any $k_1, k_2 \in \mathbb{N}$, $k_3 < k$. Then, we have

$$\alpha_3^k \cdot x^{k_1}z^{k_2} = q^c x^{k_1}z^{k_2}y^k + (\text{lower } y\text{-degree terms}),$$

where $c \in \mathbb{Z}$. This implies that $x^{k_1}z^{k_2}y^k \in \mathscr{S}_q(\partial H_2)_\gamma \cdot X$.

The first case in Case 2: We only prove the case when $\mathrm{DT}(\gamma) = (1, n, n+m, m+1, n+1, m, 0, 0, 0, 1, 1, 0)$. We will use the elements $\alpha_1, \alpha_2, \alpha_3, \alpha_{12}, \alpha_{13}, \alpha_{23}, \alpha_{123} \in \mathscr{S}_q(\partial H_3)_\gamma$ which are shown in Figures 8, 9, 10.

We set $X = \{\varnothing\}$. For any $k_1, k_2, k_3 \in \mathbb{N}$, we have $\alpha_1^{k_1}\alpha_2^{k_2}\alpha_3^{k_3} \cdot \varnothing = s_1^{k_1}s_2^{k_2}s_3^{k_3} \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X$.

For any $\vec{k} \in \Lambda$, we define

$$\deg(s^{\vec{k}}) := \mathrm{s}'(\vec{k}),$$

where $\mathrm{s}'(\vec{k})$ was defined in (2-1). We will use mathematical induction on $\deg(s^{\vec{k}})$ to show $s^{\vec{k}}$ is an element of $\mathscr{S}_q(\partial H_3)_\gamma \cdot X$ for any $\vec{k} \in \Lambda$.

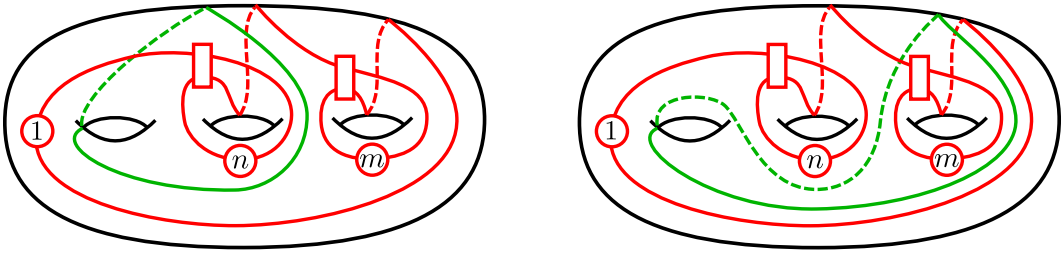


Figure 9: The green curve in the left (resp. right) picture is α_{12} (resp. α_{23}).

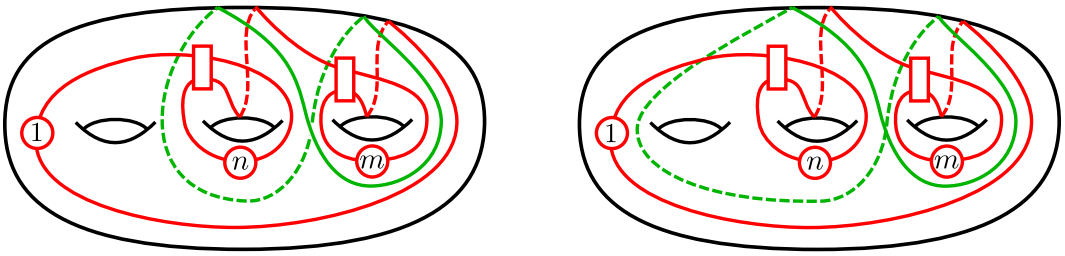


Figure 10: The green curve in the left (resp. right) picture is α_{23} (resp. α_{123}).

We will use Lemmas 2.14 and 2.15 in the rest of the proof, and will prove them in Section 3.

Lemma 2.14 *For any $\vec{k}=(k_1,k_2,k_3,k_{12},k_{13},k_{23},k_{123})\in\mathbb{N}^7$, suppose there exists a finite subset Λ_0 of Λ such that*

$$\alpha_{123}^{k_{123}}\alpha_{12}^{k_{12}}\alpha_{13}^{k_{13}}\alpha_{23}^{k_{23}}\cdot(s_1^{k_1}s_2^{k_2}s_3^{k_3})=\sum_{\vec{u}\in\Lambda_0}C_{\vec{u}}s^{\vec{u}}, \tag{2-11}$$

where $0\neq C_{\vec{u}}\in\mathbb{Z}[q^{\pm1}]$ for any $\vec{u}\in\Lambda_0$. Then, we have $\deg(s^{\vec{u}})\leq s'(\vec{k})$ for any $\vec{u}\in\Lambda_0$.

Lemma 2.15 *Suppose $q=1$. For any positive integers k_{12},k_{13},k_{23}, we have*

$$s_{12}^{k_{12}}s_{13}^{k_{13}}s_{23}^{k_{23}}=\sum_{\vec{u}\in\Lambda_0}C_{\vec{u}}s^{\vec{u}}+(\text{terms with degrees less than } 2k_{12}+3(k_{13}+k_{23})),$$

where Λ_0 is a finite subset of Λ such that, for any $\vec{u}=(u_1,u_2,u_3,u_{12},u_{13},u_{23},u_{123})\in\Lambda_0$, we have $\deg(\vec{u})=2k_{12}+3(k_{13}+k_{23})$ and $u_{13}u_{23}<k_{13}k_{23}$.

Assume that $s^{\vec{k}}\in\mathscr{S}_q(\partial H_3)_\gamma\cdot X$ for any $\vec{k}\in\Lambda$ with $\deg(s^{\vec{k}})<k$. For any

$$\vec{v}=(v_1,v_2,v_3,v_{12},v_{13},v_{23},v_{123})\in\Lambda$$

with $\deg(s^{\vec{v}})=k$, Lemma 2.14 implies that

$$\alpha_{123}^{v_{123}}\alpha_{12}^{v_{12}}\alpha_{13}^{v_{13}}\alpha_{23}^{v_{23}}\cdot(s_1^{v_1}s_2^{v_2}s_3^{v_3})=\sum_{\substack{\vec{u}\in\Lambda\\ \deg(s^{\vec{u}})=k}}C_{\vec{v},\vec{u}}s^{\vec{u}}+(\text{lower degree terms}). \tag{2-12}$$

We consider a total order on $\mathbb{N}^7$ defined by the lexicographic order with respect to

$$(v_{13}v_{23},v_{123},v_{23},v_{13},v_{12},v_3,v_2,v_1),$$

where $(v_1,v_2,v_3,v_{12},v_{13},v_{23},v_{123})\in\mathbb{N}^7$. Then this order induces the total order on $\{s^{\vec{u}}\,|\,\vec{u}\in\Lambda,\ \deg(s^{\vec{u}})=k\}$. When $q=1$, note that the left-hand side of (2-12) is equal to

$$\begin{aligned}(s_{123}+s_{12}s_3)^{v_{123}}s_{12}^{v_{12}}s_{13}^{v_{13}}(s_{23}+s_2s_3)^{v_{23}}s_1^{v_1}s_2^{v_2}s_3^{v_3}\\ =\sum_{i=0}^{v_{123}}\sum_{j=0}^{v_{23}}\binom{v_{123}}{i}\binom{v_{23}}{j}s_{123}^{v_{123}-i}s_{12}^{v_{12}+i}s_{13}^{v_{13}}s_{23}^{v_{23}-j}s_1^{v_1}s_2^{v_2+j}s_3^{v_3+i+j},\end{aligned} \tag{2-13}$$

where the coefficients denote binomial coefficients.

In the case when $v_{12}=0$ and $v_{13}v_{23}\neq 0$, we have

$$\sum_{i=0}^{v_{123}}\sum_{j=0}^{v_{23}}\binom{v_{123}}{i}\binom{v_{23}}{j}s_{123}^{v_{123}-i}\,s_{12}^{v_{12}+i}\,s_{13}^{v_{13}}\,s_{23}^{v_{23}-j}\,s_1^{v_1}\,s_2^{v_2+j}\,s_3^{v_3+i+j}$$
$$=\sum_{j=0}^{v_{23}}\binom{v_{23}}{j}s_{123}^{v_{123}}\,s_{13}^{v_{13}}\,s_{23}^{v_{23}-j}\,s_1^{v_1}\,s_2^{v_2+j}\,s_3^{v_3+i+j}$$
$$+\sum_{i=1}^{v_{123}}\sum_{j=0}^{v_{23}}\binom{v_{123}}{i}\binom{v_{23}}{j}s_{123}^{v_{123}-i}\,s_{12}^{i}\,s_{13}^{v_{13}}\,s_{23}^{v_{23}-j}\,s_1^{v_1}\,s_2^{v_2+j}\,s_3^{v_3+i+j}.$$

The highest term of

$$\sum_{j=0}^{v_{23}}\binom{v_{23}}{j}s_{123}^{v_{123}}\,s_{13}^{v_{13}}\,s_{23}^{v_{23}-j}\,s_1^{v_1}\,s_2^{v_2+j}\,s_3^{v_3+i+j}$$

with respect to the total order is $s_{123}^{v_{123}}s_{13}^{v_{13}}s_{23}^{v_{23}}s_1^{v_1}s_2^{v_2}s_3^{v_3}$. Lemma 2.15 implies the highest term of

$$\sum_{i=1}^{v_{123}}\sum_{j=0}^{v_{23}}\binom{v_{123}}{i}\binom{v_{23}}{j}s_{123}^{v_{123}-i}\,s_{12}^{i}\,s_{13}^{v_{13}}\,s_{23}^{v_{23}-j}\,s_1^{v_1}\,s_2^{v_2+j}\,s_3^{v_3+i+j}$$

with respect to the total order is lower than $s_{123}^{v_{123}}s_{13}^{v_{13}}s_{23}^{v_{23}}s_1^{v_1}s_2^{v_2}s_3^{v_3}$. Thus the highest term of (2-13) with respect to the total order is $s_{123}^{v_{123}}s_{13}^{v_{13}}s_{23}^{v_{23}}s_1^{v_1}s_2^{v_2}s_3^{v_3}$.

In the case when $v_{13}v_{23}=0$, it is easy to see that the highest term of (2-13) with respect to the total order is $s_{123}^{v_{123}}s_{12}^{v_{12}}s_{13}^{v_{13}}s_{23}^{v_{23}}s_1^{v_1}s_2^{v_2}s_3^{v_3}$.

The above two cases imply the matrix $(C_{\vec{v},\vec{u}})_{\vec{v},\vec{u}\in\Lambda,\deg(s^{\vec{v}})=\deg(s^{\vec{u}})=k}$ is a lower triangular matrix whose diagonal entries are 1 when $q=1$. Then, using the same technique as in the proof of the first case of Case 1, one can show $s^{\vec{v}}\in\mathscr{S}_q(\partial H_3)_\gamma\cdot X$ for any $\vec{v}\in\Lambda$. Thus $\iota_*(\mathscr{S}_q(\partial H_3)_\gamma\cdot X)=\mathscr{S}_q(H_3^\gamma)$.

The second case in Case 2. In the remainder of the proof, we will use $\vec{k}\in\mathbb{N}^7$ to denote the vector $(k_1,k_2,k_3,k_{12},k_{23},k_{13},k_{123})$ and use the diagrams in Figures 11, 12, and 13.

We set $X=\{\varnothing\}$. First,

$$\{s_1^{k_1}s_2^{k_2}s_3^{k_3}\mid k_1,k_2,k_3\in\mathbb{N}\}\subset\mathscr{S}_q(\partial H_3)_\gamma\cdot X \tag{2-14}$$

follows from

$$\beta_2^{k_2}\beta_3^{k_3}\beta_1^{k_1}\cdot\varnothing=\beta_2^{k_2}\beta_3^{k_3}\cdot s_1^{k_1}=s_1^{k_1}s_2^{k_2}s_3^{k_3}\in\mathscr{S}_q(\partial H_3)_\gamma\cdot X.$$

Next, we will show

$$\{s_1^{k_1}s_2^{k_2}s_3^{k_3}s_{13}^{k_{13}}\mid k_1,k_2,k_3,k_{13}\in\mathbb{N}\}\subset\mathscr{S}_q(\partial H_3)_\gamma\cdot X. \tag{2-15}$$

When $k_{13}=0$, $s_1^{k_1}s_2^{k_2}s_3^{k_3}s_{13}^{k_{13}}\in\mathscr{S}_q(\partial H_3)_\gamma\cdot X$ follows from (2-14). Suppose $k_{13}>0$ and

$$s_1^{k_1}s_2^{k_2}s_3^{k_3}s_{13}^{l}\in\mathscr{S}_q(\partial H_3)_\gamma\cdot X$$

holds for $0 \leq l < k_{13}$. We have

$$\beta_3^{k_3}\beta_{13}^{k_{13}} \cdot (s_1^{k_1}s_2^{k_2}) = s_1^{k_1}s_2^{k_2}s_3^{k_3}s_{13}^{l} + \sum_{0\leq l<k_{13}} s_1^{k_1}s_2^{k_2}s_3^{k_3}s_{13}^{l} \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X.$$

From the assumption, we have $\sum_{0\leq l<k_{13}} s_1^{k_1}s_2^{k_2}s_3^{k_3}s_{13}^{l} \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X$. Thus we have

$$s_1^{k_1}s_2^{k_2}s_3^{k_3}s_{13}^{k_{13}} \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X.$$

Next, we will show

$$\{s^{\vec{k}} \mid \vec{k} \in \Lambda,\ k_{12} = 0\} \subset \mathscr{S}_q(\partial H_3)_\gamma \cdot X. \tag{2-16}$$

When $k_{123} = 0$, we have

$$\beta_{23}^{k_{23}} \cdot (s_1^{k_1}s_2^{k_2}s_3^{k_3}s_{13}^{k_{13}}) = s_1^{k_1}s_2^{k_2}s_3^{k_3}s_{13}^{k_{13}}s_{23}^{k_{23}} \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X.$$

Suppose $k_{123} > 0$ and $s_1^{k_1}s_2^{k_2}s_3^{k_3}s_{13}^{k_{13}}s_{23}^{k_{23}}s_{123}^{l} \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X$ holds for any $0 \leq l < k_{123}$. We have

$$\begin{aligned}\beta_2^{k_2}\beta_{23}^{k_{23}}(\beta'_{123})^{k_{123}} \cdot (s_1^{k_1}s_3^{k_3}s_{13}^{k_{13}}) \\ = s_1^{k_1}s_2^{k_2}s_3^{k_3}s_{13}^{k_{13}}s_{23}^{k_{23}}s_{123}^{k_{123}} + \sum_{0\leq l<k_{123}} s_1^{k_1}s_2^{k_2}s_3^{k_3}s_{13}^{k_{13}}s_{23}^{k_{23}}s_{123}^{l} \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X.\end{aligned}$$

From the assumption, we have $\sum_{0\leq l<k_{123}} s_1^{k_1}s_2^{k_2}s_3^{k_3}s_{13}^{k_{13}}s_{23}^{k_{23}}s_{123}^{l} \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X$. Thus we have $s_1^{k_1}s_2^{k_2}s_3^{k_3}s_{13}^{k_{13}}s_{23}^{k_{23}}s_{123}^{k_{123}} \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X$.

By replacing the role of $\beta_{13}, s_{13}, \beta'_{123}$ with $\beta_{12}, s_{12}, \beta_{123}$, similar to the above, one can show

$$\{s^{\vec{k}} \mid \vec{k} \in \Lambda,\ k_{13} = 0\} \subset \mathscr{S}_q(\partial H_3)_\gamma \cdot X. \tag{2-17}$$

We will use the following lemma for the rest of proof and will prove it in Section 3.

Lemma 2.16 *Let β be a diagram in Σ_0^4 such that it intersects each C_i (see Figure 5), $1 \leq i \leq 6$, at m_i points. Suppose there exists a finite subset Λ_0 of Λ such that*

$$\beta = \sum_{\vec{u}\in\Lambda_0} C_{\vec{u}} s^{\vec{u}}, \tag{2-18}$$

where $0 \neq C_{\vec{u}} \in \mathbb{Z}[q^{\pm 1}]$ for any $\vec{u} \in \Lambda_0$. Then, we have $\mathrm{s}(\vec{u}) \leq m_1 + \cdots + m_6$ for any $\vec{u} \in \Lambda_0$.

Consider $\Lambda' := \{\vec{k} \in \Lambda \mid k_{23} = 0\}$. In this paragraph, we will show

$$\{s^{\vec{k}} \mid \vec{k} \in \Lambda'\} \subset \mathscr{S}_q(\partial H_3)_\gamma \cdot X \tag{2-19}$$

using induction on $\mathrm{s}(\vec{k})$, defined in (2-1). Trivially $s^{\vec{k}} = \varnothing \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X$ when $\mathrm{s}(\vec{k}) = 0$. Suppose

$$s^{\vec{k}} \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X \quad \text{when } \mathrm{s}(\vec{k}) < \mathrm{s}(\vec{u}) = l \text{ and } \vec{k} \in \Lambda', \tag{2-20}$$

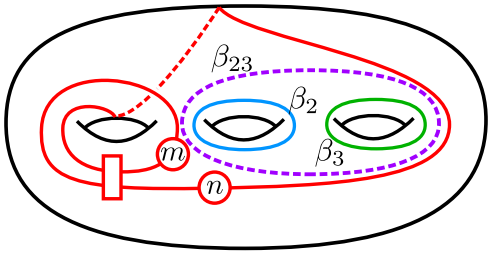


Figure 11: The red curve is the gluing curve γ. The blue curve is β_2 the green curve is β_3, and the purple curve is β_{23}. Here, β_{23} is on the back side.

where $\vec{u} = (u_1, u_2, u_3, u_{12}, u_{13}, 0, u_{123}) \in \Lambda'$. We have

$$\beta_{12}^{u_{12}} \cdot (s_1^{u_1} s_2^{u_2} s_3^{u_3} s_{13}^{u_{13}} s_{123}^{u_{123}}) \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X.$$

We can regard $\beta_{12}^{u_{12}} \cdot (s_1^{u_1} s_2^{u_2} s_3^{u_3} s_{13}^{u_{13}} s_{123}^{u_{123}})$ as a diagram in Σ_4^0 with crossings such that it intersects with $C_1 \cup \cdots \cup C_6$ at $\mathrm{s}(\vec{u})$ points. Then Lemma 2.16 implies

$$\beta_{12}^{u_{12}} \cdot (s_1^{u_1} s_2^{u_2} s_3^{u_3} s_{13}^{u_{13}} s_{123}^{u_{123}}) = \sum_{\substack{\vec{v}\in\Lambda' \\ \mathrm{s}(\vec{v})=\mathrm{s}(\vec{u})}} C_{\vec{u},\vec{v}} s^{\vec{v}} + \sum_{\substack{\vec{a}\in\Lambda' \\ \mathrm{s}(\vec{a})<\mathrm{s}(\vec{u})}} C_{\vec{u},\vec{a}} s^{\vec{a}} + \sum_{\vec{b}\in\Lambda\setminus\Lambda'} C_{\vec{u},\vec{b}} s^{\vec{b}}. \tag{2-21}$$

The previous argument implies $\sum_{\vec{b}\in\Lambda\setminus\Lambda'} C_{\vec{u},\vec{b}} s^{\vec{b}} \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X$. The assumption (2-20) implies $\sum_{\vec{a}\in\Lambda',\mathrm{s}(\vec{a})<\mathrm{s}(\vec{u})} C_{\vec{u},\vec{a}} s^{\vec{a}} \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X$. Thus we have

$$\sum_{\substack{\vec{v}\in\Lambda' \\ \mathrm{s}(\vec{v})=\mathrm{s}(\vec{u})}} C_{\vec{u},\vec{v}} s^{\vec{v}} \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X.$$

By substituting $q = 1$, equation (2-21) will turn into $\beta_{12}^{u_{12}} \cdot (s_1^{u_1} s_2^{u_2} s_3^{u_3} s_{23}^{u_{23}} s_{123}^{u_{123}}) = s^{\vec{u}}$. Then the square matrix $(C_{\vec{u},\vec{v}})_{\vec{u}\in\Lambda',\mathrm{s}(\vec{u})=l;\,\vec{v}\in\Lambda',\mathrm{s}(\vec{v})=l}$ is the identity matrix by substituting $q = 1$. This implies that $(C_{\vec{u},\vec{v}})_{\vec{u}\in\Lambda',\mathrm{s}(\vec{u})=l;\,\vec{v}\in\Lambda',\mathrm{s}(\vec{v})=l}$ is an invertible matrix in $\mathbb{Q}(q)$. Thus we have $s^{\vec{u}} \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X$ for $\vec{u} \in \Lambda', \mathrm{s}(\vec{u}) = l$.

From the previous argument, $s^{\vec{v}} \in \mathscr{S}_q(\partial H_3)_\gamma \cdot X$ for any $\vec{v} \in \Lambda$, i.e., $\iota_*(\mathscr{S}_q(\partial H_3)_\gamma \cdot X) = \mathscr{S}_q(H_3^\gamma)$. □

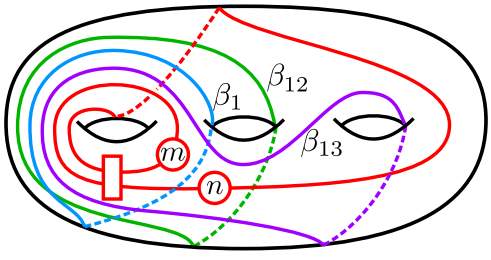


Figure 12: The red curve is the gluing curve γ. The blue curve is β_1, the green curve is β_{12}, and the purple curve is β_{13}.

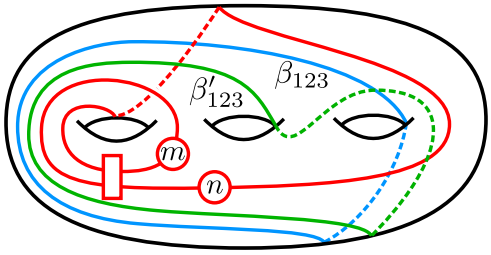


Figure 13: The red curve is the gluing curve γ. The blue curve is β_{123} and the green curve is β'_{123}.

Remark 2.17 When m and n are coprime, using Seifert–van Kampen theorem, we have

$$\pi_1(H_3^\gamma) \cong \langle a, b, c \mid a^{m+n} b^n \rangle$$

for $\mathrm{DT}(\gamma) = (n+m, m, 0, n, n, m+2n, t, 0, 0, 0, 0, 0)$ in the second case in Case 2 of Theorem 2.13. This implies that $H_3^\gamma \ncong H_g$ for any g and H_3^γ is a nontrivial example with a boundary component of genus 2. Hence the higher genus case is not covered by [2; 12; 13; 14; 15].

3 Proofs of lemmas

For any $\vec{k} \in \Lambda$, recall that $\deg(s^{\vec{k}})$ is defined to be $\mathrm{s}'(\vec{k})$; see (2-1).

Proof of Lemma 2.14 Fix $\vec{k} = (k_1, k_2, k_3, k_{12}, k_{13}, k_{23}, k_{123}) \in \mathbb{N}^7$. From Lemma 2.9, there is $\Lambda_0 \subset \Lambda$ satisfying the assumption. Put $k = \max\{\deg(s^{\vec{u}}) \mid \vec{u} \in \Lambda_0\}$. We will show $k \le \mathrm{s}'(\vec{k})$. Assume $k > \mathrm{s}'(\vec{k})$ on the contrary. There is $\vec{u_0} \in \Lambda_0$ such that $\deg(s^{\vec{u_0}}) = k$ and $\mathrm{s}(\vec{u_0}) = \max\{\mathrm{s}(\vec{u}) \mid \deg(s^{\vec{u}}) = k, \vec{u} \in \Lambda_0\}$, where $\mathrm{s}(\vec{u})$ was defined in (2-1).

For any $\vec{u} = (u_1, u_2, u_3, u_{12}, u_{13}, u_{23}, u_{123}) \in \Lambda_0$, we have

$$s^{\vec{u}} = \sum_{\gamma \in \mathrm{Multi}(\vec{u})} C_{\gamma,\vec{u}} \gamma,$$

where $\mathrm{Multi}(\vec{u})$ is a finite subset of the set of isotopy classes of essential multicurves in Σ_0^4 and $C_{\vec{u},\gamma} \in \mathbb{Z}[q^{\pm}] \setminus \{0\}$. The following statement comes from, e.g., [1, Theorem 3.2]. Using the symbols (2-2) and (2-3), there is $\gamma_{\vec{u}} \in \mathrm{Multi}(\vec{u})$ such that $\vec{m}(\gamma_{\vec{u}})$ is equal to

$$u_1\vec{m}(s_1) + u_2\vec{m}(s_2) + u_3\vec{m}(s_3) + u_{12}\vec{m}(s_{12}) + u_{13}\vec{m}(s_{13}) + u_{23}\vec{m}(s_{23}) + u_{123}\vec{m}(s_{123}) \tag{3-1}$$

and, if $\gamma_{\vec{u}} \ne \gamma \in \mathrm{Multi}(\vec{u})$ then $\mathrm{sum}(\gamma) < \mathrm{sum}(\gamma_{\vec{u}})$. Note that $\mathrm{s}(\vec{u}) = \mathrm{sum}(\gamma_{\vec{u}})$.

Suppose $\vec{u} \in \Lambda_0$ such that $\deg(s^{\vec{u}}) < k$. For any $\gamma \in \mathrm{Multi}(\vec{u})$, $m_i(\gamma) \le m_i(\gamma_{\vec{u}})$ $(i = 1, \dots, 6)$. We also have

$$m_1(\gamma_{\vec{u}}) + m_2(\gamma_{\vec{u}}) + m_4(\gamma_{\vec{u}}) + m_5(\gamma_{\vec{u}}) \le \deg(s^{\vec{u}}) < k = m_1(\gamma_{\vec{u_0}}) + m_2(\gamma_{\vec{u_0}}) + m_4(\gamma_{\vec{u_0}}) + m_5(\gamma_{\vec{u_0}}).$$

Hence $\gamma \ne \gamma_{\vec{u_0}}$ for any $\gamma \in \mathrm{Multi}(\vec{u})$.

Suppose $\vec{u} \in \Lambda_0$ such that $\deg(s^{\vec{u}}) = k$ and $\mathrm{s}(\vec{u}) < \mathrm{s}(\vec{u_0})$. Trivially, $\gamma \ne \gamma_{\vec{u_0}}$ for any $\gamma \in \mathrm{Multi}(\vec{u})$.

Suppose $\vec{u} \in \Lambda_0$ such that $\deg(s^{\vec{u}}) = k$, $\mathrm{s}(\vec{u}) = \mathrm{s}(\vec{u_0})$, and $\vec{u} \neq \vec{u_0}$. Lemma 2.11 implies $\gamma_{\vec{u}} \neq \gamma_{\vec{u_0}}$. For any $\gamma_{\vec{u}} \neq \gamma \in \mathrm{Multi}(\vec{u})$, we have $\mathrm{sum}(\gamma) < \mathrm{sum}(\gamma_{\vec{u}}) = \mathrm{s}(\vec{u}) = \mathrm{s}(\vec{u_0}) = \mathrm{sum}(\gamma_{\vec{u_0}})$, which shows $\gamma \neq \gamma_{\vec{u_0}}$.

Now we expand (2-11) as

$$\alpha_{123}^{k_{123}}\alpha_{12}^{k_{12}}\alpha_{13}^{k_{13}}\alpha_{23}^{k_{23}} \cdot (s_1^{k_1}s_2^{k_2}s_3^{k_3}) = \sum_{\vec{u}\in\Lambda_0} C_{\vec{u}}s^{\vec{u}} = \sum_{\gamma\in\mathrm{Multi}(\vec{k})} C_{\vec{k},\gamma}\gamma,$$

where $\mathrm{Multi}(\vec{k})$ is a finite subset of the set of isotopy classes of essential multicurves on Σ_0^4 and $C_{\vec{k},\gamma} \neq 0$ for any $\gamma \in \mathrm{Multi}(\vec{k})$. Then the above argument implies that the most right-hand side contains $\gamma_{\vec{u_0}}$ and its coefficient is nonzero. Since the geometric intersection number does not increase when we resolve crossings, we have

$$m_1(\gamma) + m_2(\gamma) + m_4(\gamma) + m_5(\gamma) \leq \mathrm{s}'(\vec{k})$$

for any $\gamma \in \mathrm{Multi}(\vec{k})$. This contradicts the assumption

$$\deg(s^{\vec{u_0}}) = m_1(\gamma_{\vec{u_0}}) + m_2(\gamma_{\vec{u_0}}) + m_4(\gamma_{\vec{u_0}}) + m_5(\gamma_{\vec{u_0}}) > \mathrm{s}'(\vec{k}). \qquad \square$$

Lemma 3.1 *Suppose $q = 1$. For any two $\vec{u}, \vec{v} \in \Lambda$, $s^{\vec{u}}s^{\vec{v}}$ is a linear sum of basis elements in $\{s^{\vec{k}} \mid \vec{k} \in \Lambda\}$ with degree less than or equal to $\deg(s^{\vec{u}}) + \deg(s^{\vec{v}})$.*

Before we prove Lemma 3.1, we show the following lemma.

Lemma 3.2 *Suppose $q = 1$. For any $k_{12}, k_{13}, k_{23} \in \mathbb{N}$, we have*

$$s_{12}^{k_{12}}s_{13}^{k_{13}}s_{23}^{k_{23}} = \sum_{\vec{u}\in\Lambda_0} C_{\vec{u}}s^{\vec{u}}, \tag{3-2}$$

where Λ_0 is a finite subset of Λ such that $\deg(s^{\vec{u}}) \leq 2k_{12} + 3(k_{13} + k_{23})$ for any $\vec{u} \in \Lambda_0$.

Proof If $k_{12}k_{13}k_{23} = 0$, then since $s_{12}^{k_{12}}s_{13}^{k_{13}}s_{23}^{k_{23}}$ is a basis element, $s_{12}^{k_{12}}s_{13}^{k_{13}}s_{23}^{k_{23}}$ satisfies the claim.

If $k_{12}k_{13}k_{23} \neq 0$ then we apply Lemma 2.10 for $s_{12}^{k_{12}-1}s_{13}^{k_{13}-1}s_{23}^{k_{23}-1}s_{12}s_{13}s_{23}$. Since each term on the right-hand side of the equation in Lemma 2.10 has the degree no greater than $\deg(s_{12}s_{13}s_{23})$. By applying Lemma 2.10 repeatedly until we have (3-2), we can conclude that $s_{12}^{k_{12}}s_{13}^{k_{13}}s_{23}^{k_{23}}$ is a linear sum of basis elements whose degrees are not greater than $2k_{12} + 3(k_{13} + k_{23})$. $\square$

Proof of Lemma 3.1 If $s^{\vec{u}}s^{\vec{v}}$ does not contain $s_{12}s_{13}s_{23}$, the claim is trivial. If $s^{\vec{u}}s^{\vec{v}}$ contains $s_{12}s_{13}s_{23}$, the claim follows from Lemma 3.2. $\square$

Proof of Lemma 2.15 To prove the claim, we use mathematical induction on k_{12}. By applying Lemmas 2.10 and 3.1 to $s_{12}s_{13}^{k_{13}}s_{23}^{k_{23}}$, it is equal to

$$\begin{aligned}&(s_{13}s_2s_{123} + s_{23}s_1s_{123} + \text{terms with degrees less than } 8)s_{13}^{k_{13}-1}s_{23}^{k_{23}-1}\\ &\qquad = s_{13}^{k_{13}}s_{23}^{k_{23}-1}s_2s_{123} + s_{13}^{k_{13}-1}s_{23}^{k_{23}}s_1s_{123} + (\text{terms with degrees less than } 2 + 3(k_{13} + k_{23})).\end{aligned}$$

Thus the claim holds for $s_{12}s_{13}^{k_{13}}s_{23}^{k_{23}}$.

Fix $k \geq 2$ and suppose the claim holds for any $k_{12} < k$. Then

$$\begin{aligned} s_{12}^k s_{13}^{k_{13}} s_{23}^{k_{23}} &= s_{12}\Big(\sum_{\vec{u}\in\Lambda'} C_{\vec{u}} s^{\vec{u}} + (\text{terms with degrees less than } 2(k-1)+3(k_{13}+k_{23}))\Big) \\ &= \sum_{\vec{u}\in\Lambda'} C_{\vec{u}} s_{12} s^{\vec{u}} + (\text{terms with degrees less than } 2k+3(k_{13}+k_{23})), \end{aligned}$$

where Λ' is a finite subset of Λ such that, for any $\vec{u} = (u_1, u_2, u_3, u_{12}, u_{13}, u_{23}, u_{123}) \in \Lambda'$, we have $\deg(s^{\vec{u}}) = 2(k-1)+3(k_{13}+k_{23})$ and $u_{13}u_{23} < k_{13}k_{23}$. For $\vec{u} \in \Lambda'$, if $u_{13}u_{23} = 0$ then $s_{12}s^{\vec{u}}$ is a basis element and $u_{13}u_{23} = 0 < k_{13}k_{23}$. If $u_{13}u_{23} \neq 0$, we have

$$\begin{aligned} s_{12}s^{\vec{u}} &= s_1^{u_1} s_2^{u_2} s_3^{u_3} s_{123}^{u_{123}} s_{12} s_{13}^{u_{13}} s_{23}^{u_{23}} \\ &= s_1^{u_1} s_2^{u_2} s_3^{u_3} s_{123}^{u_{123}} \Big(\sum_{\vec{u}\in\Lambda''} C_{\vec{v}} s^{\vec{v}} + (\text{terms with degrees less than } 2+3(u_{13}+u_{23}))\Big), \end{aligned}$$

where Λ'' is a finite subset of Λ such that, for any $\vec{v} = (v_1, v_2, v_3, v_{12}, v_{13}, v_{23}, v_{123}) \in \Lambda''$, we have

$$\deg(s^{\vec{v}}) = 2+3(u_{13}+u_{23}) \quad \text{and} \quad v_{13}v_{23} < u_{13}u_{23} < k_{13}k_{23}.$$

Then

$$s_{12}s^{\vec{u}} = \sum_{\vec{u}\in\Lambda''} C_{\vec{v}} s_1^{u_1} s_2^{u_2} s_3^{u_3} s_{123}^{u_{123}} s^{\vec{v}} + (\text{terms with degrees less than } 2+\deg(s^{\vec{u}})).$$

Hence the claim holds for $k_{12} = k$. □

Proof of Lemma 2.16 Assume that $\mathrm{s}(\vec{u}) > m_1 + \cdots + m_6$ for some $\vec{u} \in \Lambda_0$ on the contrary. Then

$$m = \max\{\mathrm{s}(\vec{u}) \mid \vec{u} \in \Lambda_0\} > m_1 + \cdots + m_6.$$

As in the proof of Lemma 2.14, for any $\vec{u} = (u_1, u_2, u_3, u_{12}, u_{13}, u_{23}, u_{123}) \in \Lambda_0$, we have

$$s^{\vec{u}} = \sum_{\gamma\in\mathrm{Multi}(\vec{u})} C_{\gamma,\vec{u}}\gamma.$$

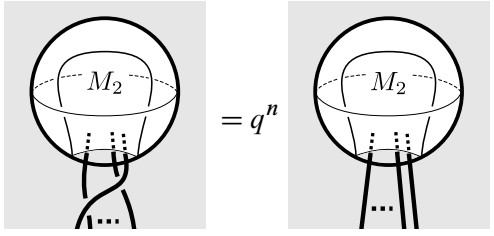
There is $\gamma_{\vec{u}} \in \mathrm{Multi}(\vec{u})$ such that $\vec{m}(\gamma_{\vec{u}})$ is equal to the formula in (3-1). Take $\vec{u}_0 \in \Lambda_0$ such that $\mathrm{s}(\vec{u}_0) = m$.

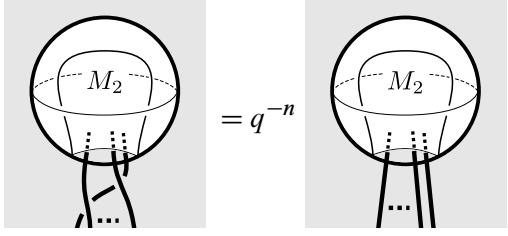
For any $\vec{v} \in \Lambda_0$ such that $\mathrm{s}(\vec{v}) < m$, we have $\mathrm{s}(\gamma) < \mathrm{s}(\gamma_{\vec{u}_0})$ for $\gamma \in \mathrm{Multi}(\vec{v})$. Thus $\gamma \neq \gamma_{\vec{u}_0}$. For any $\vec{v} \in \Lambda_0$ such that $\mathrm{s}(\vec{v}) = m$, we have $\mathrm{s}(\gamma) < \mathrm{s}(\gamma_{\vec{u}_0})$ for $\gamma \in \mathrm{Multi}(\vec{v}) \setminus \{\gamma_{\vec{v}}\}$. Thus $\gamma \neq \gamma_{\vec{u}_0}$ for $\gamma \in \mathrm{Multi}(\vec{v}) \setminus \{\gamma_{\vec{v}}\}$. Lemma 2.11 implies $\gamma_{\vec{v}} \neq \gamma_{\vec{u}_0}$ for $\vec{v} \in \Lambda_0 \setminus \{\vec{u}_0\}$. Thus if we present $\sum_{\vec{u}\in\Lambda_0} C_{\vec{u}} s^{\vec{u}}$ as a linear sum of essential multicurves, it contains $\gamma_{\vec{u}_0}$.

For a diagram α, by resolving its crossings, α can be presented as a linear sum of essential multicurves. This implies that this linear sum does not contain $\gamma_{\vec{u}_0}$ since the intersection number between $\gamma_{\vec{u}_0}$ and $C_1 \cup \cdots \cup C_6$ is strictly greater than $m_1 + \cdots + m_6$. Then we have a contradiction. □

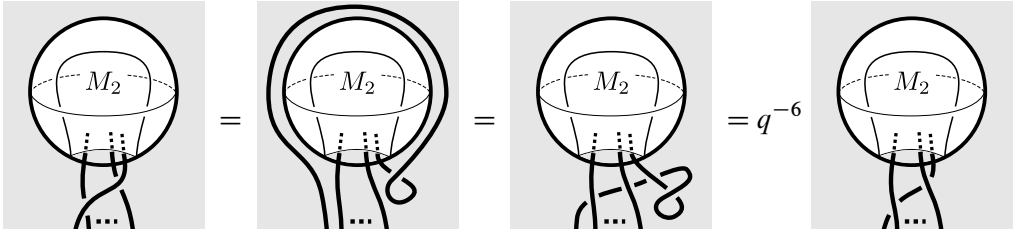
Appendix: Sphere sliding

We recall the sphere sliding given in [11]. The *degree* of a skein in $M_1 \#_{\boldsymbol{D}} M_2$ is the geometric intersection number with the skein and D_1 (equivalently that of the skein and D_2), illustrated here:

(A-1) 

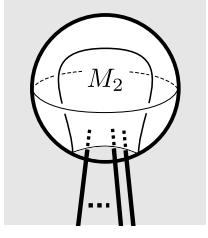
(A-2) 

where the shaded part outside of the sphere is the inside of $M_1^{C_1}$ and the shaded part on the sphere is the attaching region D_1. We also have

(A-3) 

where the last equality follows from  $= -q^{-3}$. Note that a similar idea is originally given in [11], where we call the technique the *sphere sliding*.

From (A-3) with (A-1) and (A-2),

(A-4) 

$= \dfrac{1}{q^n - q^{-n-6}}$(lower degree terms).

Note that (A-4) implies the degree of a framed link in $M_1 \#_{\boldsymbol{D}} M_2$ can decrease, where the degree of a framed link in $M_1 \#_{\boldsymbol{D}} M_2$ is the geometric intersection number of the framed link and the properly embedded disk D. Since the degree is even, one can apply similar way repeatedly until all the degrees are 0.

Acknowledgements

The authors are grateful to Thang Lê and Roland van der Veen for valuable comments. Karuo was partially supported by JSPS KAKENHI Grant Numbers JP22K20342, JP23K12976. Wang was supported by the NTU research scholarship and the Ph.D. scholarship from the University of Groningen.

References

[1] **N Abdiel**, **C Frohman**, *The localized skein algebra is Frobenius*, Algebr. Geom. Topol. 17:6 (2017) 3341–3373 MR

[2] **J R Aranda**, **N Ferguson**, *Generating sets for the Kauffman skein module of a family of Seifert fibered spaces*, New York J. Math. 28 (2022) 44–68 MR

[3] **J W Barrett**, *Skein spaces and spin structures*, Math. Proc. Cambridge Philos. Soc. 126:2 (1999) 267–275 MR

[4] **D Bullock**, *Rings of* $\mathrm{SL}_2(\mathbb{C})$*-characters and the Kauffman bracket skein module*, Comment. Math. Helv. 72:4 (1997) 521–542 MR

[5] **D Bullock**, **J H Przytycki**, *Multiplicative structure of Kauffman bracket skein module quantizations*, Proc. Amer. Math. Soc. 128:3 (2000) 923–931 MR

[6] **A Carrega**, *Nine generators of the skein space of the* 3*-torus*, Algebr. Geom. Topol. 17:6 (2017) 3449–3460 MR

[7] **R Detcherry**, *Infinite families of hyperbolic* 3*-manifolds with finite-dimensional skein modules*, J. Lond. Math. Soc. (2) 103:4 (2021) 1363–1376 MR

[8] **P M Gilmer**, **J M Harris**, *On the Kauffman bracket skein module of the quaternionic manifold*, J. Knot Theory Ramifications 16:1 (2007) 103–125 MR

[9] **S Gunningham**, **D Jordan**, **P Safronov**, *The finiteness conjecture for skein modules*, Invent. Math. 232:1 (2023) 301–363 MR

[10] **J Hoste**, **J H Przytycki**, *The* $(2,\infty)$*-skein module of lens spaces*; *a generalization of the Jones polynomial*, J. Knot Theory Ramifications 2:3 (1993) 321–333 MR

[11] **J Hoste**, **J H Przytycki**, *The Kauffman bracket skein module of* $S^1 \times S^2$, Math. Z. 220:1 (1995) 65–73 MR

[12] **T T Q Lê**, *The colored Jones polynomial and the A-polynomial of knots*, Adv. Math. 207:2 (2006) 782–804 MR

[13] **T T Q Le**, **A T Tran**, *The Kauffman bracket skein module of two-bridge links*, Proc. Amer. Math. Soc. 142:3 (2014) 1045–1056 MR

[14] **T T Q Le**, **A T Tran**, *On the AJ conjecture for knots*, Indiana Univ. Math. J. 64:4 (2015) 1103–1151 MR

[15] **J Marché**, *The skein module of torus knots*, Quantum Topol. 1:4 (2010) 413–421 MR

[16] **M Mroczkowski**, *Kauffman bracket skein module of a family of prism manifolds*, J. Knot Theory Ramifications 20:1 (2011) 159–170 MR

[17] **R C Penner**, **J L Harer**, *Combinatorics of train tracks*, Annals of Mathematics Studies 125, Princeton Univ. Press (1992) MR

[18] **J H Przytycki**, *Kauffman bracket skein module of a connected sum of* 3*-manifolds*, Manuscripta Math. 101:2 (2000) 199–207 MR

[19] **J H Przytycki**, **A S Sikora**, *On skein algebras and* $\mathrm{Sl}_2(\mathbb{C})$*-character varieties*, Topology 39:1 (2000) 115–148 MR

[20] **T Saito**, **M Scharlemann**, **J Schultens**, *Lecture notes on generalized Heegaard splittings* (2005) arXiv math/0504167

HIROAKI KARUO hiroaki.karuo@gakushuin.ac.jp
Department of Mathematics, Gakushuin University, Tokyo, Japan

ZHIHAO WANG zhihaowang@kias.re.kr
School of Mathematics, Korea Institute for Advanced Study (KIAS), 85 Hoegi-ro, Dongdaemun-gu, Seoul 02455, South Korea

Received: September 2, 2024 Revised: December 7, 2024

msp

Algebraic & Geometric Topology **26**:3 (2026) 1115–1154
DOI: 10.2140/agt.2026.26.1115
Published: April 1, 2026

L-spaces, taut foliations and fibred hyperbolic two-bridge links

DIEGO SANTORO

We prove that if M is a rational homology sphere that is Dehn surgery on a fibred hyperbolic two-bridge link, then M is not an L-space if and only if M supports a co-orientable taut foliation. As a corollary we show that if K' is obtained by a nontrivial knot K as a result of an operation called *two-bridge replacement*, then all nonmeridional surgeries on K' support co-orientable taut foliations. This operation generalises Whitehead doubling and as a particular case we deduce that all nonmeridional surgeries on Whitehead doubles of a nontrivial knot support co-orientable taut foliations.

1 Introduction

In recent years the field of low-dimensional topology has seen a growing interest in the study of the so-called L-space conjecture. This conjecture predicts that the following notions of "complexity" are all equivalent:

Conjecture 1.1 (L-space conjecture) *For an irreducible oriented rational homology 3-sphere M, the following are equivalent*:

(1) *M supports a co-oriented taut foliation.*

(2) *M is not an L-space, i.e., its Heegaard Floer homology is not minimal.*

(3) *M is left-orderable, i.e., $\pi_1(M)$ is left-orderable.*

The equivalence between (1) and (2) was conjectured by Juhász [27], while the equivalence between (2) and (3) was conjectured by Boyer, Gordon and Watson [4]. This conjecture predicts strong connections among geometric, dynamical, Floer homological, and algebraic properties of 3-manifolds. Despite its boldness, as a result of [3; 4; 5; 8; 14; 22; 40] it is now known that the conjecture holds for all the graph manifolds, i.e., the manifolds whose JSJ decomposition includes only Seifert fibred pieces. Moreover the results of Oszváth–Szabó [47], Bowden [2] and Kazez–Roberts [30] imply that in general manifolds supporting co-orientable taut foliations are not L-spaces.

A natural way to investigate this conjecture is by using Dehn surgery descriptions of 3-manifolds. For instance, it is known that if a nontrivial knot K in S^3 has a positive surgery that is an L-space, then K is prime [31], fibred [20; 45] and strongly quasipositive [23]. Moreover, the r-framed surgery on such a knot K is an L-space if and only if $r \geq 2g(K)-1$, where $g(K)$ denotes the genus of K [33]. Taut foliations on manifolds obtained as surgery on knots in S^3 are constructed, for example, in [11; 12; 32; 51; 52] and it is possible to prove the left-orderability of some of these manifolds by determining which of these foliations have vanishing Euler class, as done in [25]. Another approach to study the left-orderability of surgeries on knots is via representation-theoretic methods, as presented in [9; 13].

MSC2020: primary 57M50, 57K10, 57K18, 57K32, 57R30; secondary 57K10.

When it comes to investigate surgeries on links, the story becomes more mysterious. For instance, there is no generalisation of the result of [33] we cited in the previous paragraph — even if it holds in some cases, see, for example, [53, Lemma 2.6] — and links admitting L-space surgeries need not to be fibred [41, Example 3.9] nor quasipositive [7, Proposition 1.5]. Concerning foliations, Kalelkar and Roberts [28] constructed co-orientable taut foliations on some fillings of 3-manifolds that fibre over the circle and in particular their methods can also be applied to surgeries on fibred links. In [53], taut foliations on all the surgeries on the Whitehead link that are not L-spaces are constructed.

In this paper we study the L-space conjecture for manifolds that can be obtained as surgery on two-bridge links. A two-bridge link is either hyperbolic or isotopic to the $(2n, 2)$ torus link, for some integer n. In the latter case the exterior is a Seifert fibred manifold and since the conjecture has been proven for graph manifolds — in particular, see [4; 14; 26; 39; 44] for the case of Seifert fibred manifolds — we focus our study on hyperbolic two-bridge links. The main theorem of this paper is the following:

Theorem 1.2 *Let L be a fibred hyperbolic two-bridge link and let M be a manifold obtained as Dehn surgery on L. Then M admits a co-orientable taut foliation if and only if M is not an L-space.*

Remark 1.3 In contrast to the case of knots, the property of being fibred for a link depends on the choice of an orientation of the link. This happens, for instance, in the case of the $(2n, 2)$ torus link for $n > 1$, see, for example, [1, Example 3.1]. On the other hand, changing orientations of the components of L has no effects on the study of the L-space conjecture for the surgeries on L. For this reason we will consider links as unoriented and say that a link is fibred if there exists an orientation for which it is a fibred link.

Remark 1.4 Theorem 1.2 is a result about the exterior of the links. Links are not uniquely determined by their complement, hence it can be a priori possible that many nonisotopic hyperbolic two-bridge links share the same exterior. However it follows from [43, Theorem 1.4] that the exteriors of hyperbolic two-bridge links (with two components) are not even commensurable.

We will be able to completely determine for each fibred hyperbolic two-bridge link L the set of surgeries on L that are L-spaces. We denote by $\mathcal{L}(L)$ the set of slopes on L that produce L-spaces. Recall that since L is a link in S^3 there is a canonical identification between the set of slopes on L and $\overline{\mathbb{Q}} \times \overline{\mathbb{Q}}$, where $\overline{\mathbb{Q}} = \mathbb{Q} \cup \{\infty\}$, obtained by considering on each component of L its canonical meridian and longitude basis. Also, notice that we can reduce our study to the rational surgeries. In fact the components of two-bridge links are unknotted, so when one of the two surgery coefficients is infinite the only rational homology spheres that can be obtained are S^3 and lens spaces. We will prove the following proposition, where the link L_n is shown in Figure 1, for $n \geq 1$.

Proposition 1.5 *Let L be a fibred hyperbolic two-bridge link.*

- *If L is isotopic as an unoriented link to L_n, then $\mathcal{L}(L) \cap \mathbb{Q}^2 = \big([n, +\infty) \times [n, +\infty)\big) \cap \mathbb{Q}^2$.*
- *If L is isotopic as an unoriented link to the mirror of L_n, then $\mathcal{L}(L) \cap \mathbb{Q}^2 = \big((-\infty, -n] \times (-\infty, -n]\big) \cap \mathbb{Q}^2$.*
- *If L is not isotopic as an unoriented link to any of the links L_n or their mirrors, then $\mathcal{L}(L) \cap \mathbb{Q}^2 = \varnothing$.*

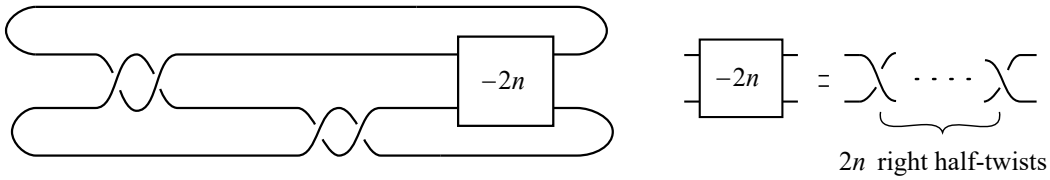


Figure 1: The link L_n.

We observe that L_1 is the Whitehead link. The L-space conjecture for surgeries on the Whitehead link was studied by Santoro [53]. As a consequence of the previous proposition we have the following Dehn surgery characterisation of the Whitehead link:

Corollary 1.6 *Let L be a fibred hyperbolic two-bridge link and suppose that the $(1,1)$-surgery on L is an L-space. Then L is isotopic, as an unoriented link, to the Whitehead link.*

We observe that all the links $\{L_n\}_{n\geq 1}$ can be obtained as surgery on a 3-component link, see Figure 33 on page 1151. On the other hand we have the following:

Proposition 1.7 *It is not possible to obtain the exteriors of all the hyperbolic fibred two-bridge links as Dehn filling on a fixed cusped hyperbolic manifold N. In particular there exists no hyperbolic link L such that every hyperbolic fibred two-bridge link is surgery on L.*

Proof It follows from the main result of [34] that there exists a family of fibred hyperbolic two-bridge links whose volumes grow to infinity (this is the family of links associated to $L(a_1,\dots,a_n)=L(2,2,\dots,2)$ in the notation introduced in Section 4). Volume decreases under hyperbolic Dehn filling [55], hence we obtain the thesis. □

As two-bridge links have tunnel number one, all surgeries on these links have at most Heegaard genus two. It has been proven by Li [37] that if a closed orientable irreducible three manifold with Heegaard genus two has left-orderable fundamental group, then it admits a co-orientable taut foliation. As a consequence of this result together with Proposition 1.5 we have:

Corollary 1.8 *Let M be obtained as (r_1,r_2)-surgery on the link L_n, with $(r_1,r_2)\in[n,+\infty)\times[n,+\infty)$, and suppose that M is irreducible. Then M is not left-orderable.*

Applications to satellites on knots and links We briefly recall the satellite operation. Suppose that P is a knot inside the standard solid torus $V=\mathbb{D}^2\times S^1$ and assume that P is not contained in a 3-ball of V. Let K be a knot in S^3 and let (μ_K,λ_K) be a meridian-longitude basis of K. Consider the orientation preserving diffeomorphism ϕ between V and a tubular neighbourhood of K mapping, as oriented curves, the meridian μ_V and longitude λ_V of V to μ_K and λ_K, respectively. The image of P under ϕ is a knot S in S^3, called a *satellite* of K. The knot K is called the *companion* of S and the knot P is called the *pattern* of S.

Let L be a fibred hyperbolic two-bridge link, denote by $\mathcal{K}$ one of its component and orient it arbitrarily. Since two-bridge links have unknotted components, the exterior of $\mathcal{K}$ in S^3 is a solid torus V and we can

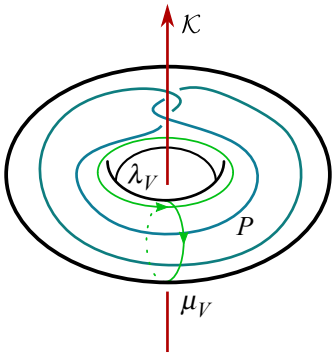


Figure 2: The (positive clasped) Whitehead pattern. The meridian μ_V is given by the longitude of the knot K_0 and the longitude λ_V by its meridian. By considering the mirror of the Whitehead link one obtains the negative clasped Whitehead pattern.

use the other component as pattern P for producing satellite knots. We also fix a meridian-longitude basis for V given by $(\mu_V, \lambda_V) = (\lambda_{\mathcal{K}}, \mu_{\mathcal{K}})$, where $\mu_{\mathcal{K}}$ and $\lambda_{\mathcal{K}}$ are the canonical meridian and longitude of $\mathcal{K}$.

Example 1.9 If L is the Whitehead link we obtain the Whitehead pattern. This is the pattern used to define Whitehead doubles of knots, see Figure 2.

We now define an operation, that we call *two-bridge replacement*, that generalises Whitehead doubling.

Definition 1.10 Let K be a knot in S^3 and let $L = \mathcal{K} \sqcup P$ be a fibred hyperbolic two-bridge link. A knot K' in S^3 is a *two-bridge replacement* of K if K' is a satellite knot of K with pattern P. More generally, if $\mathcal{L} = K_1 \sqcup \cdots \sqcup K_d$ is a link with d components, we say that $\mathcal{L}' = K_1' \sqcup \cdots \sqcup K_d'$ is a *two-bridge replacement* of $\mathcal{L}$ if each knot K_i' is a two-bridge replacement of K_i, for $i = 1, \dots, d$.

Notice that in a two-bridge replacement of a link $\mathcal{L}$ it is allowed to act on different components on $\mathcal{L}$ by two-bridge replacement with different two-bridge links. We also remark that in the definition of two-bridge replacement we ask L to be fibred and hyperbolic.

Recall that the genus of a link $\mathcal{L}$ is the minimal genus of a connected Seifert surface for $\mathcal{L}$, and that if $\mathcal{L}$ is fibred (see Definition 2.3) then a Seifert surface has minimal genus if and only if it is a fibred surface [15, Chapter 1.4]. The proofs of Theorem 1.2 and of the main theorem of [53], together with results from [28; 38], imply the following theorem.

Theorem 1.11 *Let $\mathcal{L}$ be any nontrivial knot or any fibred link with positive genus, and let $\mathcal{L}'$ denote a two-bridge replacement of $\mathcal{L}$. Then all manifolds obtained by doing surgery on each component of $\mathcal{L}'$ along a nonmeridional slope support a co-orientable taut foliation.*

Proof • We first analyse the case where $\mathcal{L}$ is a nontrivial knot. We denote it by K and we denote by K' its two-bridge replacement. We use the notation introduced above, and so we denote by $L = \mathcal{K} \sqcup P$ the fibred hyperbolic two-bridge link used in the definition of two-bridge replacement, and V is the exterior of $\mathcal{K}$ in S^3. Recall that V is a solid torus. We also denote by $E_{K'}$, E_K and E_L the exteriors in S^3 of K', K and L, respectively. Let ϕ be the map from V to a tubular neighbourhood of K given by the definition

of the satellite operation. Our aim is to prove that all nontrivial surgeries on K' support a co-orientable taut foliation. The key observation is that $E_{K'}$ is obtained by gluing E_K and E_L. More precisely we have

$$E_{K'} \cong E_K \cup_\varphi E_L,$$

where φ is the restriction of ϕ to $\partial V \subset \partial E_L$.

In order to construct foliations on $E_{K'}$ we can therefore reduce ourselves to construct foliations on E_K and E_L and to study the gluing map φ. We fix the canonical meridian-longitude basis (μ_K, λ_K) for the knot K and we use it to identify slopes on K with $\overline{\mathbb{Q}}$. By definition ϕ maps the meridian μ_V of V to μ_K and the longitude λ_V of V to some longitude of K, i.e.,

$$\phi(\lambda_\mathcal{K}) = \phi(\mu_V) = \mu_K, \quad \phi(\mu_\mathcal{K}) = \phi(\lambda_V) = l\mu_K + \lambda_K$$

for some integer $l \in \mathbb{Z}$. Given two coprime integers p, q the map ϕ satisfies

$$p\mu_K + q\lambda_K = \phi((p - ql)\lambda_\mathcal{K} + q\mu_\mathcal{K})$$

and therefore its restriction φ acts on the slopes by identifying the slope $\frac{p}{q}$ on K with the slope $\left(\frac{p}{q} - l\right)^{-1}$ on $\mathcal{K}$. Since K is a nontrivial knot, it follows by [38, Theorem 1.1] that there exists an interval $(-a, b)$, where $a, b > 0$, such that for every slope $s \in (-a, b)$ there exists a co-orientable taut foliation on E_K intersecting the boundary torus in a collection of circles of slope s. The slope 0 on K corresponds, via the identification given by φ, to the slope $-\frac{1}{l}$ on $\mathcal{K}$. Hence the interval $(-a, b)$ is identified with a neighbourhood U_1 of $-\frac{1}{l} \subset \overline{\mathbb{Q}}$. The proof of [53, Theorem 1.1] when L is the Whitehead link and the proof of Theorem 1.2 when L is any other fibred hyperbolic two-bridge link imply the following fact: for every integer $l \in \mathbb{Z}$ and every neighbourhood U of $-\frac{1}{l} \subset \overline{\mathbb{Q}}$ there exists a slope $r \in U$ such that for every nonmeridional slope r' on P there is a co-orientable taut foliation on E_L intersecting the boundary tori in circles of slopes r and r', respectively.

By choosing a slope $r \in U_1$ guaranteed by the previous observation we are able to find for each nonmeridional slope r' in $E_{K'}$ taut foliations $\mathcal{F}$ on E_K and $\mathcal{F}'$ on E_L that can be glued along φ to define a co-orientable taut foliation in $E_{K'}$ intersecting the boundary in parallel curves of slope r'. By capping off with meridional discs, these foliations extend to the surgeries on K'.

• When $\mathcal{L} = K_1 \sqcup \cdots \sqcup K_d$ is a fibred link with multiple components and positive genus we can proceed in an analogous way. Let S denote the fibre surface for $\mathcal{L}$. By intersecting S with the boundaries of tubular neighbourhoods of the knots $K_1, \dots, K_d$ we obtain longitudes $\lambda_1^S, \dots, \lambda_d^S$. We use them to define meridian-longitude bases for the components of $\mathcal{L}$ and to identify slopes on the exterior of $\mathcal{L}$ with $\overline{\mathbb{Q}}^d$. It follows by [28, Theorem 1.1] that for every multislope $(r_1, \dots, r_d)$ in a neighbourhood of $0 \in \overline{\mathbb{Q}}^d$ there exists a co-orientable taut foliation in the exterior of $\mathcal{L}$ intersecting the boundary tori in parallel curves of slopes $r_1, \dots, r_d$, respectively. The statement now follows by applying to each component of $\mathcal{L}$ the same reasoning as in the previous case, where we never made use of the fact that λ_K was the canonical longitude of K. □

Two-bridge replacement generalises Whitehead doubling and we emphasise the following corollary.

Corollary 1.12 *Let K be a nontrivial knot and let K' be any Whitehead double of K. Then all nontrivial surgeries on K' support a co-orientable taut foliation.* □

Structure of the paper In Section 2 we recall some notions on two-bridge links and describe some properties of fibred two-bridge links that will be used in the subsequent sections. Section 3 is devoted to the construction of taut foliations and will take most of the paper. In Section 3.1 we introduce branched surfaces and recall some of their basic properties, together with the main result of [36]. In Section 3.2 we recall a general method of constructing branched surfaces in fibred manifolds with boundary. In Section 3.3 we briefly discuss the boundary train tracks of these branched surfaces and from Section 3.4 we start focussing our attention on surgeries on fibred hyperbolic two-bridge links: we subdivide them in four families that we analyse in Sections 3.5–3.8. In Section 4 we recall part of the main result of [50] and then we use it to study the L-space surgeries on the links $\{L_n\}_{n\geq 1}$.

2 Background on two-bridge links

In this section we briefly recall the definition of fibred link and some facts about two-bridge links that we will often use in the paper. We refer to [6] for proofs and details regarding two-bridge links.

A *two-bridge link* can be described by a rational number $\frac{p}{q}$, where p and q are coprime integers, $p>0$, q is odd and $0<|q|<p$, in the following way. We fix a sequence of integers $(a_1,\dots,a_n)$ such that

$$(*)\qquad \frac{p}{q} = a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}}$$

and consider the link defined by the diagram in Figure 3. We denote this link by $L(a_1,\dots,a_n)$.

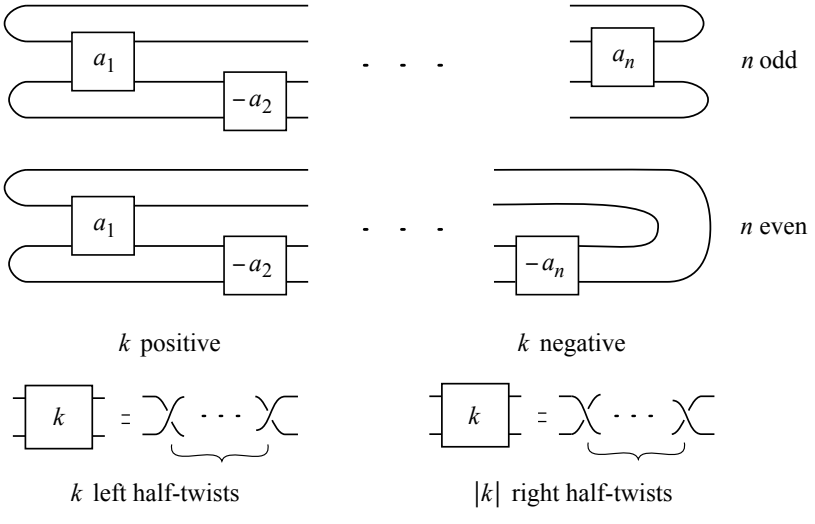


Figure 3: The two-bridge knot or link $L(a_1,\dots,a_n)$.

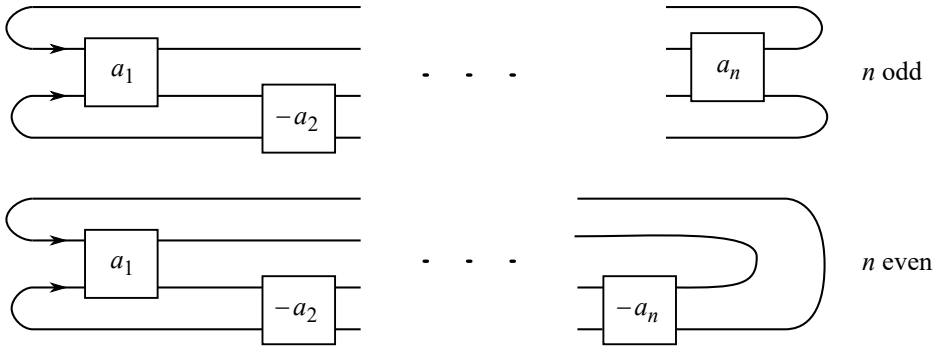


Figure 4: The oriented two-bridge link $L(a_1, \dots, a_n)$.

We are interested in the case when $L(a_1, \dots, a_n)$ has two components. This happens exactly when the fraction $\frac{p}{q}$ has numerator p even [6, Proposition 12.3]. When $L(a_1, \dots, a_n)$ is a link we orient the components as in Figure 4.

A priori it could happen that the isotopy class of the two-bridge link associated to $\frac{p}{q}$ depends on the choice of the continued fraction representation of $\frac{p}{q}$. This is not the case, by the following theorem.

Theorem 2.1 ([54], see also [6]) *Let $L = L(a_1 \dots, a_n)$ and $L' = L(b_1, \dots, b_m)$ be two oriented two-bridge links and let $\frac{p}{q}$ and $\frac{p'}{q'}$ be the rational numbers defined as in $(*)$. Then the links L and L' are isotopic if and only if $p = p'$ and $q' \equiv q^{\pm 1} \bmod 2p$. If $p = p'$ and $q' \equiv q + p \bmod 2p$ or $qq' \equiv 1 + p \bmod 2p$, then L and L' are isotopic after reversing the orientation of one of the components.*

We denote by $b(p, q)$ the two-bridge link associated to the rational number $\frac{p}{q}$. We also recall, for later reference, the following fact [29, Exercise 2.1.16].

Lemma 2.2 *Two-component two-bridge links are symmetric, i.e., there exists an ambient isotopy of S^3 interchanging their components.* □

We are interested in studying fibred hyperbolic two-bridge links. For this reason we recall the definition of fibred link.

Given an oriented surface S with boundary and an orientation-preserving homeomorphism $h : S \to S$ fixing ∂S pointwise, we denote the mapping torus of h by

$$M_h = \frac{S \times [0, 1]}{(h(x), 0) \sim (x, 1)}.$$

We orient $S \times [0, 1]$ as a product and M_h with the orientation induced by $S \times [0, 1]$. We also identify S with its image in M_h via the map

$$S \to S \times \{0\} \subset M_h, \quad x \mapsto (x, 0).$$

The homeomorphism h is called the *monodromy* of M_h.

Definition 2.3 Let L be an oriented link in S^3. We say that L is *fibred* if there exists a Seifert surface S for L, an orientation preserving homeomorphism h of S fixing ∂S pointwise and an orientation preserving homeomorphism

$$\chi : S^3 \setminus \mathrm{int}(N_L) \to M_h,$$

where N_L denotes a tubular neighbourhood of L in S^3, so that

- $\chi_{|S}$ is the inclusion $S \subset M_h$;
- $\chi(m_i) = \{x_i\} \times [0,1]$, where m_i is a meridian for the i-th component of L and $x_i \in \partial S$ is a point.

In the case of fibred links, we refer to the homeomorphism h as the *monodromy of the link*.

The following lemma yields a useful characterisation of fibred hyperbolic two-bridge links.

Lemma 2.4 *Let be L a two-bridge link with two components. Then L is fibred if and only if $L = L(2b_1, \dots, 2b_n)$ as an unoriented link, where $|b_i| = 1$ for all i and n is odd. Moreover L is fibred and hyperbolic if and only if it admits such a description with $(b_1, \dots, b_n) \neq \pm(1, -1, 1, \dots, (-1)^{n-1})$.*

Proof It follows from [29, Exercise 2.1.14] that any two-bridge link L with two components can be written as $L = L(2b_1, \dots, 2b_n)$, where b_i is a nonzero integer and n is odd. Moreover it follows from [18, Proposition 2][1] that L is fibred if and only if we can find such a description with $|b_i| = 1$ for all i. This proves the first part of the lemma. Since two-bridge links are nonsplit, prime, alternating links (see [6]), as a consequence of [42, Corollary 2], a two-bridge link is hyperbolic if and only if is it not a torus link. The only nontrivial torus links that are two-components two-bridge links are the $(2m, 2)$ torus links, with m a nonzero integer, and by using Theorem 2.1 one can see that L is a torus link if and only if any description of L as $L(2b_1, \dots, 2b_n)$ with all $|b_i| = 1$ satisfies $(b_1, \dots, b_n) = \pm(1, -1, 1, \dots, (-1)^{n-1})$. □

Suppose $L = L(2b_1, \dots, 2b_n)$ where $|b_i| = 1$ for all i and $n = 2k + 1$ is odd. Then it is possible to draw an explicit fibre surface S for L. This surface is obtained by starting with the boundary connected sum of k Hopf bands, and then plumbing other $k + 1$ Hopf bands to this surface. The sign of the Hopf bands is determined in a straightforward way from the coefficients $(b_1, \dots, b_n)$. One example is described in Figure 5. We also fix an orientation of S, so that in the figure the positive side is coloured in pink, and this induces an orientation of the link.

From this very easy description of the fibre surface of L we are able to determine the monodromy of L. More precisely, S can be described in a more abstract way as in Figure 6 and the monodromy is given by the diffeomorphism (to be read from right to left)

$$(**) \qquad h = \tau_2^{\varepsilon_2} \tau_4^{\varepsilon_4} \dots \tau_{2k}^{\varepsilon_{2k}} \tau_1^{\varepsilon_1} \tau_3^{\varepsilon_3} \dots \tau_{2k+1}^{\varepsilon_{2k+1}},$$

where τ_i denotes the positive (i.e., the right) Dehn twist along the curve γ_i shown in Figure 6 and

$$\varepsilon_i = \begin{cases} -\mathrm{sgn}(b_i) & \text{when } i \text{ is even,} \\ \mathrm{sgn}(b_i) & \text{when } i \text{ is odd.} \end{cases}$$

[1] The proof presented there is for knots, but the same proof works also for links.

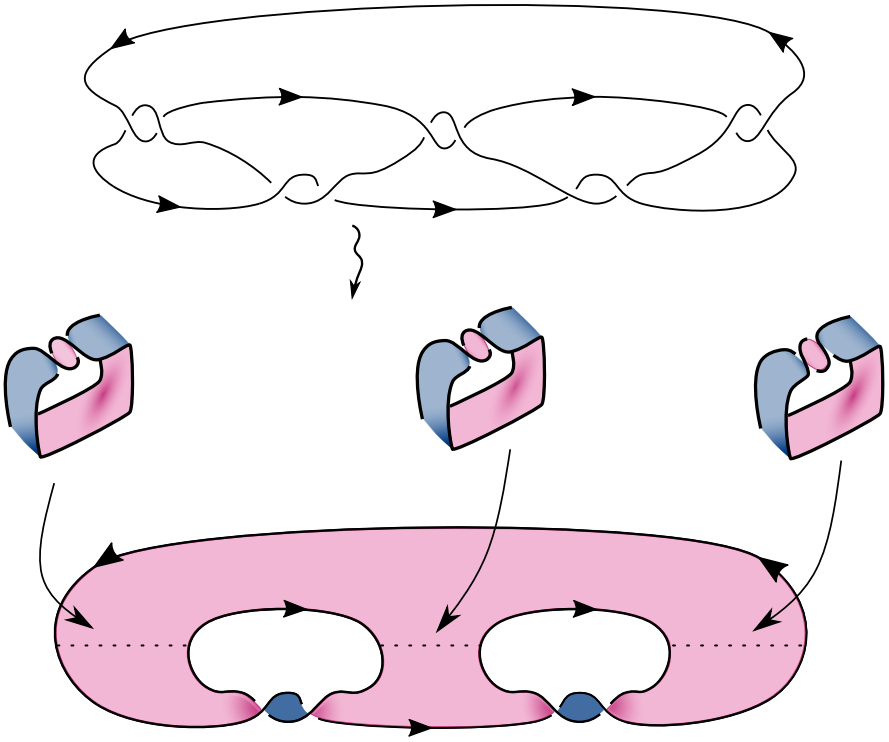


Figure 5: The fibre surface of the link $L(-2,-2,-2,2,2)$. The positive side is coloured in pink.

This follows from the fact that the monodromy of the boundary of a positive (resp. negative) Hopf band is a positive (resp. negative) Dehn twist along its core and from the way the monodromy of a plumbing or a boundary connected sum (or more generally a Murasugi sum) behaves with respect to the monodromies of the summands, see [17, Corollary 1.4].

To ease the exposition of some lemmas in the next section we fix the following notation. With reference to Figure 6 we say that a Dehn twist along one of the curves $\gamma_1, \gamma_3, \dots, \gamma_n$ is a *bridge twist*, and a Dehn twist along one of the curves $\gamma_2, \gamma_4, \dots, \gamma_{n-1}$ is a *river twist*.

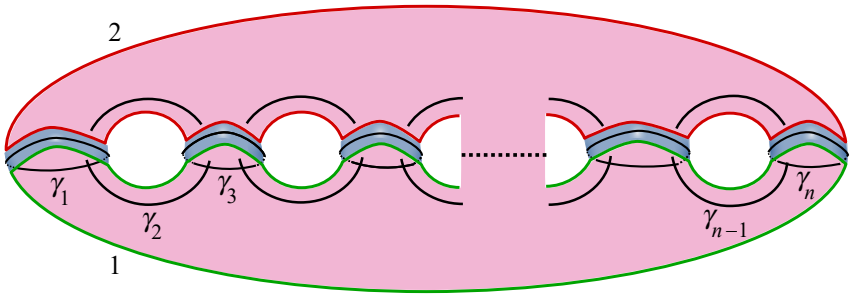


Figure 6: An abstract drawing of the fibre surface S together with the curves γ_i. We have also coloured the two boundary components of S.

3 Taut foliations

In this section we study the existence of taut foliations on the surgeries on fibred hyperbolic two-bridge links, proving the foliation part of Theorem 1.2. Branched surfaces will be our main tool. In Section 3.1 we introduce them and recall some of their basic properties, together with the main result of [36]. In Section 3.2 we recall a general method to construct branched surfaces in fibred manifolds with boundary and in Section 3.3 we discuss their boundary train tracks. In Section 3.4 we focus our attention on surgeries on fibred hyperbolic two-bridge links: we subdivide them in four families with Lemma 3.10 and then study each of these families separately in Sections 3.5–3.8.

3.1 Background

In this and in the next sections we assume familiarity with the basic notions of the theory of train tracks; see [49] for a reference. In the cases of our interest train tracks can also have bigons as complementary regions.

We now recall some basic facts about branched surfaces. We refer to [16; 46] for more details.

Definition 3.1 A *branched surface with boundary* in a 3-manifold M (possibly with boundary) is a closed subset $B \subset M$ that is locally diffeomorphic to one of the models in $\mathbb{R}^3$ of Figure 7(a) or to one of the models in the closed half space of Figure 7(b), where $\partial B := B \cap \partial M$ is represented with a bold line.

Branched surfaces generalise the concept of train tracks from surfaces to 3-manifolds. When the boundary of B is nonempty it defines a train track ∂B in ∂M.

If B is a branched surface it is possible to identify two subsets of B: the *branch locus* and the set of *triple points*. The branch locus is defined as the set of points where B is not locally homeomorphic to a surface. It is self-transverse and intersects itself in double points only. The set of triple points of B can be defined as the points where the branch locus is not locally homeomorphic to an arc. For example, the rightmost model of Figure 7(a) contains a triple point.

The complement of the branch locus in B is a union of connected surfaces. The abstract closures of these surfaces under any path metric on M are called the *branch sectors* of B. Analogously, the complement of the set of the triple points inside the branch locus is a union of 1-dimensional connected manifolds. Moreover, to each of these manifolds we can associate an arrow in B pointing in the direction of the smoothing, as in Figure 7. We call these arrows *cusp directions*.

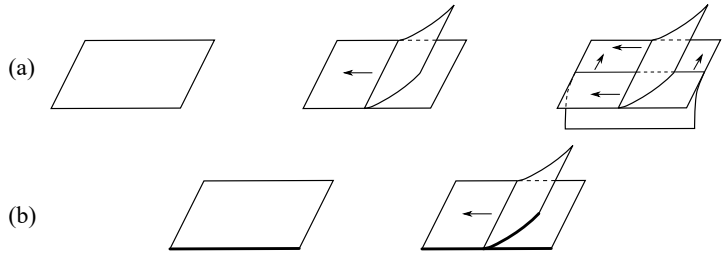


Figure 7: Local models for a branched surface, with cusp directions. The bold regions lie in ∂M.

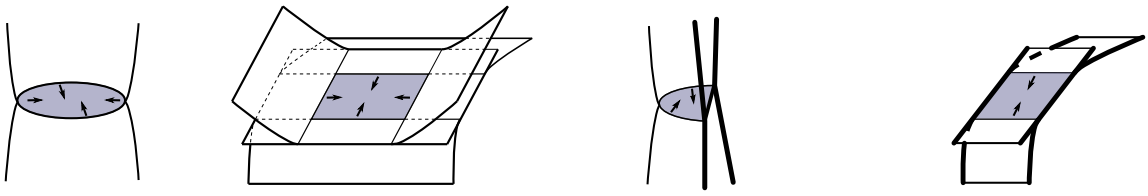


Figure 8: Examples of sink discs.

Li [35] introduced the notion of *sink disc*.

Definition 3.2 Let B be a branched surface in M and let S be a branch sector in B. We say that S is a *sink disc* if S is a disc (possibly with $S \cap \partial M \neq \varnothing$) and the branch direction of any smooth curve or arc in $\partial S \setminus \partial M$ points into S.

In Figure 8 some examples of sink discs are depicted. The bold lines represent the intersection of the branched surface with ∂M. Notice that the intersection $\partial S \cap \partial M$ can also be disconnected.

Li [35] introduced the definition of laminar branched surface and in [36] he generalised this definition to branched surfaces with boundary as follows. For the definition of trivial bubble we refer the reader to [35; 36].

Definition 3.3 [35; 36] Let B be a branched surface in a 3-manifold M. We say that B is *laminar* if B has no trivial bubbles and the following hold:

(1) $\partial_h N_B$ is incompressible and ∂-incompressible in $M \setminus \mathrm{int}(N_B)$, and no component of $\partial_h N_B$ is a sphere or a properly embedded disc in M.

(2) There is no monogon in $M \setminus \mathrm{int}(N_B)$, i.e., no disc $D \subset M \setminus \mathrm{int}(N_B)$ such that $\partial D = D \cap N_B = \alpha \cup \beta$, where α is in an interval fibre of $\partial_v N_B$ and β is an arc in $\partial_h N_B$.

(3) $M \setminus \mathrm{int}(N_B)$ is irreducible and $\partial M \setminus \mathrm{int}(N_B)$ is incompressible in $M \setminus \mathrm{int}(N_B)$.

(4) B contains no Reeb branched surfaces (see [19] for the definition).

(5) B has no sink discs.

The key property of laminar branched surfaces is that they fully carry essential laminations[2] [35] and when the ambient manifold M has torus boundary, they can be used to construct essential laminations in fillings of M, as the following theorem shows.

Theorem 3.4 [36] *Let M be an irreducible and orientable 3-manifold whose boundary is union of k incompressible tori $T_1, \dots, T_k$. Suppose that B is a laminar branched surface in M such that $\partial M \setminus \partial B$ is a union of bigons. Then for any multislope $(s_1, \dots, s_k) \in \overline{\mathbb{Q}}^k$ that is realised by the train track ∂B, if B does not carry a torus that bounds a solid torus in $M(s_1, \dots, s_k)$, there exists an essential lamination Λ in M fully carried by B that intersects ∂M in parallel simple curves of multislope $(s_1, \dots, s_k)$. Moreover this lamination extends to an essential lamination of the filled manifold $M(s_1, \dots, s_k)$.*

[2]For the definition of essential laminations see [19], but we will not need their properties for our purposes.

A train track *realises* a slope r if it can be split into finitely many copies of that slope, see [49]. An equivalent, but more computationally useful definition, is given in Section 3.3.

Remark 3.5 In [36] the statement of the theorem is given for M with connected boundary but, as already observed in [28], if M has multiple boundary components we can split B in a neighbourhood of each boundary tori T_i and the same proof of [36] works.

Remark 3.6 The statement of Theorem 3.4 is slightly more detailed than the version of [36]. The details we have added come from the proof of Theorem 3.4. In fact the idea of the proof is to split the branched surface B in a neighbourhood of ∂M so that it intersects T_i in parallel simple closed curves of slopes s_i, for $i = 1, \dots k$. In this way, when gluing the solid tori, we can glue meridional discs of these tori to B to obtain a branched surface $B(s_1, \dots, s_k)$ in $M(s_1, \dots, s_k)$ that is laminar and that by [35, Theorem 1] fully carries an essential lamination. In particular, this essential lamination is obtained by gluing the meridional discs of the solid tori to an essential lamination in M that intersects T_i in parallel simple closed curves of slopes s_i, for $i = 1, \dots, k$.

3.2 Constructing branched surfaces in fibred manifolds

In this section we recall a general method to build branched surfaces in mapping tori. This will be the starting point to construct taut foliations on surgeries on fibred two-bridge links.

Let S be a connected oriented surface with boundary, let h be an orientation preserving homeomorphism of S fixing ∂S pointwise and let M_h be the mapping torus associated to (S, h).

We consider pairwise disjoint properly embedded arcs $\alpha_1, \dots, \alpha_k$ in S and discs $\overline{D}_i = \alpha_i \times [0, 1]$ contained in $S \times [0, 1]$. Each of these discs has a "bottom" boundary, $\alpha_i \times \{0\}$, and a "top" boundary, $\alpha_i \times \{1\}$. When we consider the images of these discs in M_h under the projection map

$$S \times [0, 1] \to M_h$$

we have that the bottom and top boundaries become, respectively, $\bigcup_i \alpha_i \subset S$ and $\bigcup_i h(\alpha_i) \subset S$.

We perturb the discs $\overline{D}_i$ in a neighbourhood of $S \times \{1\} \subset S \times [0, 1]$ so that when projected to M_h their top boundaries define a family of arcs, that we still denote $h(\alpha_i)$, in S such that for each $i, j \in \{1, \dots, k\}$ the intersection between α_i and $h(\alpha_j)$ is transverse and the endpoints of α_i and $h(\alpha_i)$ are disjoint. We also denote by D_i the projected perturbed disc contained in M_h and we refer to these discs as *product discs*. If we assign (co)orientations to these discs, since S is (co)oriented, we can smoothen $S \cup D_1 \cup \dots \cup D_k$ to a branched surface B by imposing that the smoothing preserves the co-orientation of S and of the discs. In particular, each disc has two possible co-orientations and hence it can be smoothed in two different ways. This operation is demonstrated in Figure 9, where S is a torus with an open disc removed.

The following proposition shows that in this setting, under very mild hypothesis on the branched surface B, we can get taut foliations on fillings on M_h. The proof is obtained by combining Lemmas 3.16 and 3.23 of [53] and uses as a fundamental result Theorem 3.4.

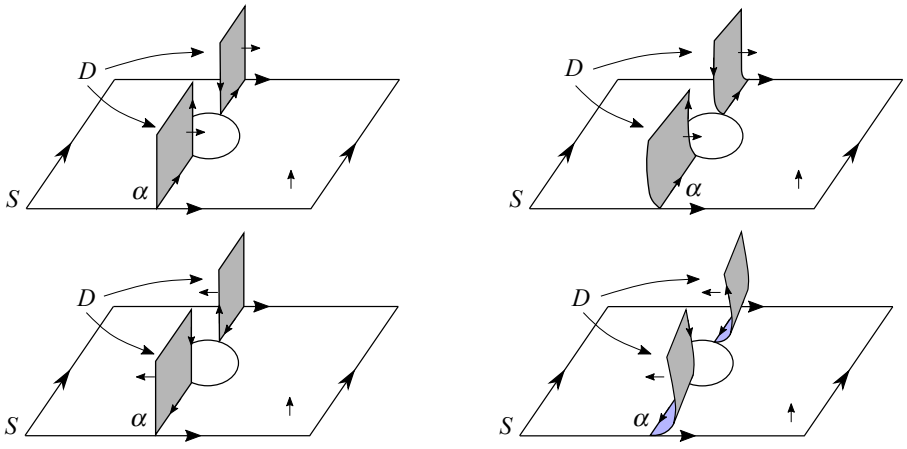


Figure 9: How to smoothen $S \cup D$ according to the co-orientations.

Proposition 3.7 *Suppose that B is a branched surface constructed as described above. Suppose also that B has no sink discs and that $S \setminus \bigcup_{i=1}^{k} \alpha_i$ has no disc components. If $(r_1, \dots, r_n)$ is a multislope realised by ∂B then $M_h(r_1 \dots, r_n)$ contains a co-orientable taut foliation. More precisely there exists a co-orientable taut foliation in M_h intersecting the boundary component T_i in a foliation by curves of slopes r_i, for $i = 1, \dots, n$.* □

3.3 Boundary train tracks

We now briefly discuss the boundary train tracks of the branched surfaces constructed with the procedure described above. We start by describing an explicit way to compute the slopes realised by a train track.

Let τ be an oriented train track in a torus T. Fix a meridian-longitude basis μ, λ and use it to identify slopes on T with elements in $\overline{\mathbb{Q}}$. It is possible to compute the slopes realised by a train track τ by endowing it with *weight systems*. A weight system w on τ is the assignment of a positive rational number, called *weight*, to each sector of τ so that at each branch point of τ the sum of the weights of the incoming sectors is equal to the sum of the outgoing one. Given that τ is oriented, we can associate to such a weight system the rational number $\frac{w_\mu}{w_\lambda}$, where w_μ and w_λ are the *weighted* intersections of the train track with our fixed meridians μ and longitudes λ, as we would do with oriented simple closed curves. This quotient can be interpreted as a *slope* in T and it can be proved that the slopes $\frac{p}{q}$ obtained in this way are exactly those realised by the train track. For details, see [49]. Figure 10 shows an example of a train track with weight systems.

We now focus our attention on a branched surface B in a mapping torus M_h obtained by adding product discs to the fibre S. We start by fixing a meridian μ_j and a longitude λ_j for each boundary component T_j of M_h. We take λ_j to be $S \cap T_j$, with the orientation induced by S, and as meridian the curve $\mu_j = \frac{\{x_j\}\times[0,1]}{\sim_h}$, where $x_j \in S \cap T_j$, oriented in the direction of ascending $t \in [0,1]$. Notice that when T_j is oriented as the boundary of M_h, the algebraic intersection $< \mu_j, \lambda_j >$ is negative. By using this meridian-longitude basis we can identify slopes on T_j with elements in $\overline{\mathbb{Q}}$.

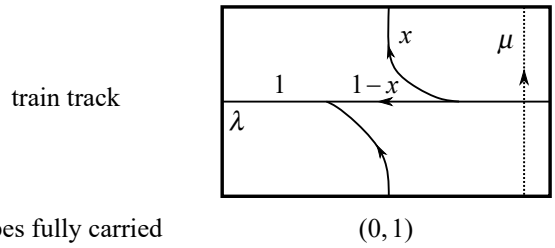


Figure 10: An example of train track τ with a weight system w_x, together with the set of slopes realised by τ. In order for the weights to be positive, x must be a rational number in $(0,1)$.

Given a properly embedded oriented arc α in S we denote its endpoints with $\alpha(0)$ and $\alpha(1)$, so that α goes from $\alpha(0)$ to $\alpha(1)$. Let us now consider a product disc D in the branched surface B, spanned by an arc α. The fixed orientation on D induces an orientation on α and on the two arcs of intersections $D \cap \partial M_h$. More precisely the arc in $D \cap \partial M_h$ containing $\alpha(1)$ is oriented in the direction of ascending $t \in [0,1]$, and the one containing $\alpha(0)$ is oriented in the direction of descending $t \in [0,1]$ (see Figure 9 for an example). For a fixed boundary component T of M_h the four possible configurations of one arc of $D \cap T$ are described in Figure 11, together with the set of slopes fully carried by the train tracks given by one component of $D \cap T$ and the longitude of T.

In the course of the paper, we will mostly encounter train tracks of the four types described in Figure 11, or those that we call of *type* $t_1 + \cdots + t_m$, where t_i is (a), (b), (c) or (d), for $i = \{1, \dots, m\}$. These are obtained by juxtaposing train tracks of type t_i, for $i = \{1, \dots, m\}$, preserving the cyclic order. See Figure 12 for an example.

The next lemma — which follows by a direct computation with weight systems — will be useful when computing the sets of slopes realised by our train tracks.

Lemma 3.8 *Let τ be a train track of type $t_1 + \cdots + t_m$. Then, a slope r is realised by τ if and only if $r = r_1 + \cdots + r_m$, where r_i is a slope realised by the train track of type t_i.* □

In many of the cases of our interest, the boundary train tracks will satisfy the hypothesis of Lemma 3.8 and so we can use it to compute the slopes that they realise. Two examples of train tracks that do not satisfy the hypothesis of the previous lemma can be found in Figure 31 on page 1148.

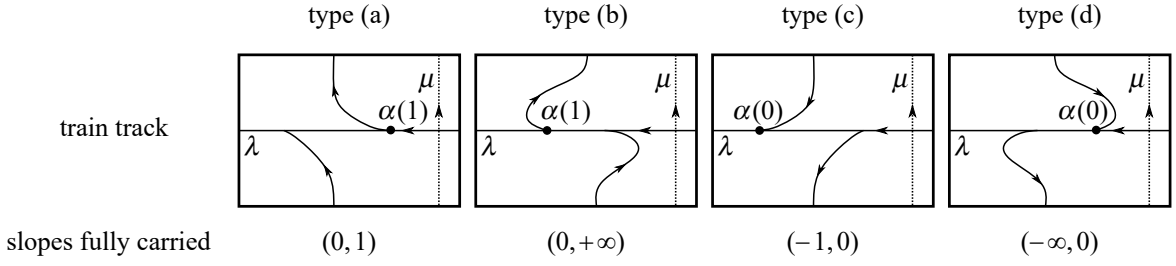


Figure 11: Possible train tracks given by the longitude of T and one arc of intersection in $T \cap D$, together with the intervals of slopes they fully carry.

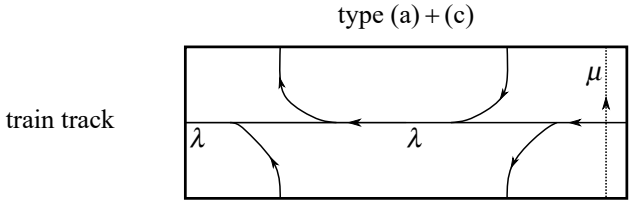


Figure 12: A train track of type (a) + (c).

3.4 Fibred hyperbolic two-bridge links

We are now ready to construct foliations on surgeries on the hyperbolic fibred two-bridge links. The general strategy is simple: we have from (∗∗) explicit descriptions of the monodromies of these links and we want to construct branched surfaces in the way described in Section 3.2. If we are able to construct these branched surfaces so that they have no sink discs, then by Proposition 3.7 we can deduce that all the surgeries corresponding to the multislopes realised by these branched surfaces contain co-orientable taut foliations. For this reason, we will have to study which multislopes are realised by the boundary train tracks.

Before we begin, we make the following remark and establish some conventions.

Remark 3.9 In [52] it is proved that all nontrivial surgeries on fibred hyperbolic knots with fractional Dehn twist coefficient zero contain a co-orientable taut foliation. A similar result does not hold for links. In fact, all fibred hyperbolic two-bridge links have fractional Dehn twist coefficient zero (on both boundary components). This can be proved by finding appropriate arcs moved to the left and to the right and by using [24, Proposition 3.1]. Nonetheless, this family contains infinitely many L-space links.

Conventions Thorough the section we will use the following conventions:

• We fix a fibre surface S for the two-bridge link $L = L(2b_1, \dots, 2b_n)$ and we fix its orientation as in Figure 5. With the induced orientation, L has linking number

$$\mathrm{lk}(L) = \sum_{i=0}^{k} b_{2i+1},$$

where $n = 2k + 1$.

• When a link is fibred there is a natural choice of meridians and longitudes that is in general different from the one induced by the ambient manifold S^3, obtained as follows. We identify $S^3 \setminus \mathrm{int}(N_L) \cong \frac{S \times [0,1]}{\sim_h}$, where N_L is a tubular neighbourhood of L. We fix a point x_i in each boundary component of S and we consider the curves $\mu_i = \frac{\{x_i\} \times [0,1]}{\sim_h}$ oriented in the direction of ascending $t \in [0, 1]$ as meridians and the boundary components λ_i of S as longitudes. By definition of fibred link, the meridians defined in this way coincide with the usual meridians of the link. On the other hand these longitudes do not coincide in general with the canonical longitudes of the link components. In fact, letting l_i denote the canonical

longitude of K_i, we have that

$$(\star) \qquad \lambda_i + \sum_{j \neq i} \mathrm{lk}(K_i, K_j)\mu_i = l_i$$

as elements in $H_1(\partial N_{K_i}, \mathbb{Z})$, where N_{K_i} is the connected component of N_L containing K_i.

From now on we will refer to the bases (μ_i, λ_i) as the *Seifert framing*, and to the bases (μ_i, l_i) as the *canonical framing*. Notice that it follows from $(\star)$ that the longitudes given by the Seifert framing do not depend on the choice of the Seifert surface. Unless otherwise stated we use Seifert framings.

- We will always suppose $n > 1$, because when $n = 1$ the only links obtained in this way are the Hopf links and we are interested in hyperbolic links.
- We construct branched surfaces by considering *oriented* arcs in the fibre surface S and then by attaching product discs as in Section 3.2. We will always co-orient the discs in the following way: we orient them so that the orientations on their boundaries induce the given orientation on the arcs and then we use the orientation of the ambient manifold to co-orient them. Analogously, the co-orientation of the fibre S is obtained by using the orientation of S and of the ambient manifold. A good way to keep in mind the cusps directions of branched surfaces constructed in the way is the following: looking at the positive side of S, the cusps directions is pointing right along the arcs α and β with respect to their orientations and is pointing left along the oriented arcs $h(\alpha)$ and $h(\beta)$ with respect to their orientations. See Figure 14.
- Drawings will usually show the positive side of the (abstract) fibre surface S.
- We will usually omit 1-handles from the drawings, and we draw the attaching arcs of the 1-handles in pink. Compare Figure 6 with Figure 19 on page 1136.
- The arcs $\alpha, \beta, \ldots$ lie on the positive side of S and are depicted with solid lines; their images $h(\alpha), h(\beta), \ldots$ via the monodromy h lie on the negative side of S and are depicted with dashed lines.
- We will usually denote the sectors of a branched surface in S with letters $\mathcal{A}, \mathcal{B}, \ldots$.

With the following lemma we partition fibred hyperbolic two-bridge links in few families and study all of them separately.

Lemma 3.10 *Let $L = L(2b_1, \ldots, 2b_n)$ with $|b_i| = 1$ for all i and n odd. If L is not a torus link then, up to mirroring L, $(b_1, \ldots, b_n)$ satisfies*:

- *Family 1*: *There exist indices l and m such that $b_{2l} \neq b_{2m}$.*
- *Family 2*: *$(b_1, \ldots, b_n) = (b_1, -1, b_3, \ldots, -1, b_n)$ and there exist indices $j \neq l \neq m$ such that $b_{2j+1} = b_{2m+1} = -1$ and $b_{2l+1} = 1$.*
- *Family 3*: *$(b_1, \ldots, b_n) = (-1, -1, -1, \ldots, -1, -1)$.*
- *Family 4*: *$(b_1, \ldots, b_n) = (b_1, -1, b_3, \ldots, -1, b_n)$ where exactly one b_{2j+1} is -1.*

Proof Suppose that L does not belong to family 1. Then, up to mirroring, we can suppose that $b_{2l} = -1$ for all indices l. We now consider the integers b_{2j+1}:

- If there are at least two of them that are negative and one that is positive, then $(b_1, \dots, b_n)$ belongs to family 2.
- If none of them is positive, then $(b_1, \dots, b_n)$ belongs to family 3.
- If exactly one of them is negative, then $(b_1, \dots, b_n)$ belongs to family 4.
- If none of them is negative, then $L = (2, -2, 2, \dots, -2, 2)$ and is a torus link. □

Before beginning the study of these families, we briefly summarise what we will prove and the general strategy of the proofs. For links in families $1, 2$ and 3 we will construct taut foliations on all rational surgeries. The links in family 4 will be divided into two subfamilies, the first containing the links with $b_{2j+1} = -1$ for some j such that $2j + 1 \neq 1, n$ and the second containing those with $b_1 = -1$ or $b_n = -1$. The links in the first subfamily will have taut foliations on all rational surgeries, while those in the second will be isotopic to the links L_n in Figure 1 and we will construct taut foliations on all surgeries in $\big((-\infty, n) \times \mathbb{Q}\big) \cup \big(\mathbb{Q} \times (-\infty, n)\big)$.

For families 1 and 2, the taut foliations will be constructed by using branched surfaces obtained by adding product discs to the fibre surface. For families 3 and 4 we need a more elaborate strategy, and we will drill out one or two carefully chosen loops in the link complements to get new fibred 3-component and 4-component links. We will then study these new links, and construct branched surfaces in their exterior by adding product discs to their fibre surfaces.

3.5 Study of the links in family 1

Recall from the end of Section 2 the definition of river twist and bridge twist.

Lemma 3.11 *Let $L = L(2b_1, \dots, 2b_n)$ with n odd and $|b_i| = 1$ for all i, and let h denote its monodromy as in $(**)$. Let M denote the exterior of L.*

(1) *If there is at least one positive (resp. negative) river twist in the factorisation of the monodromy h, then the manifold $M(r_1, r_2)$ contains a co-orientable taut foliation for every multislope $(r_1, r_2) \in (-\infty, 1)^2$ (resp. for all $(r_1, r_2) \in (-1, +\infty)^2$); see Figure 13*(a)–(b).

(2) *If there are two river twists with different exponents in the factorisation of the monodromy h, then the manifold $M(r_1, r_2)$ contains a co-orientable taut foliation for every multislope $(r_1, r_2) \in \big((-1, +\infty) \times (-\infty, 1)\big) \cup \big((-\infty, 1) \times (-1 + \infty)\big)$; see Figure 13*(c).

Proof (1) Suppose that there is a positive river twist along the curve γ_i. We consider the arcs α and β as in Figure 14. The oriented arcs α and β determine a co-oriented branched surface B obtained by attaching two discs to the fibre surface S as described in Section 3.2. Since $n > 1$, S is not an annulus and therefore the complement of $\alpha \cup \beta$ in S has no disc components. Due to the fact that we have chosen α and β so that they are disjoint from γ_j for $j \neq i$ it follows that $h(\alpha) = \tau_i(\alpha)$ and $h(\beta) = \tau_i(\beta)$, as depicted in Figure 14. In Figure 14 we have also labelled the branch locus of B with the cusps directions and denoted with capital letters $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ the four sectors of the branched surface B in S and none of

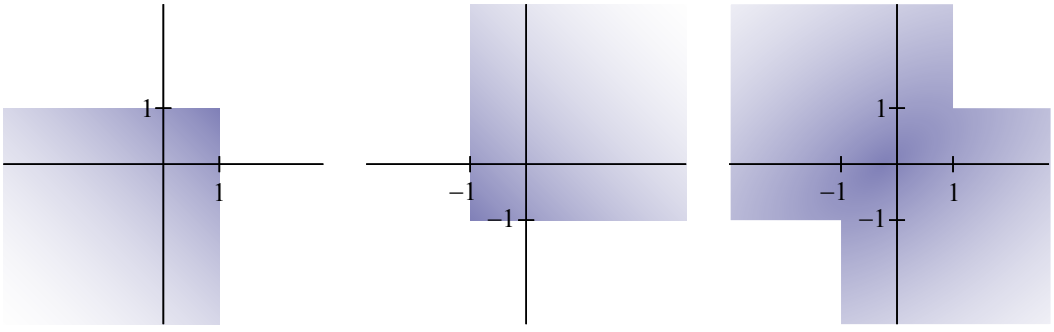


Figure 13: From left to right, the slopes (r_1, r_2) in the coloured region yield manifolds with co-orientable taut foliations in the case where there is, respectively, at least one positive river twist, at least one negative river twist, two river twists with different exponents in the factorisation of the monodromy h.

them is a sink disc. For this reason we can apply Proposition 3.7 and deduce that $M(r_1, r_2)$ supports a co-orientable taut foliation for all the multislopes (r_1, r_2) realised by ∂B. On each boundary component of M we obtain a train track, depicted in Figure 16(a), of type (a) + (d). Therefore by using Lemma 3.8 we have that the boundary train tracks of B realise all the multislopes in $(-\infty, 1)^2$ and by applying Proposition 3.7 we obtain taut foliations on $M(r_1, r_2)$ for all $(r_1, r_2) \in (-\infty, 1)^2$.

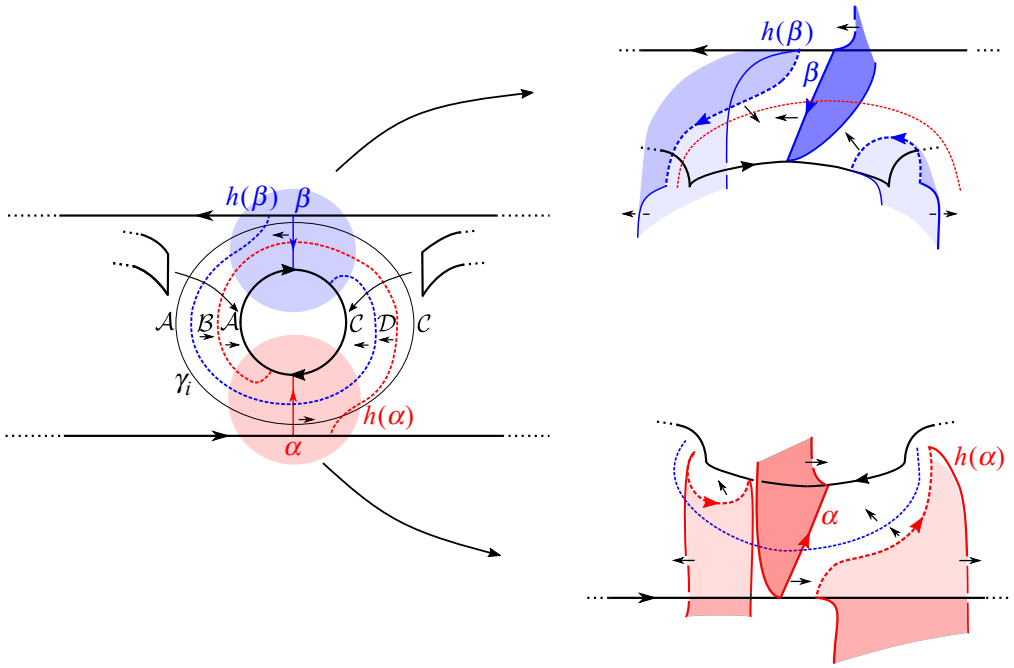


Figure 14: The arcs α and β and the co-oriented discs spanned by them. The letters $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ denote the sectors of B in S.

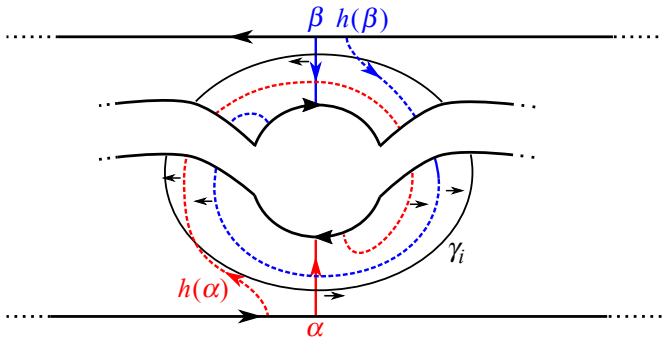


Figure 15: The arcs α and β when the river twist is negative.

If there is a negative river twist, we consider the same oriented arcs α and β (see Figure 15), and on each of the two boundary components we obtain the train track depicted in Figure 16(b), that is of type (b) + (c). This train track realises all the slopes in $(-1,+\infty)$ and so we obtain taut foliations on $M(r_1,r_2)$ for all $(r_1,r_2)\in(-1,+\infty)^2$.

(2) Suppose now that there are two river twists with different exponents in the factorisation of h and suppose that the positive one is along the curve γ_i and the negative one is along γ_j. We suppose $i<j$ but the proof does not change if $j<i$. We choose now α and β as in Figure 17 and as before we have $h(\alpha)=\tau_i(\alpha)$ and $h(\beta)=\tau_j(\beta)$. Also in this case the complement of $\alpha\cup\beta$ contains no disc components. Moreover the complement of $\alpha\cup\beta\cup h(\alpha)\cup h(\beta)$ in S is connected and this implies that there are no sink discs in the branched surface associated to α and β.

The boundary train tracks are shown in Figure 17. The one on the boundary component of M containing the boundary component of S labelled with 1 is of type (b) + (c) and by Lemma 3.8 it realises all slopes in $(-1,+\infty)$. On the other boundary component we get a train track of type (a) + (d) that hence realises all slopes in $(-\infty,1)$. Therefore as a consequence of Proposition 3.7 we have taut foliations in $M(r_1,r_2)$ for all $(r_1,r_2)\in(-1,+\infty)\times(-\infty,1)$. As by Lemma 2.2 two-bridge links are symmetric, we deduce that there are taut foliations also on the surgeries associated to coefficients $(r_1,r_2)\in(-\infty,1)\times(-1,+\infty)$. □

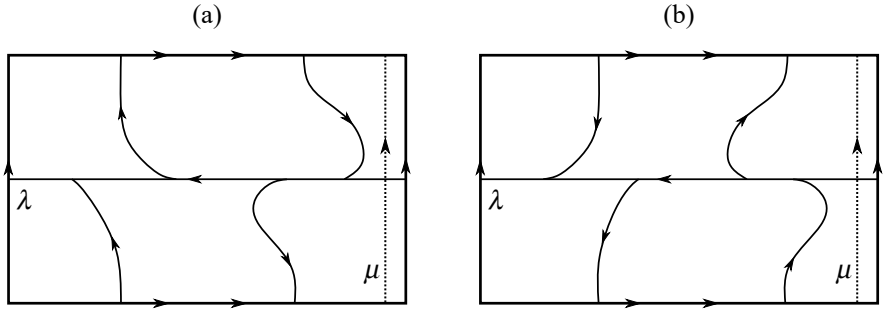


Figure 16: The boundary train tracks in the case where (a) there is a positive river twist and (b) there is a negative river twist.

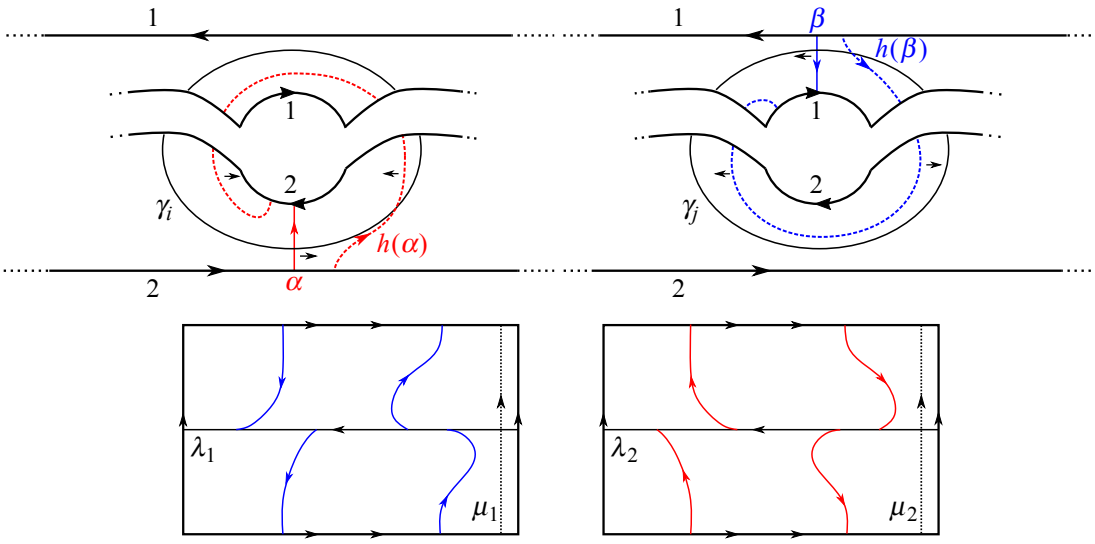


Figure 17: This picture describes the choice of the arcs α and β when the twist along the curve γ_i is positive and the one along γ_j is negative, together with the boundary train tracks of the associated branched surface.

Remark 3.12 Recall that we are working with Seifert framings. However we have already noticed that the meridians of the Seifert framing coincide with the canonical meridians of L. This implies that a surgery coefficient (r_1, r_2) on L is rational with respect to the Seifert framing if and only if it is rational with respect to the canonical framing.

Corollary 3.13 *If the factorisation of the monodromy h has two river twists with different exponents, i.e., if L belongs to family* 1*, then all the rational surgeries on the link L contain co-orientable taut foliations.*

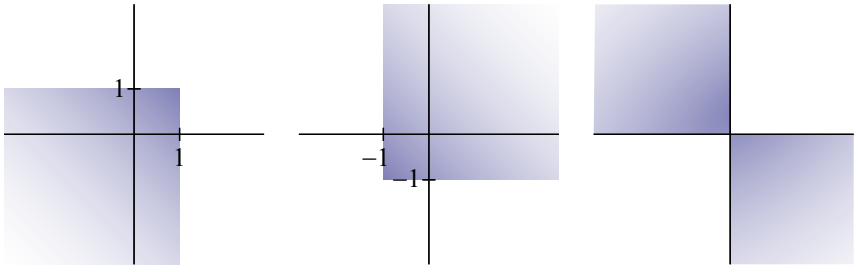
Proof It follows from the first part of Lemma 3.11 that there are co-orientable taut foliations on $M(r_1, r_2)$ for $(r_1, r_2) \in (-\infty, 1)^2 \cup (-1, +\infty)^2$ and it follows from the second part of Lemma 3.11 that there are co-orientable taut foliations on $M(r_1, r_2)$ for $(r_1, r_2) \in \big((-1, +\infty) \times (-\infty, 1)\big) \cup \big((-\infty, 1) \times (-1, +\infty)\big)$. The union of these sets is exactly the set of all rational multislopes. □

3.6 Study of the links in family 2

As a consequence of Corollary 3.13, by taking mirrors if necessary, we can reduce our study to the case where the river twists are all positives, i.e., to links of the form $L = L(2b_1, -2, 2b_3, \dots, -2, 2b_n)$ with n odd.

Lemma 3.14 *Let $L = L(2b_1, \dots, 2b_n)$ with n odd and $|b_i| = 1$ for all i, and let h denote its monodromy as in $(**)$. Let M denote the exterior of L.*

(1) *If there are at least two positive (resp. negative) bridge twists in the factorisation of the monodromy h, then the manifold $M(r_1, r_2)$ contains a co-orientable taut foliation for every multislope $(r_1, r_2) \in (-\infty, 1)^2$ (resp. for all $(r_1, r_2) \in (-1, +\infty)^2$); see Figure 18*(a)–(b).



(a) (b) (c)

Figure 18: From left to right, the slopes (r_1, r_2) in the coloured region yield manifolds with co-orientable taut foliations in the case where there are, respectively, at least two positive bridge twists, at least two negative bridge twists, two bridge twists with different exponents in the factorisation of the monodromy h.

(2) *If there are two bridge twists with different exponents in the factorisation of the monodromy h, then the manifold $M(r_1, r_2)$ contains a co-orientable taut foliation for every multislope $(r_1, r_2) \in \big((0, +\infty) \times (-\infty, 0)\big) \cup \big((-\infty, 0) \times (0, +\infty)\big)$; see Figure 18*(c).

Proof (1) Suppose that the positive bridge twists are along the curves γ_i and γ_j. We consider the oriented arc α and β as in Figure 19. We have $h(\alpha) = \tau_i(\alpha)$: in fact

$$h = \underbrace{\tau_2^{\varepsilon_2}\tau_4^{\varepsilon_4}\dots\tau_{2k}^{\varepsilon_{2k}}}_{\text{river twists}}\underbrace{\tau_1^{\varepsilon_1}\tau_3^{\varepsilon_3}\dots\tau_{2k+1}^{\varepsilon_{2k+1}}}_{\text{bridge twists}}$$

and the only bridge twist that has effect on α is τ_i and the river twists have no effect on $\tau_i(\alpha)$. The same reasoning proves that $h(\beta) = \tau_j(\beta)$. Also in this case we obtain a branched surface that satisfies the hypotheses of Proposition 3.7. In fact this branched surface has at most two sectors[3] $\mathcal{A}$ and $\mathcal{B}$ in S and each of them has some cusp direction on its boundary pointing out of it, as Figure 19 shows. Therefore we just need to study the multislopes realised by the boundary train tracks of B. Both of these are of type (a) + (d) and hence the multislopes realised are the ones in $(-\infty, 1)^2$.

The case where we have two negative bridge twists is analogous: we choose α and β in the same way but now so that they turn right when they meet the curves γ_i and γ_j. Everything works in the same way but now the multislopes realised by the boundary train tracks are the ones in $(-1, +\infty)^2$.

(2) Suppose that are two bridge twists with different exponents in the factorisation of h and suppose that the positive one is along the curve γ_i and the negative one is along γ_j. We choose α and β as in Figure 20. Also in this case there are at most two sectors $\mathcal{A}$ and $\mathcal{B}$ of the resulting branched B surface in S and none of them is a sink disc. Moreover, the boundary train tracks of B realise all the slopes in $(0, +\infty)$ and $(-\infty, 0)$. In fact on one boundary component we have a train track of type (a) + (b) and on the other a train track of type (c) + (d).

Using the fact that two-bridge links are symmetric, we obtain the statement. □

[3]There are two sectors when $i = 1$ and $j = n = 3$ and there is only one sector otherwise.

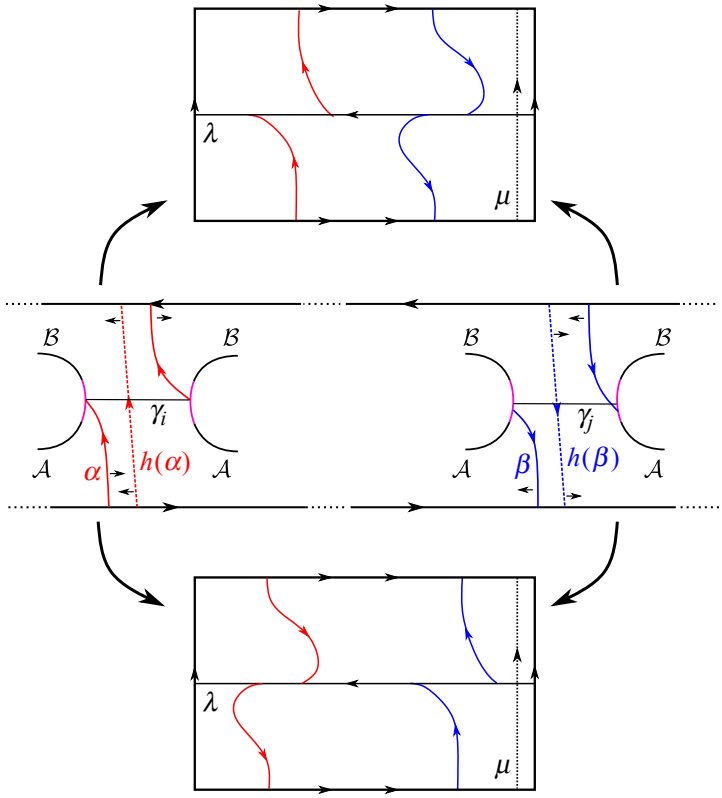


Figure 19: The arcs α and β, together with their image via the monodromy h and the cusps directions, are depicted. We also describe the train tracks obtained on the boundary of M. To simplify the picture we do not draw the 1-handles; we understand that the pink-coloured lines are pairwise identified in the obvious way.

Corollary 3.15 *Let $L = L(2b_1, -2, 2b_3, \dots, -2, 2b_n)$ with n odd and $|b_i| = 1$ for all i, and let h denote its monodromy as in $(**)$. If there are at least two negative bridge twists and one positive bridge twist in the factorisation of h, i.e., if L belongs to family 2, then all the rational surgeries on the link L contain co-orientable taut foliations.*

Proof As a consequence of the fact that the factorisation of h contains positive river twists, by Lemma 3.11 we know that $M(r_1, r_2)$ contains a taut foliation for all the multislopes $(r_1, r_2) \in (-\infty, 1)^2$. Moreover, since there are two negative bridge twists it follows from the first part of Lemma 3.14 that $M(r_1, r_2)$ contains a taut foliation for all the multislopes $(r_1, r_2) \in (-1, +\infty)^2$. As there is also at least one positive bridge twist we can apply the second part of Lemma 3.14 and deduce that $M(r_1, r_2)$ contains a taut foliation for all the multislopes $(r_1, r_2) \in \big((0, +\infty) \times (-\infty, 0)\big) \cup \big((-\infty, 0) \times (0, +\infty)\big)$. The union of these sets is exactly the set of all rational multislopes. □

3.7 Study of the links in family 3

We now focus our attention on the links composing *family* 3.

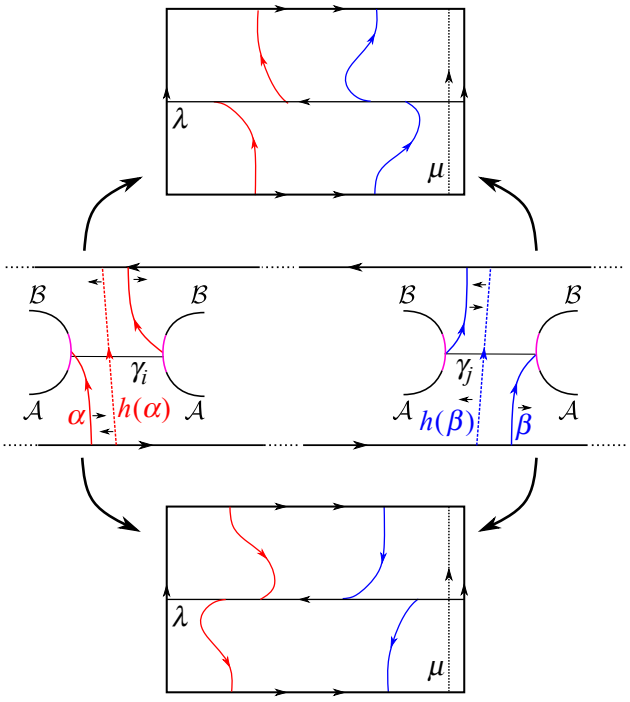


Figure 20: The arcs α and β in the case where there are two bridge twist with different exponents and the boundary train tracks realised on ∂M.

Proposition 3.16 *Let L be a two-bridge link of the form $L = L(-2,-2,-2,\dots,-2,-2)$, i.e., belonging to family 3. Then all the rational Dehn surgeries on L support a co-orientable taut foliation.*

Proof It follows by Lemmas 3.11 and 3.14 that, as the monodromy of L has (at least) two negative bridge twists and (at least) one positive river twist, all the surgery coefficients contained in $(-\infty,1)^2 \cup (-1,+\infty)^2$ yield manifolds with co-orientable taut foliations. We recall that these coefficients are associated to the Seifert framing. We now consider two cases:

• L is not the link $L(-2,-2,-2)$: We construct a branched surface B whose boundary train tracks realise all the multislopes in $(-\infty,1)\times(0,+\infty)$. Two-bridge links are symmetric, hence this will imply the statement. This branched surface is constructed by considering the arcs α and β in Figure 21 and satisfies the hypotheses of Proposition 3.7. In fact the complement of $\alpha \cup \beta$ in S is not a disc, and there is only one sector of B in S. Hence B has no sink discs. We can therefore use B to construct foliations on all the surgeries associated to the multislopes realised by its boundary train tracks: one of these is of type (b), and so realises all slopes in $(0,+\infty)$, and the other is of type (a) + (c) + (d) and hence realises all slopes in $(-\infty,1)$.

• $L = L(-2,-2,-2)$: To study this case we use an idea that will be useful also later on. We construct taut foliations on all the (r,s)-surgeries on L, where $r<0$ or $s<0$. This is enough because we already know from Lemma 3.14 that the surgeries associated to $(r,s)\in(-1,+\infty)^2$ contain taut foliations. Observe

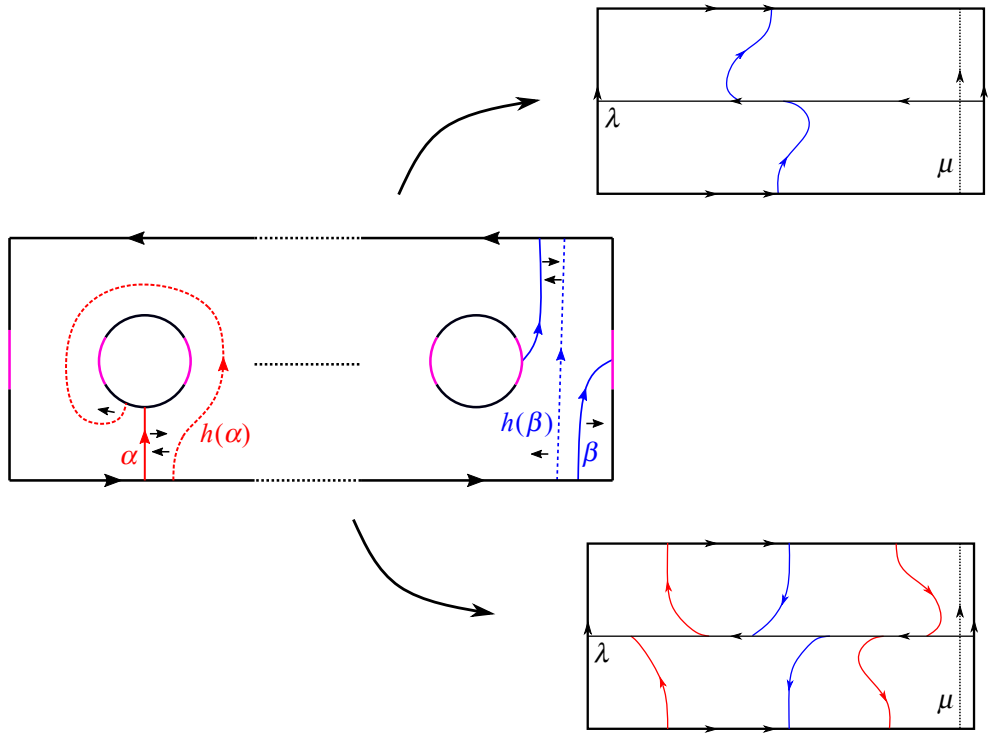


Figure 21: The arcs α and β that we consider when L is not the link $L(-2,-2,-2)$, and the boundary train track of the associated branched surface.

that L can be described as surgery on a 3-components link $\mathcal{L}$, as in Figure 22. The link $\mathcal{L}$ is also fibred, because it is boundary of a surface obtained via a sequence of Hopf plumbing, as described in Figure 22.

Moreover the monodromy of the link $\mathcal{L}$ is given by $h = \tau_4\tau_3^{-1}\tau_2\tau_1^{-1}$, where τ_i denotes the positive Dehn twist along the curve c_i shown in Figure 23.

This description of L will help us to construct the desired taut foliations. The idea is to find a branched surface in the exterior of $\mathcal{L}$ so that the boundary train tracks realise slope -1 on the boundary component

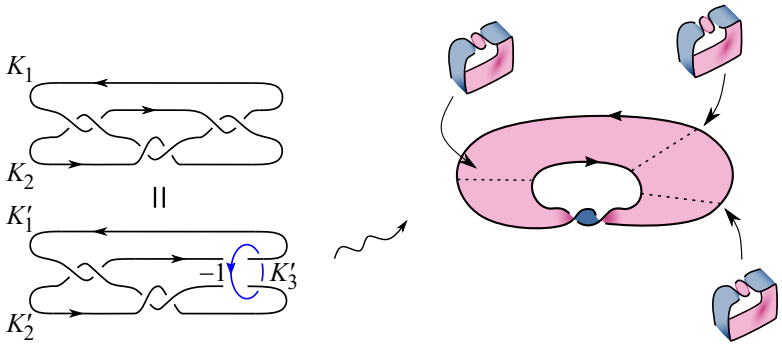


Figure 22: How to obtain the link $L(-2,-2,-2)$ as surgery on a 3-component link $\mathcal{L}$. We also describe a fibre surface for $\mathcal{L}$, obtained via a sequence of Hopf plumbings.

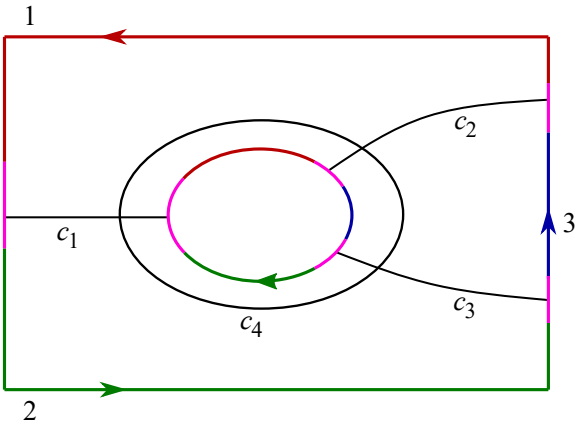


Figure 23: An abstract drawing of the fibre surface for the link $\mathcal{L}$, together with the curves c_i. We have also coloured the three boundary components of S, the boundary component labelled i corresponding to the component K_i' of the link, for $i = 1, 2, 3$.

associated to K_3'. To do this is important to pay attention to how the surgery coefficients change when passing from $\mathcal{L}$ to L. Recall that the coefficients of the slopes are written by using the identification given by the Seifert framing. The $(a, b, -1)$-surgery on $\mathcal{L}$ coincides with the $(a-1, b+1)$-surgery on L, as the following diagram suggests:

$$\begin{array}{ccc} \overbrace{(a,b,-1)}^{\text{Seifert framing for } \mathcal{L}} & \dashrightarrow & \overbrace{(a-1,b+1)}^{\text{Seifert framing for } L} \\ \downarrow & & \uparrow \\ \underbrace{(a,b+2,-1)}_{\text{canonical framing for } \mathcal{L}} & \longrightarrow & \underbrace{(a+1,b+3)}_{\text{canonical framing for } L} \end{array}$$

The changes of coefficients indicated by the vertical arrows are a consequence of formula ($\star$) and the fact that

$$\mathrm{lk}(K_1, K_2) = -2, \quad \mathrm{lk}(K_1', K_2') = \mathrm{lk}(K_2', K_3') = -1, \quad \mathrm{lk}(K_1', K_3') = 1.$$

We construct two branched surfaces B_i in the exterior of $\mathcal{L}$, associated to the arcs α_i, β_i and γ_i, for $i = 1, 2$, as described in Figure 24. It can be checked by direct inspection that for $i = 1, 2$ the complement of $\alpha_i \cup \beta_i \cup \gamma_i$ contains no disc components in S, and that there are no sink discs. Hence we can apply Proposition 3.7 and deduce that these branched surfaces carry laminations that extend to taut foliations on the manifolds obtained by Dehn filling the boundary tori along the multislopes realised by the boundary train tracks. First, we consider the boundary train tracks of B_1. Notice that the one contained in the boundary component of M labelled with 1 does not satisfy the hypothesis of Lemma 3.8, and so we cannot use it to compute the slopes it realises. Nonetheless, a direct computation with weight systems shows that it realises all slopes in $(-\infty, 1)$. The two other boundary train tracks of B_1 are of type (b) and type (c)+(d)

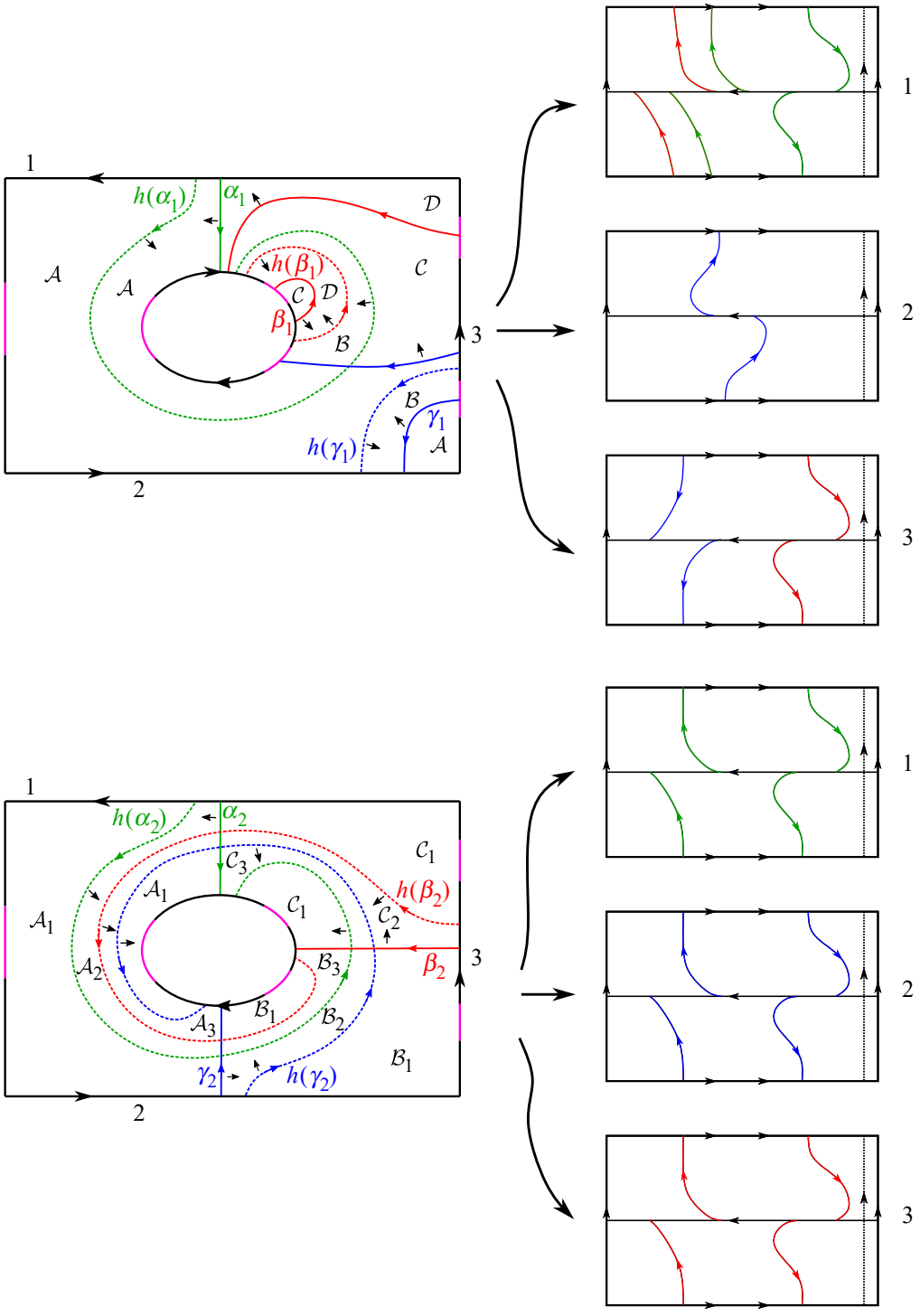


Figure 24: The arcs α_i, β_i, γ_i and their images via the monodromy h, together with the cusp directions of the associated branched surfaces and their boundary train tracks. We have also indicated the sectors of the branched surfaces in S.

and so realise all slopes in $(0,+\infty)$ and $(-\infty,0)$ on the corresponding boundary components. Summing up, the boundary train tracks of B_1 realise all the multislopes in $(-\infty,1)\times(0,+\infty)\times(-\infty,0)$. The boundary train tracks associated to B_2 are all of type (a) + (d) and hence realise all multislopes $(-\infty,1)^3$. In particular, we have taut foliations on $S^3_{r,s,-1}(\mathcal{L}) = S^3_{r-1,s+1}(L)$ for all $(r,s)\in(-\infty,1)\times\mathbb{R}$.

Since L is symmetric, and since all multislopes in $(-1,+\infty)^2$ are covered by Lemma 3.14, the statement follows. □

3.8 Study of the links in family 4

We now focus on the links of *family* 4, i.e., on the links of the form $L = L(2b_1,-2,2b_3,\dots,-2,2b_m)$ where exactly one b_i is -1 and all the others are equal to 1. We first study the case when $b_i = 1$ for some $i \neq 1,m$. We write $m = 2n+1$ for some positive integer n.

Lemma 3.17 *Let $L = L(2b_1,-2,2b_3,\dots,-2,2b_m)$ where exactly one b_{2k+1} is -1 and all the others are equal to 1 and suppose that $2k+1 \neq 1,m$. Then L is isotopic as an unoriented link to $L(-2k,-2,2,-2,-2h)$, where $h = n-k$.*

Proof We will prove this algebraically. We start by computing the fraction associated to the link $L(-2k,-2,2,-2,-2h)$. We have

$$-2k+\cfrac{1}{-2+\cfrac{1}{2+\cfrac{1}{-2-\cfrac{1}{2h}}}} = -2k+\cfrac{1}{-2+\cfrac{1}{2-\cfrac{2h}{4h+1}}} = -2k+\cfrac{1}{-2+\cfrac{4h+1}{6h+2}}$$
$$= -2k+\frac{6h+2}{-(8h+3)} = \frac{16kh+6k+6h+2}{-(8h+3)}$$

and this implies $L(-2k,-2,2,-2,-2h) = b(16kh+6k+6h+2,-(8h+3))$, where $b(p,q)$ denotes the two-bridge link associated to the rational $\frac{p}{q}$.

We now study the fraction corresponding to L. This fraction depends on k and n, or equivalently on k and $h = n-k$, and we denote its reduced representative by $\frac{\alpha_{k,h}}{\beta_{k,h}}$. Then we have

$$\frac{\alpha_{k,h}}{\beta_{k,h}} = 2+\cfrac{1}{-2+\cfrac{1}{2+\cfrac{1}{\ddots+\cfrac{1}{-2+\cfrac{1}{-2+\cfrac{q_h}{p_h}}}}}},$$

where the last -2 corresponds to $2b_{2k+1}$, and where $\frac{p_h}{q_h}$ is defined as

$$\frac{p_h}{q_h} = \overbrace{-2+\cfrac{1}{2+\cfrac{1}{-2+\cfrac{1}{\ddots+\cfrac{1}{2}}}}}^{\text{length } 2h}.$$

One can check by induction on $h \geq 1$ that $\frac{p_h}{q_h} = \frac{2h+1}{-2h}$, and hence $p_h = 2h+1$ and $q_h = -2h$.

We now prove by induction on k that

$$\alpha_{k,h} = 16kh + 6k + 6h + 2,$$
$$\beta_{k,h} = 16kh - 2h + 6k - 1$$

for every h.

- Case $k = 1$: We have the equality

$$\frac{\alpha_{1,h}}{\beta_{1,h}} = 2 + \cfrac{1}{-2+\cfrac{1}{-2+\cfrac{q_h}{p_h}}} = 2 + \cfrac{1}{-2-\cfrac{1+2h}{6h+2}} = 2 - \frac{6h+2}{14h+5} = \frac{22h+8}{14h+5}.$$

Since $(22h+8, 14h+5) = (q_h, p_h) = 1$ we deduce that $\alpha_{1,h} = 22h+8$ and $\beta_{1,h} = 14h+5$.

- Case $k > 1$: We can use the equality

$$\frac{\alpha_{k,h}}{\beta_{k,h}} = 2 + \cfrac{1}{-2+\cfrac{\beta_{k-1,h}}{\alpha_{k-1,h}}} = 2 + \frac{\alpha_{k-1,h}}{-2\alpha_{k-1,h} + \beta_{k-1,h}} = \frac{3\alpha_{k-1,h} - 2\beta_{k-1,h}}{2\alpha_{k-1,h} - \beta_{k-1,h}}$$

and the fact that

$$(3\alpha_{k-1,h} - 2\beta_{k-1,h}, 2\alpha_{k-1,h} - \beta_{k-1,h}) = (\alpha_{k-1,h}, \beta_{k-1,h}) = 1$$

to deduce that $\alpha_{k,h} = 3\alpha_{k-1,h} - 2\beta_{k-1,h}$ and that $\beta_{k,h} = 2\alpha_{k-1,h} - \beta_{k-1,h}$. Therefore we have

$$\alpha_{k,h} - \beta_{k,h} = \alpha_{k-1,h} - \beta_{k-1,h} = 8h + 3,$$
$$\alpha_{k,h} - \alpha_{k-1,h} = 2(\alpha_{k-1,h} - \beta_{k-1,h}) = 16h + 6.$$

These equalities imply

$$\alpha_{k,h} = \alpha_{k-1,h} + 16h + 6 = 16kh + 6k + 6h + 2,$$
$$\beta_{k,h} = \alpha_{k,h} - 8h - 3 = 16kh - 2h + 6k - 1$$

and this proves the claim.

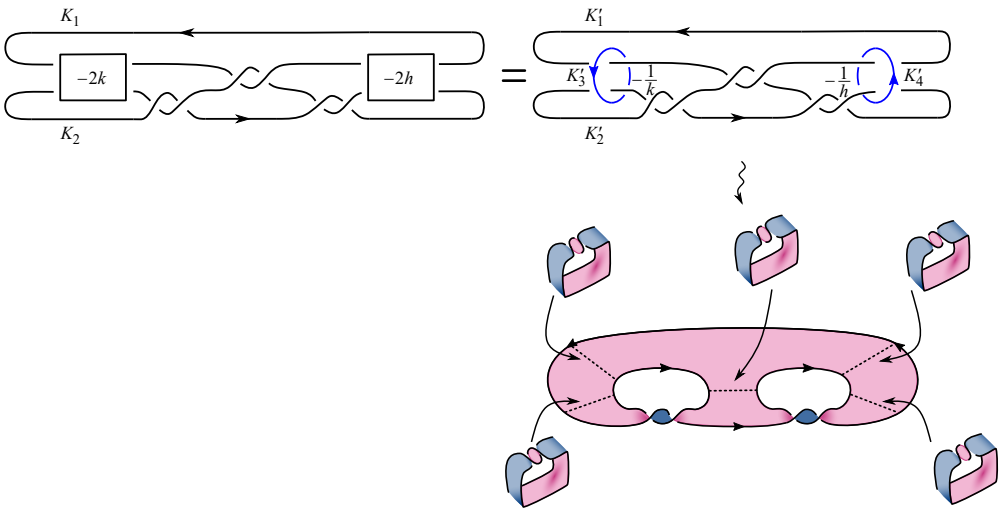


Figure 25: How to obtain the link $L(-2k,-2,2,-2,-2h)$ as surgery on a 4-component link $\mathcal{L}$. We also describe a fibre surface for $\mathcal{L}$, obtained as a sequence of Hopf plumbings.

To conclude the proof of the lemma we just have to recall from Theorem 2.1 that if $\beta' \equiv \alpha + \beta \bmod 2\alpha$ then the links $b(\alpha,\beta)$ and $b(\alpha,\beta')$ are isotopic after reversing the orientation of one of the components. In the case of our interest we have

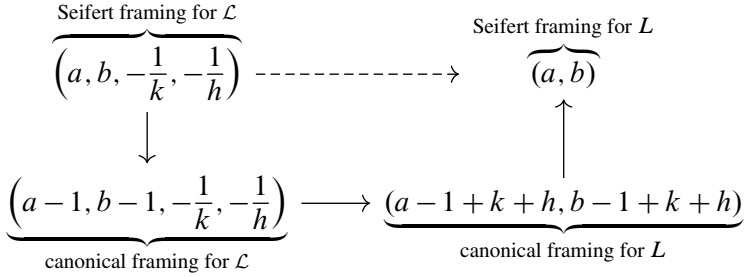
$$\alpha_{k,h} + \beta_{k,h} \equiv -\alpha_{k,h} + \beta_{k,h} \equiv -(8h+3) \mod 2\alpha_{k,h}. \qquad \square$$

The description given by the previous lemma allows us to prove:

Proposition 3.18 *Let $L = L(2b_1,-2,2b_3,\dots,-2,2b_m)$ where exactly one b_{2k+1} is -1 and all the others are equal to 1 and suppose that $2k+1 \neq 1,m$. Then all the rational Dehn surgeries on L support co-orientable taut foliations.*

Proof By virtue of Lemma 3.17 it is equivalent to study surgeries on links of the form $L_{k,h} = L(-2k,-2,2,-2,-2h)$ where $h>0$ and $k>0$. These links can be obtained as surgeries on a 4-component fibred link $\mathcal{L}$, as described in Figure 25. Our aim now is to construct foliations on enough surgeries on $\mathcal{L}$.

The monodromy of the link $\mathcal{L}$ is given by $h = \tau_5\tau_3\tau_7\tau_6^{-1}\tau_4\tau_2\tau_1^{-1}$, where τ_i denotes the positive Dehn twist along the curve c_i shown in Figure 26. If we label the components of L and $\mathcal{L}$ as described in Figure 25, the surgery coefficients change in the following way:



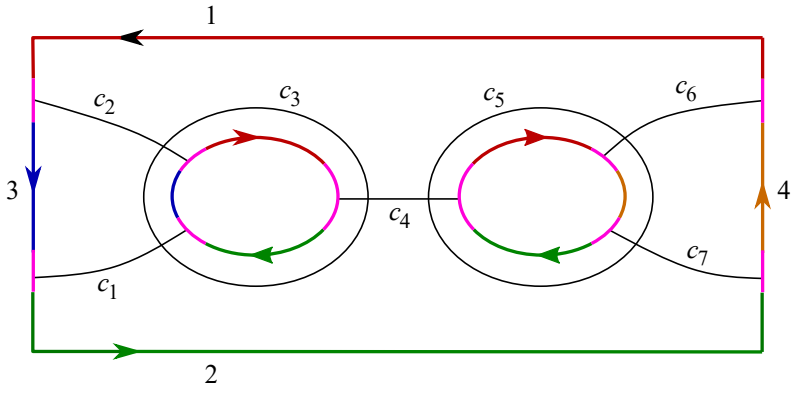


Figure 26: An abstract drawing of the fibre surface for the link $\mathcal{L}$, together with the curves c_i. We have also coloured the boundary components of S, the one labelled with i corresponding to the component K_i' of the link, for $i = 1, 2, 3, 4$.

As usual, when constructing foliations it is more natural to work with the framings given by the Seifert surfaces.

We construct two branched surfaces in the exterior of $\mathcal{L}$. The first one is associated to the arcs α, β, γ, δ, ϵ depicted in Figure 27, where we also describe the types of the boundary train tracks. The complement of these arcs in the fibre surface is not a disc (it is easier to see this by considering the complement of the images of these arcs via the diffeomorphism h) and the branched surface does not contain sink discs. In fact, there are five sectors in S, labelled with capital letters in Figure 27 and none of them is a sink disc. Therefore we can apply Proposition 3.7 and deduce that there exist taut foliations on all the surgeries on $\mathcal{L}$ corresponding to multislopes in $(0, +\infty) \times \mathbb{R} \times (-\infty, 0) \times (-\infty, 0)$.

The second branched surface is the one associated to the arcs described in Figure 28. In this case all the boundary train tracks are of type (a) + (d) and hence we are able to construct foliations on the surgeries corresponding to multislopes in $(-\infty, 1)^4$.

This implies that for every $k > 0$ and $h > 0$ all the surgeries on the link $L_{k,h}$ corresponding to multislopes in $(0, +\infty) \times \mathbb{R}$ and in $(-\infty, 1)^2$ support a co-orientable taut foliation. The conclusion follows using the fact that all these links are symmetric. □

Now we only have to study the links $L = L(2b_1, -2, 2b_3, \dots, -2, 2b_{2n+1})$ where $b_1 = -1$ and all the other b_i are 1, or where $b_{2n+1} = -1$ and all the other b_i are 1. The link $L(a_1, a_2, \dots, a_{2n+1})$ is isotopic to $L(a_{2n+1}, \dots, a_2, a_1)$, so we can reduce our study to the case when $b_{2n+1} = -1$ and we denote the corresponding link by L_n.

Lemma 3.19 *The link L_n is isotopic as unoriented link to the link $L(2, -2, -2n)$, illustrated in Figure 1.*

Proof We compute the fractions associated to these links. The one associated to $L(2, -2, -2n)$ is $\frac{6n+2}{4n+1}$. Therefore by Theorem 2.1 the link $L(2, -2, -2n)$ is isotopic, after reversing the orientation of one of the components, to the link defined by the fraction $\frac{6n+2}{-(2n+1)}$.

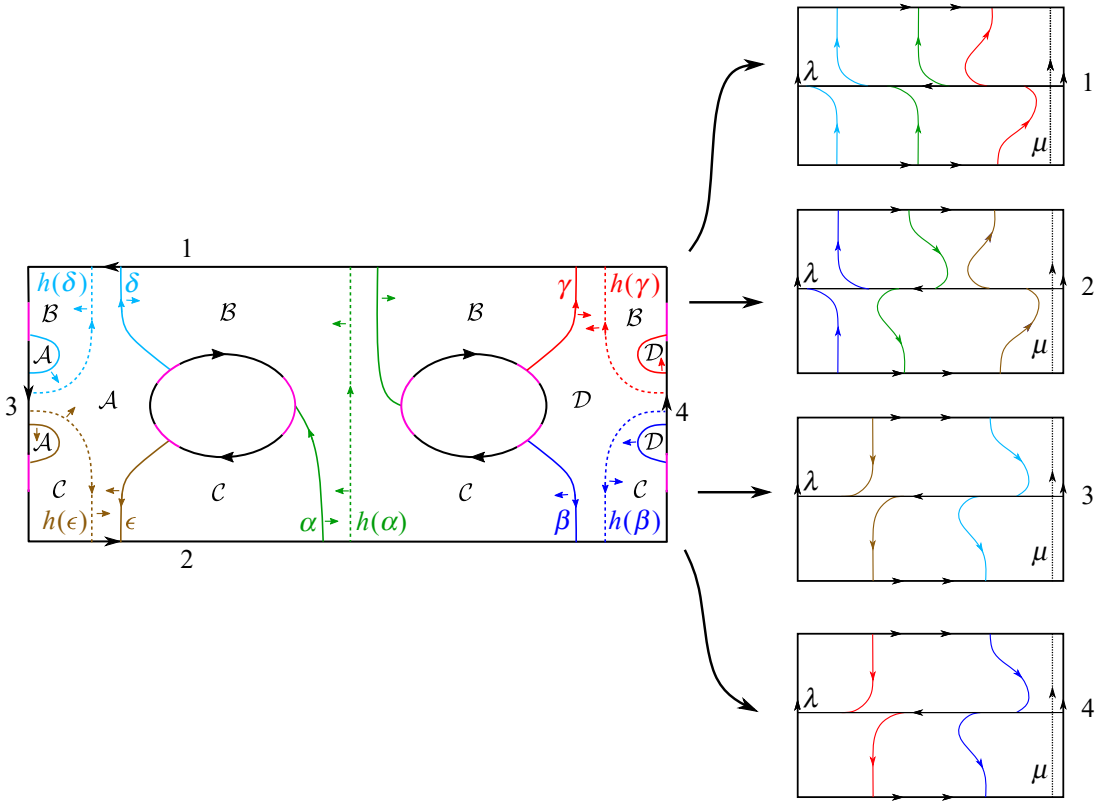


Figure 27: The arcs α, β, γ, δ, ϵ and the boundary train tracks of the associated branched surface.

The fractions $\frac{p_n}{q_n}$ associated to L_n satisfy the recursive equation

$$\frac{p_n}{q_n} = 2 + \frac{1}{-2 + \dfrac{q_{n-1}}{p_{n-1}}} = 2 + \frac{p_{n-1}}{-2p_{n-1} + q_{n-1}} = \frac{3p_{n-1} - 2q_{n-1}}{2p_{n-1} - q_{n-1}}. \tag{1}$$

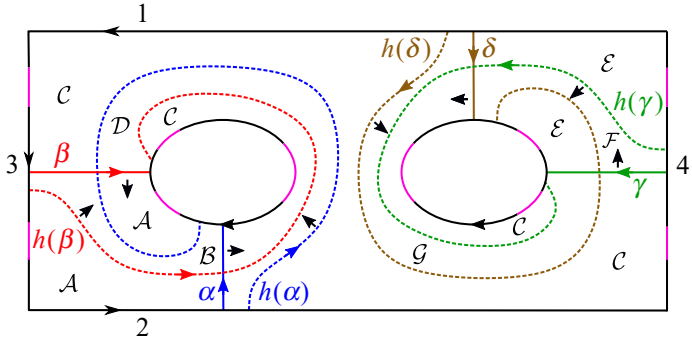


Figure 28: The arcs α, β, γ, δ used to construct the second branched surface. We have also labelled the sector contained in S and one can check that none of them is a sink disc.

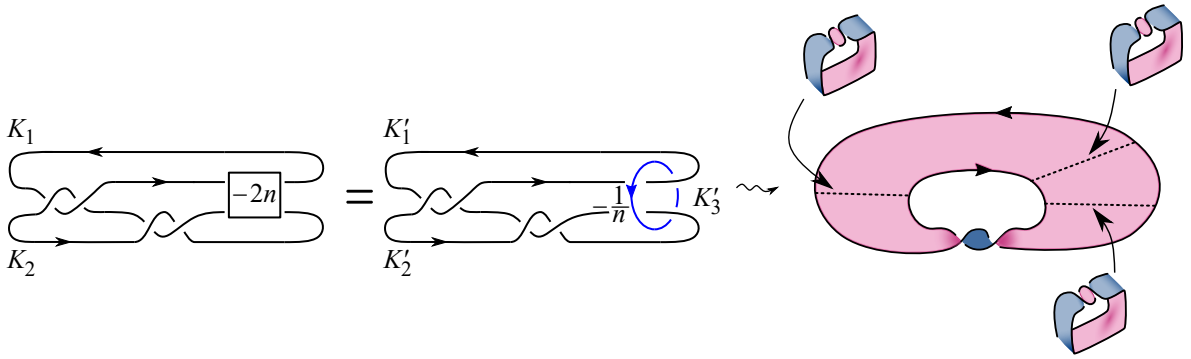


Figure 29: How to obtain the links $\{L_n\}_{n\geq 1}$ as surgery on a 3-components link $\mathcal{L}$ and a fibre surface S for $\mathcal{L}$.

Let us find an explicit formula for p_n and q_n. It follows from (1) that

$$p_n - q_n = p_{n-1} - q_{n-1}$$

and as a consequence the quantity $p_i - q_i$ does not depend on the index i. Moreover, (1) also implies

$$p_n - p_{n-1} = q_n - q_{n-1} = 2(p_{n-1} - q_{n-1})$$

and therefore also the quantity $p_i - p_{i-1} = q_i - q_{i-1}$ is constant in i. As when $n = 1$ we have $\frac{p_1}{q_1} = \frac{8}{5}$, we deduce by induction that $p_n = 8 + (n-1)6 = 6n + 2$ and $q_n = 5 + (n-1)6$. To conclude the proof is enough to observe that $-(2n+1) \equiv q_n^{-1} \bmod 2p_n$ and use again Theorem 2.1. □

We will prove in the next section (see Proposition 4.3) that when $r \geq n$, $s \geq n$ the (r,s)-surgery on L_n is an L-space, where the surgery coefficients are to be considered in the *canonical* framing. We now prove that all the other (rational) surgeries on L_n support co-orientable taut foliations.

Proposition 3.20 *Let $L_n = L(2b_1, -2, 2b_3, \dots, -2, 2b_{2n+1})$, where $b_{2n+1} = -1$ and all the other b_i are 1, and let $r < n$, $s < n$ be rational numbers. Then the (r,s)-surgery on L_n supports a co-orientable taut foliation, where the surgery coefficients are considered in the canonical framing.*

Proof We can suppose that $n \geq 2$, because L_1 is the Whitehead link, for which the corresponding result was proved in [53]. By Lemma 3.19 we have $L_n = L(2, -2, -2n)$ as unoriented links and by using this representation it is evident that $L_n = S^3_{\bullet,\bullet,-1/n}(\mathcal{L})$, where $\mathcal{L}$ is drawn in Figure 29. This figure also shows a fibre surface S for $\mathcal{L}$ obtained via a sequence of four Hopf plumbings.

We choose four triples α, β, γ of oriented arcs in S and consider the four branched surfaces in the exterior of $\mathcal{L}$ associated to these arcs, as depicted in Figures 30 and 31. Each of these triples has the property that its complement in S contains no disc components.

Moreover, the figures also illustrate the sectors of the branched surfaces contained in S and it can be checked that none of these is a sink disc. Thus, thanks to Proposition 3.7 we only need to study the boundary train tracks of these branched surfaces in order to construct the desired taut foliations. The multislopes realised by these branched surfaces in the Seifert framing of $\mathcal{L}$ are, respectively,

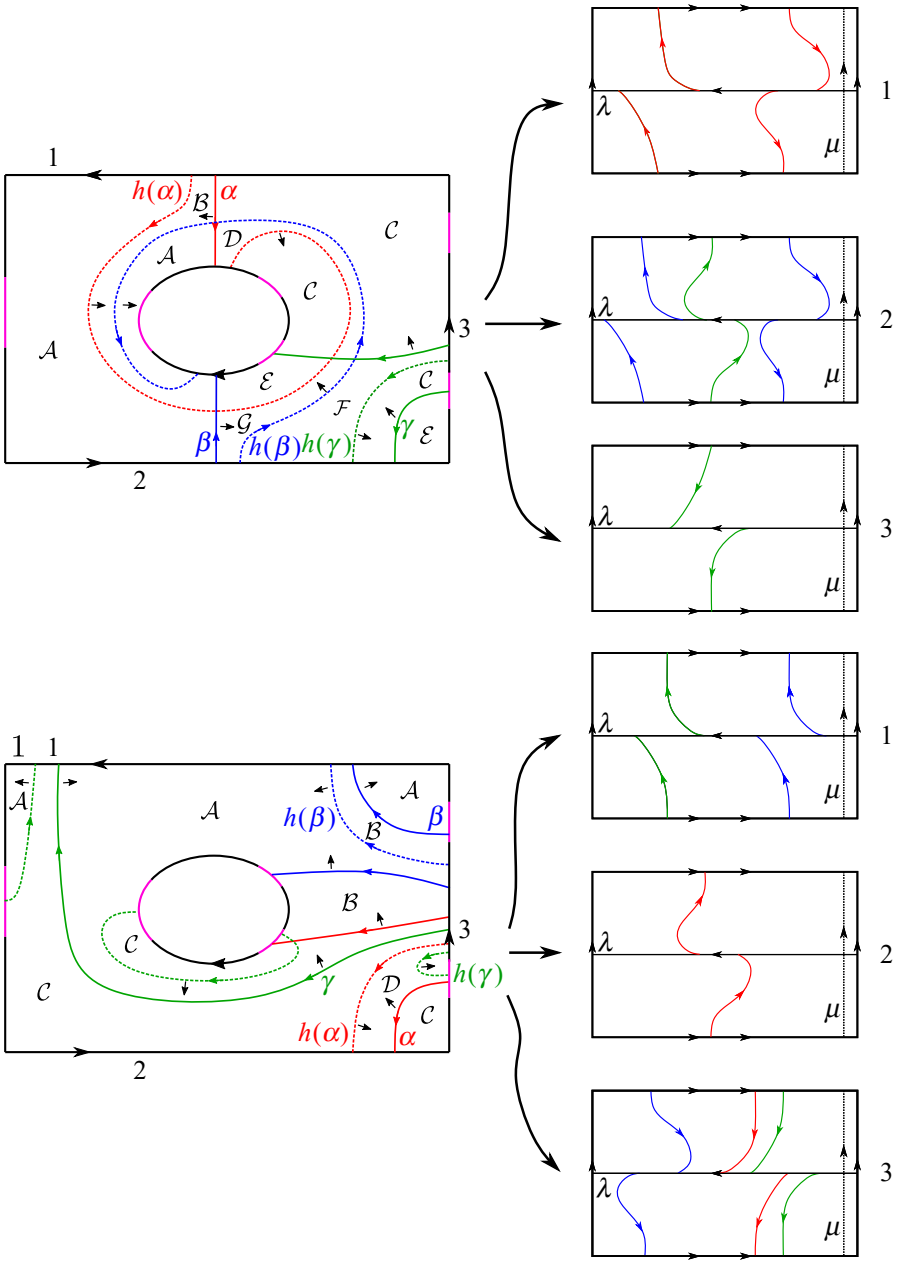


Figure 30: How to choose two of the four triples of arcs α, β, γ. The picture also represents their images via the monodromy, the boundary train tracks, and the sectors in S.

- all the multislopes in $(-\infty, 1) \times \mathbb{R} \times (-1, 0)$;
- all the multislopes in $(0, 2) \times (0, +\infty) \times (-\infty, 0)$;
- all the multislopes in $(0, 2) \times (-\infty, 0) \times (-1, 0)$;
- all the multislopes in $(-\infty, 2) \times (-1, 1) \times (-1, 0)$.

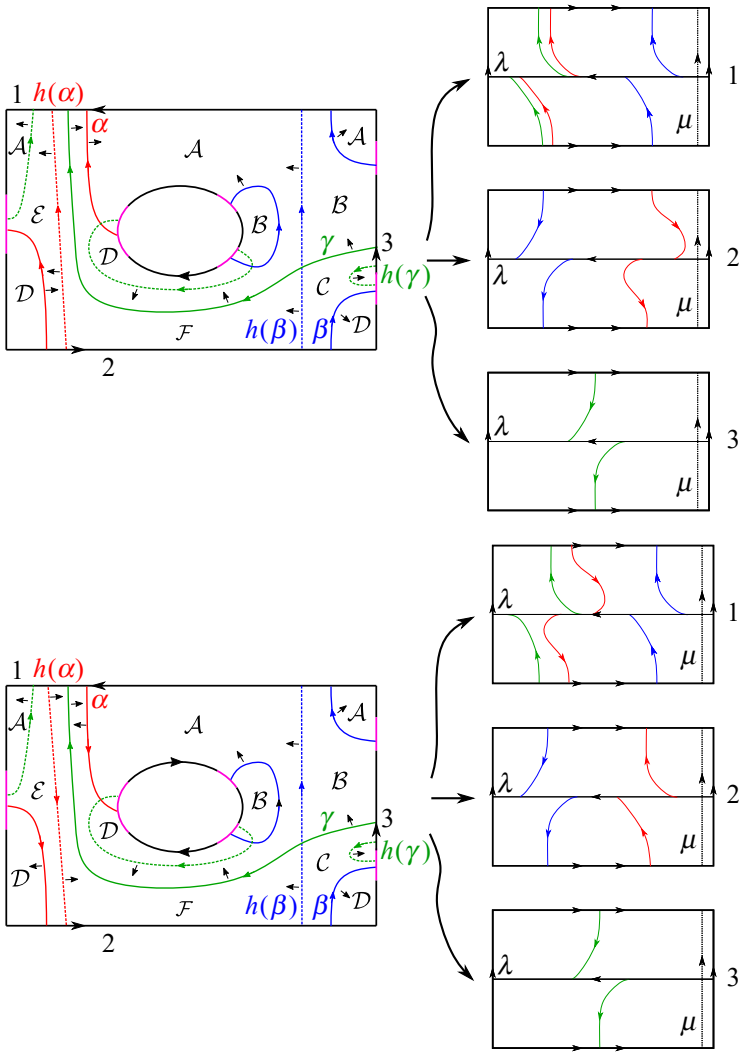


Figure 31: How to choose the two other triples of arcs α, β, γ. The picture also represents their images via the monodromy, the boundary train tracks, and the sectors in S.

We now prove that by considering L_n as $-\frac{1}{n}$ surgery on the third component of $\mathcal{L}$, we have constructed the desired foliations on the surgeries on L_n. First of all we observe that

$$\mathrm{lk}(K_1', K_2') = \mathrm{lk}(K_1', K_3') = 1, \quad \mathrm{lk}(K_2', K_3') = -1$$

and by using formula $(\star)$ we deduce the change of surgery coefficients

$$\underbrace{\left(a, b, -\frac{1}{n}\right)}_{\text{Seifert framing for } \mathcal{L}} \to \overbrace{\left(a-2, b, -\frac{1}{n}\right)}^{\text{canonical framing for } \mathcal{L}} \to \overbrace{(a+n-2, b+n)}^{\text{canonical framing for } L_n}.$$

Therefore, for every $n \geq 2$, we obtain taut foliations on all the surgeries on L_n corresponding to multislopes in

- $W = (-\infty, n-1) \times \mathbb{R}$;
- $X = (n-2, n) \times (n, +\infty)$;
- $Y = (n-2, n) \times (-\infty, n)$;
- $Z = (-\infty, n) \times (n-1, n+1)$.

We now show that these four sets are enough to deduce that, for all $n \geq 2$, all the surgeries on L_n corresponding to multislopes (r_1, r_2) where $r_1 < n$ or $r_2 < n$ support a co-orientable taut foliation. In fact suppose that we have such a pair (r_1, r_2). Since L_n is symmetric we can suppose that $r_1 < n$ and we have the following cases:

- $r_1 < n-1$: In this case the pair is contained in the set W.
- $n-1 \leq r_1 < n$: If $r_2 > n$ the pair is contained in X, if $r_2 < n$ we conclude by using the set Y and if $r_2 = n$ we use the set Z. □

4 *L*-spaces

We now study the L-space surgeries on the links L_n, illustrated in Figure 1, and conclude the proof of Theorem 1.2 and Proposition 1.5. To do this we will use the main result of [50]. We recall some definitions and fix some notation.

Let Y be a rational homology solid torus, i.e., Y is a compact oriented 3-manifold with toroidal boundary such that $H_*(Y; \mathbb{Q}) \cong H_*(\mathbb{D}^2 \times S^1; \mathbb{Q})$.

We are interested in the study of the Dehn fillings on Y. We define the *set of slopes in* Y as

$$\mathrm{Sl}(Y) = \{\alpha \in H_1(\partial Y; \mathbb{Z}) \mid \alpha \text{ is primitive}\}/\pm 1.$$

We will denote with $Y(\alpha)$ the Dehn filling of Y associated to $[\alpha] \in \mathrm{Sl}(Y)$.

Notice that as a consequence of Y being a rational homology solid torus, there is a distinguished slope in $\mathrm{Sl}(Y)$ that we call the *homological longitude* of Y and that is defined in the following way. We denote with $i : H_1(\partial Y; \mathbb{Z}) \to H_1(Y; \mathbb{Z})$ the map induced by the inclusion $\partial Y \subset Y$ and we consider a primitive element $l \in H_1(\partial Y; \mathbb{Z})$ such that $i(l)$ is torsion in $H_1(Y; \mathbb{Z})$. The element l is unique up to sign, and its equivalence class $[l] \in \mathrm{Sl}(Y)$ is the homological longitude of Y. This definition, which may seem counterintuitive, is given so that when Y is the complement of a knot in S^3, the homological longitude of Y coincides with the slope defined by the longitude of the knot.

We want to study the fillings on Y that are L-spaces. For this reason we define the set of the *L-space filling slopes* as

$$\mathcal{L}(Y) = \{[\alpha] \in \mathrm{Sl}(Y) \mid Y(\alpha) \text{ is an } L\text{-space}\}.$$

Once we fix a basis (μ, λ) for $H_1(\partial Y; \mathbb{Z})$ we can identify the set $\mathrm{Sl}(Y)$ with $\overline{\mathbb{Q}}$. The following theorem is a straightforward consequence of [50, Theorem 1.6].

Theorem 4.1 *Let Y be a rational homology solid torus and let $[\alpha] \neq [\beta]$ be two slopes in $\mathcal{L}(Y)$. Then the set $\mathcal{L}(Y)$ contains the interval in $\mathrm{Sl}(Y)$ between $[\alpha]$ and $[\beta]$ that does not contain the homological longitude $[l]$.* □

We want to use this result to study L-space surgeries on links. If L is a link in S^3, we denote by $\mathcal{L}(L)$ the set of slopes in the exterior of L such that the corresponding surgery is an L-space. For each component of the link we fix the canonical meridian and longitude, and in this way we can identify $\mathcal{L}(L)$ with a subset of $\overline{\mathbb{Q}}^d$, where d is the number of components of the link. We fix $d = 2$, i.e., we suppose that L has two components K_1 and K_2. Given $(r_1, r_2) \in \overline{\mathbb{Q}}^2$, we denote by

- $S^3_{r_1,r_2}(L)$ the (r_1, r_2)-surgery on L;
- $S^3_{r_1,\bullet}(L)$ the manifold obtained by drilling K_2 and performing r_1-surgery on K_1;
- $S^3_{\bullet,r_2}(L)$ the manifold obtained by drilling K_1 and performing r_2-surgery on K_2.

Recall that if L has two components, then by using Mayer–Vietoris one can see that the manifold $S^3_{r_1,r_2}(L)$ is not a rational homology sphere if and only if $\{r_1, r_2\} = \{0, \infty\}$ or $r_1 r_2 = \mathrm{lk}(L)^2$, where $\mathrm{lk}(L)$ denotes the linking number of the components of L. Hence if $r_1 \neq 0$ the manifold $S^3_{r_1,\bullet}(L)$ is a rational homology solid torus with homological longitude given by $\frac{\mathrm{lk}(L)^2}{r_1} \in \mathbb{Q}$. Analogously, if $r_2 \neq 0$ the manifold $S^3_{\bullet,r_2}(L)$ is a rational homology solid torus with homological longitude given by $\frac{\mathrm{lk}(L)^2}{r_2} \in \mathbb{Q}$.

Proposition 4.2 *Let L be a two-component link with two unknotted components. Suppose $(r_1, r_2) \in \mathcal{L}(L)$ with $r_1 r_2 > \mathrm{lk}(L)^2$ and $r_1 > 0, r_2 > 0$. Then $\big([r_1, \infty] \times [r_2, \infty]\big) \cap \overline{\mathbb{Q}}^2$ is contained in $\mathcal{L}(L)$. Analogously, if $r_1 r_2 > \mathrm{lk}(L)^2$ and $r_1 < 0, r_2 < 0$ then $\big([\infty, r_1] \times [\infty, r_2]\big) \cap \overline{\mathbb{Q}}^2$ is contained in $\mathcal{L}(L)$.*

Proof The proof is the straightforward adaptation of the proof of [53, Lemma 2.6]. We report it here for convenience of the reader. We prove the proposition in the case $r_1 r_2 > \mathrm{lk}(L)^2$ and $r_1 > 0$, $r_2 > 0$. The other case is analogous. We consider the manifold $Y = S^3_{r_1,\bullet}$. We have that $r_2 \in \mathcal{L}(Y)$ and since the components of L are unknotted it follows that also $\infty \in \mathcal{L}(L)$. In fact $S^3_{r_1,\infty}(L)$ is a lens space, and hence an L-space. Thus we can deduce, by virtue of Theorem 4.1, that the interval between r_2 and ∞ that does not contain the homological longitude is contained in $\mathcal{L}(Y)$. The homological longitude $\frac{\mathrm{lk}(L)^2}{r_1}$ is smaller than r_2, so we deduce that $[r_2, \infty] \cap \overline{\mathbb{Q}} \subset \mathcal{L}(Y)$. In other words we have proved that $S^3_{r_1,s}(L)$ is an L-space for all $s \geq r_2$. Now we fix $s \geq r_2$ and consider the manifold $Y_s = S^3_{\bullet,s}$. As a consequence of r_1 and ∞ belonging to $\mathcal{L}(Y_s)$, we can apply again Theorem 4.1 and deduce that the interval between r_1 and ∞ that does not contain the homological longitude is contained in $\mathcal{L}(Y_s)$. Since $r_1 \geq \frac{\mathrm{lk}(L)^2}{r_2} \geq \frac{\mathrm{lk}(L)^2}{s}$ and the last term in the chain of inequalities is the homological longitude of Y_s, we conclude that $[r_1, \infty] \cap \overline{\mathbb{Q}} \subset \mathcal{L}(Y_s)$ for all $s \geq r_2$. This is exactly equivalent to saying that $\big([r_1, \infty] \times [r_2, \infty]\big) \cap \overline{\mathbb{Q}}^2 \subset \mathcal{L}(L)$. A pictorial sketch of the proof is described in Figure 32. □

We are now ready to prove the main result of this section. Recall that since two-bridge links have unknotted components, the rational homology spheres associated to surgery coefficients (r, s) where at least one of r and s is ∞ are all L-spaces. For this reason, we will only consider the set $\mathcal{L}(L_n) \cap \mathbb{Q}^2$.

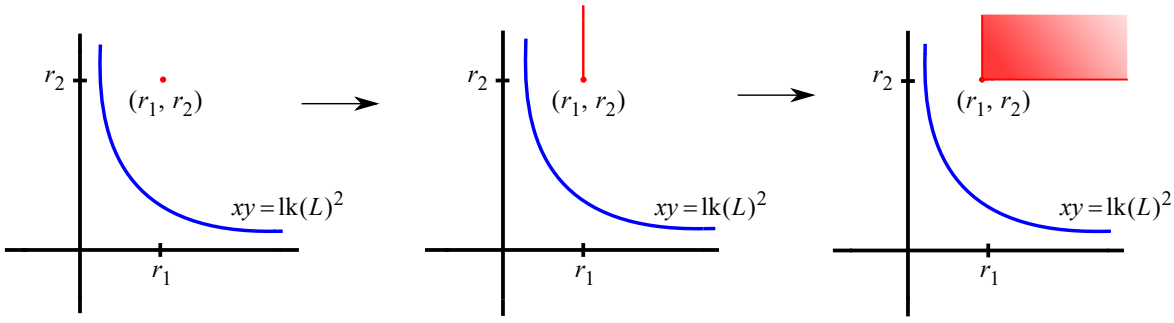


Figure 32: A pictorial sketch of the proof.

Proposition 4.3 *Let L_n be the link described in Figure 1. Then $\mathcal{L}(L_n) \cap \mathbb{Q}^2 = [n, +\infty)^2 \cap \mathbb{Q}^2$.*

Proof Since L-space do not support taut foliations, it follows from Proposition 3.20 that it suffices to prove that

$$[n, +\infty)^2 \subset \mathcal{L}(L_n).$$

The link L_n satisfies $\mathrm{lk}(L_n)^2 = (n-1)^2$ and its components are unknotted, hence by Proposition 4.2 it is enough to prove that $(n, n) \in \mathcal{L}(L_n)$. We can see the links L_n as surgeries on a three-component link $\mathscr{L}$, as represented in Figure 33. We have also fixed an orientation of this link, that we will use later in the proof.

More precisely we have that $S^3_{a,b,-1/(n-1)}(\mathscr{L}) = S^3_{a+n-1,b+n-1}(L_n)$. This implies that the statement is equivalent to proving that $S^3_{1,1,-1/(n-1)}(\mathscr{L})$ is an L-space for all $n \geq 1$ and to prove this we will apply Theorem 4.1 to the rational homology solid torus $S^3_{1,1,\bullet}(\mathscr{L})$. Denoting this manifold by Y, we have:

• $\underline{\infty \in \mathcal{L}(Y)}$: In fact $S^3_{1,1,\infty}(\mathscr{L})$ is $(1,1)$-surgery on the Whitehead link. This is the Poincaré homology sphere and manifolds with finite fundamental group are L-spaces [48, Proposition 2.2].

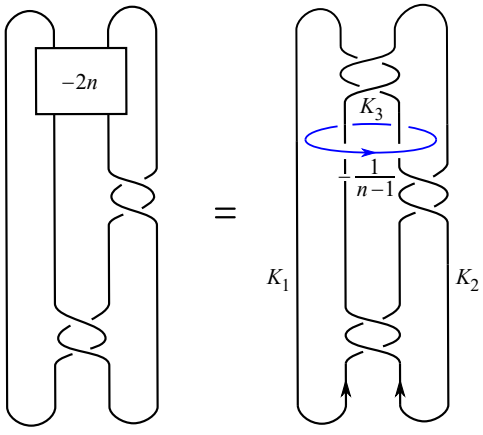


Figure 33: How to obtain the links $\{L_n\}_{n\geq 1}$ as surgeries on a 3-component link $\mathscr{L}$.

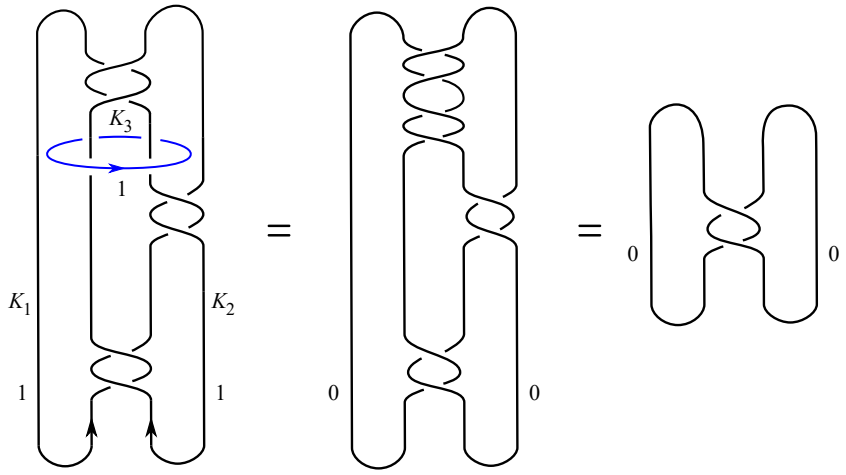


Figure 34: The $(1,1,1)$-surgery on $\mathscr{L}$ is $(0,0)$-surgery on the Hopf link.

• $1 \in \mathcal{L}(Y)$: In fact $S^3_{1,1,1}(\mathscr{L})$ is $(0,0)$-surgery on the Hopf link, see Figure 34. This manifold is S^3 and therefore an L-space.

• The homological longitude of Y is the slope 2: To prove this we have to do a simple computation. We fix an orientation for the link and we denote the components of $\mathscr{L}$ with K_1, K_2 and K_3 as in Figure 33.

We have that $\mathrm{lk}(K_1, K_3) = \mathrm{lk}(K_2, K_3) = 1$ and $\mathrm{lk}(K_1, K_2) = 0$. Consequently, a presentation matrix for $H_1(S^3_{1,1,p/q}(\mathscr{L}), \mathbb{Z})$ is given by

$$A = \begin{pmatrix} 1 & 0 & q \\ 0 & 1 & q \\ 1 & 1 & p \end{pmatrix}$$

and in particular $S^3_{1,1,p/q}(\mathscr{L})$ is not a rational homology sphere if and only if the determinant of A is zero. This happens if and only if $p = 2q$ and therefore 2 is the homological longitude of the manifold $S^3_{1,1,\bullet}(\mathscr{L})$. What we have proved implies by Theorem 4.1 that $(-\infty, 1] \cap \mathbb{Q} \subset \mathcal{L}(Y)$. In particular $S^3_{1,1,-1/(n-1)}(\mathscr{L})$ is an L-space for all $n \geq 1$ and this manifold is exactly the (n,n)-surgery on L_n. □

Remark 4.4 In the terminology of [21], the links L_n are *L-space links*. Liu [41] conjectured that a two-bridge link is an L-space link if and only if is of the form $b(pq-1,-q)$, where p and q are odd positive integers. This conjecture was proved by Dawra [10]. It is not difficult to prove that the link L_n, as an unoriented link, is isotopic to $b(6n+2,-3)$.

Acknowledgements

I warmly thank my advisors Bruno Martelli and Paolo Lisca for their support and for their useful comments on a first draft of the paper. I also thank the anonymous referees for comments and suggestions, which improved the clarity and overall organisation of the paper. The author is supported by the FWF project P 34318 "Cut and Paste Methods in Low Dimensional Topology".

References

[1] **S Baader**, **C Graf**, *Fibred links in S^3*, Expo. Math. 34:4 (2016) 423–435 MR

[2] **J Bowden**, *Approximating C^0-foliations by contact structures*, Geom. Funct. Anal. 26:5 (2016) 1255–1296 MR

[3] **S Boyer**, **A Clay**, *Foliations, orders, representations, L-spaces and graph manifolds*, Adv. Math. 310 (2017) 159–234 MR

[4] **S Boyer**, **C M Gordon**, **L Watson**, *On L-spaces and left-orderable fundamental groups*, Math. Ann. 356:4 (2013) 1213–1245 MR

[5] **M Brittenham**, **R Naimi**, **R Roberts**, *Graph manifolds and taut foliations*, J. Differential Geom. 45:3 (1997) 446–470 MR

[6] **G Burde**, **H Zieschang**, *Knots*, 2nd edition, De Gruyter Studies in Mathematics 5, Walter de Gruyter, Berlin (2003) MR

[7] **A Cavallo**, **B Liu**, *Fibered and strongly quasi-positive L-space links*, Michigan Math. J. 73:1 (2023) 209–224 MR

[8] **A Clay**, **T Lidman**, **L Watson**, *Graph manifolds, left-orderability and amalgamation*, Algebr. Geom. Topol. 13:4 (2013) 2347–2368 MR

[9] **M Culler**, **N M Dunfield**, *Orderability and Dehn filling*, Geom. Topol. 22:3 (2018) 1405–1457 MR

[10] **N Dawra**, *On the link Floer homology of L-space links* (2015) arXiv 1505.01100

[11] **C Delman**, **R Roberts**, *Taut foliations from double-diamond replacements*, from "Characters in low-dimensional topology" (O Collin, S Friedl, C Gordon, S Tillmann, L Watson, editors), Contemp. Math. 760, Amer. Math. Soc., Providence, RI (2020) 123–142 MR

[12] **C Delman**, **R Roberts**, *Persistently foliar composite knots*, Algebr. Geom. Topol. 21:6 (2021) 2761–2798 MR

[13] **N M Dunfield**, **J Rasmussen**, *A unified Casson–Lin invariant for the real forms of* SL(2), Geom. Topol. 29:8 (2025) 4055–4188 MR

[14] **D Eisenbud**, **U Hirsch**, **W Neumann**, *Transverse foliations of Seifert bundles and self-homeomorphism of the circle*, Comment. Math. Helv. 56:4 (1981) 638–660 MR

[15] **D Eisenbud**, **W Neumann**, *Three-dimensional link theory and invariants of plane curve singularities*, Annals of Mathematics Studies 110, Princeton Univ. Press (1985) MR

[16] **W Floyd**, **U Oertel**, *Incompressible surfaces via branched surfaces*, Topology 23:1 (1984) 117–125 MR

[17] **D Gabai**, *The Murasugi sum is a natural geometric operation, II*, from "Combinatorial methods in topology and algebraic geometry" (Rochester, NY, 1982) (J R Harper, R Mandelbaum, editors), Contemp. Math. 44, Amer. Math. Soc., Providence, RI (1985) 93–100 MR

[18] **D Gabai**, **W H Kazez**, *Pseudo-Anosov maps and surgery on fibred* 2*-bridge knots*, Topology Appl. 37:1 (1990) 93–100 MR

[19] **D Gabai**, **U Oertel**, *Essential laminations in* 3*-manifolds*, Ann. of Math. (2) 130:1 (1989) 41–73 MR

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

[21] **E Gorsky**, **A Némethi**, *Links of plane curve singularities are L-space links*, Algebr. Geom. Topol. 16:4 (2016) 1905–1912 MR

[22] **J Hanselman**, **J Rasmussen**, **S D Rasmussen**, **L Watson**, *L-spaces, taut foliations, and graph manifolds*, Compos. Math. 156:3 (2020) 604–612 MR

[23] **M Hedden**, *Notions of positivity and the Ozsváth–Szabó concordance invariant*, J. Knot Theory Ramifications 19:5 (2010) 617–629 MR

[24] **K Honda**, **W H Kazez**, **G Matić**, *Right-veering diffeomorphisms of compact surfaces with boundary*, Invent. Math. 169:2 (2007) 427–449 MR

[25] **Y Hu**, *Euler class of taut foliations and Dehn filling*, Comm. Anal. Geom. 31:7 (2023) 1749–1782 MR

[26] **M Jankins**, **W D Neumann**, *Rotation numbers of products of circle homeomorphisms*, Math. Ann. 271:3 (1985) 381–400 MR

[27] **A Juhász**, *A survey of Heegaard Floer homology*, from "New ideas in low dimensional topology" (L H Kauffman, V O Manturov, editors), Ser. Knots Everything 56, World Sci., Hackensack, NJ (2015) 237–296 MR

[28] **T Kalelkar**, **R Roberts**, *Taut foliations in surface bundles with multiple boundary components*, Pacific J. Math. 273:2 (2015) 257–275 MR

[29] **A Kawauchi**, *A survey of knot theory*, Birkhäuser, Basel (1996) MR

[30] **W H Kazez**, **R Roberts**, *C^0 approximations of foliations*, Geom. Topol. 21:6 (2017) 3601–3657 MR

[31] **D Krcatovich**, *The reduced knot Floer complex*, Topology Appl. 194 (2015) 171–201 MR

[32] **S Krishna**, *Taut foliations, positive 3-braids, and the L-space conjecture*, J. Topol. 13:3 (2020) 1003–1033 MR

[33] **P Kronheimer**, **T Mrowka**, **P Ozsváth**, **Z Szabó**, *Monopoles and lens space surgeries*, Ann. of Math. (2) 165:2 (2007) 457–546 MR

[34] **M Lackenby**, *The volume of hyperbolic alternating link complements*, Proc. London Math. Soc. (3) 88:1 (2004) 204–224 MR

[35] **T Li**, *Laminar branched surfaces in 3-manifolds*, Geom. Topol. 6 (2002) 153–194 MR

[36] **T Li**, *Boundary train tracks of laminar branched surfaces*, from "Topology and geometry of manifolds" (Athens, GA, 2001) (G Matić, C McCrory, editors), Proc. Sympos. Pure Math. 71, Amer. Math. Soc., Providence, RI (2003) 269–285 MR

[37] **T Li**, *Taut foliations of 3-manifolds with Heegaard genus 2*, Duke Math. J. 173:8 (2024) 1427–1475 MR

[38] **T Li**, **R Roberts**, *Taut foliations in knot complements*, Pacific J. Math. 269:1 (2014) 149–168 MR

[39] **P Lisca**, **A I Stipsicz**, *Ozsváth–Szabó invariants and tight contact three-manifolds, I*, Geom. Topol. 8 (2004) 925–945 MR

[40] **P Lisca**, **A I Stipsicz**, *On the existence of tight contact structures on Seifert fibered 3-manifolds*, Duke Math. J. 148:2 (2009) 175–209 MR

[41] **Y Liu**, *L-space surgeries on links*, Quantum Topol. 8:3 (2017) 505–570 MR

[42] **W Menasco**, *Closed incompressible surfaces in alternating knot and link complements*, Topology 23:1 (1984) 37–44 MR

[43] **C Millichap**, **W Worden**, *Hidden symmetries and commensurability of 2-bridge link complements*, Pacific J. Math. 285:2 (2016) 453–484 MR

[44] **R Naimi**, *Foliations transverse to fibers of Seifert manifolds*, Comment. Math. Helv. 69:1 (1994) 155–162 MR

[45] **Y Ni**, *Knot Floer homology detects fibred knots*, Invent. Math. 170:3 (2007) 577–608 MR

[46] **U Oertel**, *Incompressible branched surfaces*, Invent. Math. 76:3 (1984) 385–410 MR

[47] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and genus bounds*, Geom. Topol. 8 (2004) 311–334 MR

[48] **P Ozsváth**, **Z Szabó**, *On knot Floer homology and lens space surgeries*, Topology 44:6 (2005) 1281–1300 MR

[49] **R C Penner**, **J L Harer**, *Combinatorics of train tracks*, Annals of Mathematics Studies 125, Princeton Univ. Press (1992) MR

[50] **J Rasmussen**, **S D Rasmussen**, *Floer simple manifolds and L-space intervals*, Adv. Math. 322 (2017) 738–805 MR

[51] **R Roberts**, *Taut foliations in punctured surface bundles, I*, Proc. London Math. Soc. (3) 82:3 (2001) 747–768 MR

[52] **R Roberts**, *Taut foliations in punctured surface bundles, II*, Proc. London Math. Soc. (3) 83:2 (2001) 443–471 MR

[53] **D Santoro**, *L-spaces, taut foliations and the Whitehead link*, Algebr. Geom. Topol. 24:6 (2024) 3455–3502 MR

[54] **H Schubert**, *Knoten mit zwei Brücken*, Math. Z. 65 (1956) 133–170 MR

[55] **W P Thurston**, *Three-dimensional geometry and topology, I*, Princeton Math. Ser. 35, Princeton Univ. Press (1997) MR

DIEGO SANTORO diego.santoro95@gmail.com
University of Vienna, Vienna, Austria

Received: September 14, 2024 Revised: March 4, 2025

msp

Algebraic & Geometric Topology **26**:3 (2026) 1155–1193
DOI: 10.2140/agt.2026.26.1155
Published: April 1, 2026

New results on tilings via cup products and Chern characters on tiling spaces

JIANLONG LIU, JONATHAN ROSENBERG AND RODRIGO TREVIÑO

We study the cohomology rings of tiling spaces Ω given by cubical substitutions. While there have been many calculations before of cohomology *groups* of such tiling spaces, the innovation here is that we use computer-assisted methods to compute the cup-product structure. This leads to examples of substitution tilings with isomorphic cohomology groups but different cohomology rings. Part of the interest in studying the cup product comes from Bellissard's *gap-labeling conjecture*, which is known to hold in dimensions ≤ 3, but where a proof is known in dimensions ≥ 4 only when the Chern character from $K^0(\Omega)$ to $H^*(\Omega, \mathbb{Q})$ lands in $H^*(\Omega, \mathbb{Z})$. Computation of the cup product on cohomology often makes it possible to compute the Chern character. We introduce a natural generalization of the gap-labeling conjecture, called the *equivariant gap-labeling conjecture*, which applies to tilings with a finite symmetry group. Again this holds in dimensions ≤ 3, but we are able to show that it *fails* in general in dimensions ≥ 4. This, plus some of our cup-product calculations, makes it plausible that the gap-labeling conjecture might fail in high dimensions.

1 Introduction

This paper is about tiling spaces Ω coming from primitive substitution tilings of $\mathbb{R}^d$. All the tilings in this paper will be self-similar and have finite local complexity. Under those hypotheses, the action of $\mathbb{R}^d$ on Ω by translations is minimal and uniquely ergodic [30]. We begin with a review of the basic definitions and literature on tilings in Section 2.

The problems that we will study here were motivated by the gap-labeling conjecture of Bellissard [7]. His idea was that in a quasicrystal (an almost periodic physical system modeled mathematically by a tiling), physical properties of the material are determined by the spectra of almost periodic Schrödinger operators which live in a C^*-algebra $\mathcal{A}_p(\Omega)$ determined by the tiling, so that gaps in the spectra can be determined by computing $K_0(\mathcal{A}_p(\Omega))$ and the map from this K-group to the reals induced by the trace. The algebra $\mathcal{A}_p(\Omega)$ is highly noncommutative and thus complicated to deal with, but it contains a natural copy of the commutative algebra $C(\Omega)$ of functions on the tiling space. Bellissard noticed that at least in many cases, the K-theory of $\mathcal{A}_p(\Omega)$ all lies in the image of the K-theory of this subalgebra, which can be computed purely topologically. The gap-labeling conjecture asserts that this should always be the case. Several papers [8; 10; 16] claimed to prove the conjecture, but they all implicitly assumed that the top-dimensional piece of the Chern character $\mathrm{ch}_d : K^{-d}(\Omega) \to \check{H}^d(\Omega; \mathbb{Q})$ gives an isomorphism $K^{-d}(\Omega) \xrightarrow{\cong} \check{H}^d(\Omega; \mathbb{Z})$. (See [1, §9] for a detailed discussion.)

MSC2020: 19L47, 19L64, 37B52, 55N45.

Our goal in this paper is to compute the structure of the integral cohomology ring $H^*(\Omega;\mathbb{Z})$ for several examples of cubical substitution tilings. This is computation-intensive and requires computer assistance for the calculation of the cup product. Calculations are done in Sections 3 and 4. In several cases, we are also able to determine at least an important part of the Chern character map $\mathrm{ch}_d : K^{-d}(\Omega) \to \check{H}^d(\Omega;\mathbb{Q})$. Among the major results are:

(i) There are cases of tiling spaces with isomorphic cohomology as groups but not as *rings* (Section 3.2).

(ii) There are cases where the Chern character does *not* give an isomorphism $K^{-d}(\Omega) \to \check{H}^d(\Omega;\mathbb{Z})$ (Section 4.1).

Unfortunately we are not able with the present techniques to produce a counterexample to the gap-labeling conjecture in high dimensions, though we would not be surprised if such a counterexample exists.

However, we show that there is a natural generalization of the gap-labeling conjecture, which we call the *equivariant gap-labeling conjecture*, that can be formulated for tiling spaces with a finite symmetry group. Roughly speaking, this involves replacing K-theory by equivariant K-theory. This conjecture is formulated and discussed in Sections 2.5 and 2.6. We show that this equivariant conjecture fails in general when $d \geq 4$ (see Theorem 2.6 and Section 4.3), which provides additional evidence that the original gap-labeling conjecture probably fails in high dimensions.

2 Tiling spaces and their cohomology

This section covers the necessary background on tiling spaces and their topological invariants.

2.1 Basics of tilings

To define a tiling, we start by assuming we are given a finite set of polytopes in $\mathbb{R}^d$ called *prototiles*. A *tile* is a translate $t+\tau$, $\tau \in \mathbb{R}^d$, of a prototile t, along with a choice of color or label $c \in \mathrm{Col}$ chosen from a set Col of *colors*. We will always assume that Col is finite. If Col has only one element, a common scenario, then one can ignore the color. A *tiling* $\mathcal{T}$ of $\mathbb{R}^d$ is a decomposition of $\mathbb{R}^d$ as a union of tiles which only intersect along their faces. We write $\mathbb{R}^d = \bigcup_j t_j$, where each t_j is the *support of the tile* (t_j, c_j), and $c_j \in \mathrm{Col}$ is the color of the tile. In a slight abuse of notation, we will often fail to distinguish between tiles and their supports, but this should not cause confusion from context. A *patch* of $\mathcal{T}$ is the union of finitely many tiles. Two tiles t_1, t_2 of a tiling $\mathcal{T}$ are *translation-equivalent* if there exists a $\tau \in \mathbb{R}^d$ such that $t_1 = t_2 - \tau$, i.e., they are translations of the same prototile, and their labels are equal. Two patches P_1, P_2 of $\mathcal{T}$ are translation equivalent if there is a $\tau \in \mathbb{R}^d$ such that $P_1 = P_2 - \tau$. The translation ϕ_τ of tiles and patches by vectors $\tau \in \mathbb{R}^d$ extends to all of a tiling $\mathcal{T}$, and we denote by $\phi_\tau(\mathcal{T})$ the tiling consisting of the union of tiles of the form $\phi_\tau(t)$, where t is a tile of $\mathcal{T}$.

Let $R_* > 0$ be the smallest number so that any ball of radius R_* in $\mathbb{R}^d$ contains a tile of $\mathcal{T}$. For any $R \geq R_*$, an R-patch of $\mathcal{T}$ is the largest patch of $\mathcal{T}$ completely contained in a fixed ball of radius R. The

tiling $\mathcal{T}$ has *finite local complexity* or *FLC* if, given any $R > 0$, the set of equivalence classes (under translation) of R-patches of the family of tilings $\{\phi_t(\mathcal{T})\}_{t\in\mathbb{R}^d}$ is finite. All tilings considered in this paper will have FLC.

Given a tiling $\mathcal{T}$, define the metric d on the set of all translates $\{\phi_t(\mathcal{T})\}_{t\in\mathbb{R}^d}$ as follows. Set

$$d'(\phi_t(\mathcal{T}),\phi_s(\mathcal{T})) = \inf\big\{\epsilon > 0 : \text{there is a } \tau \in B_\epsilon(0) \text{ such that the } \epsilon^{-1}\text{-patch of } \phi_{t+\tau}(\mathcal{T}) \text{ is equal to the } \epsilon^{-1}\text{-patch of } \phi_s(\mathcal{T})\big\}$$

and define

$$d(\phi_t(\mathcal{T}),\phi_s(\mathcal{T})) = \min\{1, d'(\phi_t(\mathcal{T}),\phi_s(\mathcal{T}))\}.$$

This is a metric [30]. The *(translational) tiling space* $\Omega_\mathcal{T}$ of $\mathcal{T}$ is the metric completion of the set of translates of $\mathcal{T}$ with respect to the metric above, i.e.,

$$\Omega_\mathcal{T} := \overline{\{\phi_t(\mathcal{T}) : t \in \mathbb{R}^d\}}.$$

If $\mathcal{T}$ is FLC, then $\Omega_\mathcal{T}$ is a compact metric space, with local product structure $B \times C$, where B is a Euclidean ball and C is a zero-dimensional compact space (usually, but not always, a Cantor set). The translation of $\mathcal{T}$ by vectors of $\mathbb{R}^d$ extends to an action of $\mathbb{R}^d$ by translations.

Let us mention two more topological spaces associated to $\mathcal{T}$, where one incorporates rotations in addition to translations, that is, $\mathrm{SE}(d) = \mathbb{R}^d \rtimes \mathrm{SO}(d)$. For details, see [27, §4.1–4.5]. The metric is first relaxed to allow ϵ-rotations, then one can define

$$\Omega_\mathcal{T}^{\mathrm{rot}} = \overline{\{\phi(\mathcal{T}) : \phi \in \mathrm{SE}(d)\}}.$$

It has a canonical transversal

$$\Omega_\mathcal{T}^0 = \Omega_\mathcal{T}^{\mathrm{rot}}/\mathrm{SO}(d) \cong \Omega_\mathcal{T}/\{\text{rotational symmetries}\},$$

which we call the *rotational tiling space* of $\mathcal{T}$.

2.2 Substitution tilings

Let $(t_1, c_1), \dots, (t_m, c_m) \subset \mathbb{R}^d \times \mathrm{Col}$ be compact, connected sets with nonempty interior, where Col is a finite set. If $A \subset \mathrm{GL}(d,\mathbb{R})$ is an expanding matrix, then a *substitution rule on* $t_1, \dots, t_m$ *with expansion* A is a decomposition of At_i as

$$At_i = \bigcup_{j=1}^{m} \bigcup_{\tau\in\Lambda_{ij}} t_j - \tau$$

for any i, where each Λ_{ij} is a finite set. The *substitution matrix* corresponding to this substitution rule is the doubly indexed collection of numbers S defined by $S_{ij} = |\Lambda_{ij}|$. The substitution is *primitive* if there is a $k \in \mathbb{N}$ such that the substitution matrix S^k has all positive entries. Any substitution rule can be

iterated: the set $A^2 t_i$ can be tiled as

$$A^2 t_i = \bigcup_{j=1}^{m} \bigcup_{\tau \in \Lambda_{ij}} A t_j - A\tau,$$

where At_j can be decomposed as above. A tiling $\mathcal{T}$ is given by a substitution rule with expansion A if any patch of $\mathcal{T}$ is translation equivalent to a subpatch of $A^k t_i$ for some $k \in \mathbb{N}$, $1 \le i \le m$, where $A^k t_i$ is a tiled set as above. In that case, a patch which is translation-equivalent to one of the form $A^k t_i$ as above is a *level-k supertile*.

Let $\mathcal{T}$ be a FLC tiling with prototiles $t_1, \dots, t_m$. For a tile $t \in \mathcal{T}$, let $\mathcal{T}(t)$ be the patch consisting of all tiles of $\mathcal{T}$ which intersect t. This type of patch is called a *collared tile* of $\mathcal{T}$. Now consider the product $\Omega_{\mathcal{T}} \times \mathbb{R}^d$ with the product topology, where $\Omega_{\mathcal{T}}$ has the discrete topology and $\mathbb{R}^d$ the traditional topology. Let $\sim_1$ be the equivalence relation on this product which declares $(\mathcal{T}_1, u_1) \sim_1 (\mathcal{T}_2, u_2)$ if $\mathcal{T}_1(t_1) - u_1 = \mathcal{T}_2(t_2) - u_2$ for some tiles t_1, t_2 with $u_i \in t_i \in \mathcal{T}_i$. The *Anderson–Putnam* complex of $\Omega_{\mathcal{T}}$ is the set $\Gamma_{\mathcal{T}} := (\Omega_{\mathcal{T}} \times \mathbb{R}^d)/\sim_1$. Let $\pi_1 : \Omega_{\mathcal{T}} \times \mathbb{R}^d \to \Gamma_{\mathcal{T}}$ be the quotient map. The space $\Gamma_{\mathcal{T}}$ is a flat branched manifold.

If $\mathcal{T}$ is a primitive substitution tiling then there exists a locally expanding affine map $\gamma : \Gamma_{\mathcal{T}} \to \Gamma_{\mathcal{T}}$ such that there is a homeomorphism of tiling spaces

$$\Omega_{\mathcal{T}} \cong \varprojlim(\Gamma_{\mathcal{T}}, \gamma), \tag{1}$$

which is the seminal result of Anderson and Putnam [2]. In this case [30], the action of $\mathbb{R}^d$ by translations is *uniquely ergodic*: there is a unique $\mathbb{R}^d$-invariant ergodic probability measure on Ω, which will be denoted by μ. By the local product structure of Ω, this measure is locally of the form Leb $\times\, \nu$, where Leb is the Lebesgue measure on the local Euclidean component and ν is a measure on the local Cantor component and will be discussed further in Section 2.4.

Remark 2.1 The reader may wonder why $\Omega_{\mathcal{T}}$ may not be recovered as the inverse limit obtained by the substitution rule directly on tiles and not necessarily on collared tiles, as required for the construction of the Anderson–Putnam complex $\Gamma_{\mathcal{T}}$ above. Although an inverse limit would be well defined in such case, it may not necessarily be homeomorphic to the tiling space $\Omega_{\mathcal{T}}$. This happens if and only if the substitution rule *forces the border* (see [2] for a proof of this result). To quote [27], "A substitution rule *forces the border* if there exists a positive integer n such that any two level-n supertiles of the same type have the same pattern of neighboring tiles". Although the concept of forcing the border will not be central in this paper, it will be mentioned in a couple of the examples below, and we direct the reader to [2; 27] to learn why in that case one can use uncollared tiles to form the desired inverse limit.

Definition 2.2 A *cubical substitution* in dimension d with m colors with expansion $\lambda \in \mathbb{N}$ is a substitution rule on $[0,1]^d \times \{1, \dots, m\}$ with expansion matrix $A = \lambda \cdot \mathrm{Id}$.

Note that there are $m^{\lambda^d + 1}$ different cubical substitution rules in dimension d with m colors and expansion λ.

2.3 Topological invariants

Suppose that $\mathcal{T}$ is a primitive substitution tiling. The Anderson–Putnam homeomorphism (1) allows us to express the cohomology and K-theory of $\Omega = \Omega_{\mathcal{T}}$ in digestible terms:

$$\check{H}^*(\Omega;\mathbb{Z}) = \varinjlim(\check{H}^*(\Gamma;\mathbb{Z}), \gamma^*) \quad \text{and} \quad K^*(\Omega) = \varinjlim(K^*(\Gamma), \gamma^*). \tag{2}$$

Recall that the Chern character $\mathrm{ch} : K^*(\Omega) \to \check{H}^*(\Omega;\mathbb{Q})$ is defined by

$$\mathrm{ch}([L]) = \sum_{k\geq 0} \frac{c_1([L])^k}{k!}, \tag{3}$$

for L a line bundle on Ω and $[L]$ its class in K-theory, where $c_1([L])$ is the first Chern class of $[L]$, and then extended to all virtual vector bundles using the "splitting principle" (which says that one can pretend all vector bundles split into direct sums of line bundles). This is a ring homomorphism, and after tensoring the domain with $\mathbb{Q}$, it becomes a rational isomorphism of rings. (Strictly speaking, we have only defined the Chern character on even-degree K-theory, with values in even-degree cohomology, but it extends by suspension to odd degrees as well.) This isomorphism in fact is derived from the isomorphism $\mathrm{ch} : K^*(\Gamma)\otimes\mathbb{Q} \xrightarrow{\cong} \check{H}^*(\Gamma;\mathbb{Q})$ at the AP complex level and (2). When the dimension of Γ is at least four, then because of the denominators in (3), ch does not necessarily come from an integral isomorphism between $K^*(\Gamma)$ and $H^*(\Gamma;\mathbb{Z})$. (It does give such an isomorphism for spheres and tori — see, for example, [14, Proposition 4.3] — but not for $\mathbb{C}\mathbb{P}^k$, $k \geq 2$ [14, §4.1].)

In Section 4 we give an explicit example of how the Chern character can fail to be an integral isomorphism for tiling spaces. This gives a negative answer to a question of Benameur and Mathai [9, p. 8].

2.4 Frequencies, traces and the gap-labeling conjecture

Suppose $\mathcal{T}$ is a primitive substitution tiling and $P \subset \mathcal{T}$ is a patch. Let us make some comments about patches.

First, the patch has a *frequency*: let $\mathrm{freq}_R(P,\mathcal{T})$ be the number of copies of P found (up to equivalence) as subpatches of $\mathcal{T}$ which are completely contained in a ball of radius R around the origin. Since the tiling came from a primitive substitution, the limit $R^{-d}\mathrm{freq}_R(P,\mathcal{T}) \to \mathrm{freq}(P,\mathcal{T}) \geq 0$ exists as $R\to\infty$, and it is called the *patch frequency* of P. Since there is a unique $\mathbb{R}^d$-invariant measure on $\Omega_{\mathcal{T}}$, this limit is actually independent of $\mathcal{T}$, meaning that the patch frequency of the patch P is the same for any tiling in Ω. This will be denoted by $\mathrm{freq}(P)$.

Secondly, if P is a patch, then P is given by a closed subset $\bar{P} \subset \Gamma$ and a $k \geq 0$ in that the set $\pi_k^{-1}(\bar{P}) \subset \Omega$ is isometric to $P \times \mathcal{C}_P$, where $\mathcal{C}_P$ is a (transversal) Cantor set. It follows from Birkhoff's ergodic theorem that $\nu(\mathcal{C}_P) = \mathrm{freq}(P)$, and for this reason ν is sometimes referred to as the *frequency measure*.

Finally, if P is a patch, then it defines a cohomology class $[P] \in \check{H}^d(\Omega;\mathbb{Z})$ as follows. Recall that there exists a closed set $\bar{P} \subset \Gamma$ such that $\pi_k^{-1}(\bar{P}) \subset \Omega$ is isometric to $P\times\mathcal{C}_P$. Then the interior of $\bar{P}$ in Γ is an

open set with a Čech cohomology class $[\bar{P}] \in \check{H}^d(\Gamma;\mathbb{Z})$ which by (2) defines a class $[P] \in \check{H}^d(\Omega;\mathbb{Z})$. The map $[P] \mapsto \mathrm{freq}(P) = \nu(\mathcal{C}_P)$ turns out to give a homomorphism $C_\mu : \check{H}^d(\Omega;\mathbb{Z}) \to \mathbb{R}$ called the *Schwartzman asymptotic cycle* or the *Ruelle–Sullivan current* [18]. The image $C_\mu(\check{H}^d(\Omega;\mathbb{Z})) \subset \mathbb{R}$ is called the *frequency module*.

Associated with any substitution tiling space there are several C^*-algebras that one can define (see [17] for details). The principal one is $\mathcal{A}_p(\Omega)$, the *unstable (punctured) algebra*. Another important one is $\mathcal{A}_{\mathrm{AF}}(\Omega)$, the AF-algebra constructed from the substitution matrix defining the tiling space Ω. Since the substitution is assumed to be primitive, it comes with a unique tracial state $\tau_{\mathrm{AF}} : K_0(\mathcal{A}_{\mathrm{AF}}(\Omega)) \to \mathbb{R}$. For a primitive substitution tiling, there is also a unique tracial state $\tau : K_0(\mathcal{A}_p(\Omega)) \to \mathbb{R}$ which can be recovered from the unique $\mathbb{R}^d$-invariant probability measure μ on the tiling space Ω.

How the images $C_\mu(\check{H}^d(\Omega;\mathbb{Z}))$ and $\tau(K_0(\mathcal{A}_p(\Omega)))$ overlap is a delicate matter. The conjecture that they are equal was proposed by Bellissard [7], and went under the name of the *gap-labeling conjecture*. There have been three papers claiming to have proved this equality [8; 10; 16]; we now briefly review what is at stake. As pointed out in [1], several of these papers invoke arguments which rely on the commutative diagram

$$\begin{array}{ccc} K_0(\mathcal{A}_p(\Omega)) \otimes \mathbb{Q} & \xrightarrow{\mathrm{ch}_d \circ \chi^{-1}} & \check{H}^d(\Omega;\mathbb{Q}) \\ \downarrow{\scriptstyle \tau} & & \downarrow{\scriptstyle C_\mu} \\ \mathbb{R} & \xrightarrow{\quad\mathrm{Id}\quad} & \mathbb{R} \end{array} \tag{4}$$

where $\chi : K^{-d}(\Omega) \to K_0(\mathcal{A}_p(\Omega))$ is an isomorphism, $\mathrm{ch} : K^*(\Omega) \otimes \mathbb{Q} \to \check{H}^*(\Omega;\mathbb{Z})$ is the isomorphism given by the Chern character in (3), and ch_d is its projection to $\check{H}^d(\Omega;\mathbb{Q})$. The arguments used in several papers have the gap that they assume that ch_d factors through $\check{H}^d(\Omega;\mathbb{Z})$ to give the diagram

$$\begin{array}{ccc} K_0(\mathcal{A}_p(\Omega)) & \xrightarrow{\mathrm{ch}_d \circ \chi^{-1}} & \check{H}^d(\Omega;\mathbb{Z}) \\ \downarrow{\scriptstyle \tau} & & \downarrow{\scriptstyle C_\mu} \\ \mathbb{R} & \xrightarrow{\quad\mathrm{Id}\quad} & \mathbb{R} \end{array} \tag{5}$$

Although this may hold in low dimensions, in Section 4 we give an explicit example where this is not true, and thus show that any argument to prove the gap-labeling conjecture using this isomorphism is incomplete.

2.5 Equivariant gap labeling

In this subsection we mention an equivariant variant of the gap-labeling conjecture which is stronger, hence more likely to fail. In fact, we produce a counterexample, though we still show that the conjecture is true in low dimensions. The reason for mentioning this equivariant conjecture here is that our methods might be usable in some other cases to disprove this equivariant conjecture, even if they do not disprove the original conjecture. Suppose $\mathcal{T}$ is a primitive substitution tiling as before, and furthermore assume that

there is a finite group G of symmetries that acts on the associated tiling space Ω. (Equivariant K-theory works just as well with G compact Lie, but that case isn't relevant for substitution tilings.) This group G will of course act on $\mathcal{A}_p(\Omega)$ and preserve both the unique invariant measure on Ω and the unique tracial state $\tau : K_0(\mathcal{A}_p(\Omega)) \to \mathbb{R}$. We have a natural inclusion map on *equivariant* K-theory

$$K_G^0(\Omega) = K_0^G(C(\Omega)) \to K_0^G(\mathcal{A}_p(\Omega)).$$

The *equivariant gap-labeling conjecture* asserts that this map induces an isomorphism on all trace invariants. But there are far more of these than in the nonequivariant case. To explain this, recall that equivariant K-theory is a module over $R(G)$, the representation ring of G, which after tensoring with $\mathbb{C}$ can be identified with the ring of linear combinations of characters (of irreducible representations of G), or with the algebra $Z(G)$ of class functions on G (this is the center of the group ring $\mathbb{C}G$). If A is a (unital, for simplicity) C^*-algebra equipped with an action of G, then $K_0^G(A)$ is the Grothendieck group of equivalence classes of G-invariant self-adjoint projections $e \in \mathrm{End}(V) \otimes A$, (V, π) a finite-dimensional unitary representation space for G [11, §11.3]. If τ is a G-invariant trace on A, it defines a map $K_0^G(A) \to Z(G) \cong R(G) \otimes \mathbb{C}$ via $[e] \mapsto (g \mapsto \tau^V\big((\pi(g) \otimes 1_A)e\big))$, where τ^V is the trace induced by τ on $\mathrm{End}(V) \otimes A$. Another way of thinking about this is to observe that $R(G) \otimes \mathbb{C}$ is a semisimple ring (a direct sum of copies of $\mathbb{C}$, one for each conjugacy class of G) and so $K_0^G(A) \otimes \mathbb{C}$, as a module over $R(G) \otimes \mathbb{C}$, is a direct sum of $\mathbb{C}$-vector spaces indexed by the conjugacy classes of G. The τ then gives a trace on each summand. For example, if $A = \mathbb{C}$ with $\tau : A \to \mathbb{C}$ the identity map, $K_0^G(A) \cong R(G)$, and for a finite-dimensional representation (V, π) of G, the class $[V] \in R(G)$ is represented by $e = 1_V \in V$, and $(\pi(g) \otimes 1_A)e = \pi(g)$, whose trace under τ^V is just the character of V. As another example, if $X = G/H$ is a transitive G-space, then $K_G^0(X) = R(H)$ (with $R(G)$-module structure via the restriction map $R(G) \to R(H)$). If τ is counting measure on G/H (of total mass $[G : H]$) and if (V, π) is a finite-dimensional representation of H, consider the induced G-vector bundle $E = G \times_H V$, which is corresponds to a finitely generated projective G-$C(X)$-module, with a class in $K_G^0(X)$. When G is abelian, the trace τ^V sends this to the function in $Z(G)$ given by the character of π on H extended to be 0 off of H. When G is not abelian, one gets instead the character of the induced representation $\mathrm{Ind}_{H\uparrow G}(\pi)$, which is supported on the union of the conjugates of H. For example, if $G = S_3$, $H = A_3$ (a cyclic group of order 3), and π is a nontrivial character of H, then $\mathrm{Ind}_{H\uparrow G}(\pi)$ is an irreducible representation with character 2 at the identity, -1 on the 3-cycles, 0 on the 2-cycles.

Now recall some ideas from [3; 29]. First of all, by [29, Proposition 3.7], every prime ideal $\mathfrak{p}$ of $R(G)$ has a *support*, which is a cyclic subgroup H of G determined up to conjugation in G. This is the smallest subgroup $H < G$ such that $\mathfrak{p}$ is the inverse image of a prime ideal in $R(H)$ under the restriction map $R(G) \to R(H)$. Segal's localization theorem [3, Theorem 1.1; 28, Proposition 4.1] says that for a compact G-space such as Ω, the localization of $K_G^0(\Omega)$ at $\mathfrak{p}$ has the property that $K_G^0(\Omega)_\mathfrak{p} \to K_G^0(\Omega^{(H)})_\mathfrak{p}$ is an isomorphism. Here $\Omega^{(H)} = G \cdot \Omega^H$ denotes the G-saturation of the subset fixed by H. (When G is abelian, this is equal to Ω^H.) In particular, if $\mathfrak{p}$ is the prime ideal of characters that vanish at a point $g \in G$ (if we choose $g = 1$, then this $\mathfrak{p}$ is just the augmentation ideal of $R(G)$), then the support of $\mathfrak{p}$ is the cyclic

subgroup generated by g, and if we invert characters that don't vanish at g, which is harmless if we are evaluating characters at g, then we can restrict to $G \cdot \Omega^g$. The trace τ then gives a linear functional defined on $K_G^0(G \cdot \Omega^g)_{\mathfrak{p}} \cong K_G^0(\Omega)_{\mathfrak{p}}$, and by varying g, we get a different functional for each conjugacy class of cyclic subgroups $\langle g \rangle$. There are similar functionals defined on $K_0^G(\mathcal{A}_p(\Omega))$, after localizing at these prime ideals (one for each element of the group). These amount to the same as the map to $Z(G)$ defined before, followed by evaluation at a group element (or conjugacy class, in the noncommutative case).

Conjecture 2.3 (equivariant gap-labeling conjecture) *Suppose $\mathcal{T}$ is a primitive substitution tiling as before, and further assume that there is a finite group G of symmetries that acts on the associated tiling space Ω. This group will of course preserve the unique tracial state $\tau : K_0(\mathcal{A}_p(\Omega)) \to \mathbb{R}$. But G also acts on $\mathcal{A}_p(\Omega)$. Then the image of $K_G^0(\Omega)$ in $Z(G)$ under τ coincides with the image of $K_0^G(\mathcal{A}_p(\Omega))$.*

Let's examine a special case to see what this actually says.

Example 2.4 Let $G = \{1, -1\}$ be cyclic of order 2. Then $R(G) = \mathbb{Z}[\chi]/(\chi^2 - 1)$, where χ is the sign character. After inverting 2, this splits as $\mathbb{Z}\left[\frac{1}{2}\right] \times \mathbb{Z}\left[\frac{1}{2}\right]$, with the two factors corresponding to the trivial representation 1 and to χ. (For a summary of the structure of $R(G)$, including the complete prime ideal structure, see [25, §2].) If A is a unital G-C^*-algebra with a G-invariant trace τ, the map $K_0^G(A) \to Z(G)$ can be viewed as the pair consisting of the usual trace, $\tau : K_0(A) \to \mathbb{R}$, along with the restriction of τ to the fixed-point algebra, giving a map $K_0(A^G) \to \mathbb{R}$. (This is because $R(G)$ has exactly two minimal prime ideals, the augmentation ideal, supported at $\{1\}$, and the kernel of χ, supported on all of G. The remaining prime ideals are all maximal ideals with finite residue field.) So the equivariant gap-labeling conjecture amounts to the usual gap-labeling conjecture *plus* the assertion that the image of the trace on the G-invariant subalgebra $\mathcal{A}_p(\Omega)^G$ coincides with the image of the trace on $K_0(C(\Omega)^G) = K^0(\Omega/G)$.

Remark 2.5 To the best of our knowledge, the equivariant gap-labeling conjecture hasn't been formulated in this way before. But versions of it appear in [23; 24; 32; 33]. Those sources don't deal with equivariant K-theory in full generality but deal either with the K-theory of the crossed product $\mathcal{A}_p(\Omega) \rtimes G$ (which is the same as the equivariant K-theory by the Green–Julg theorem — see [11, §11.7]) or else with the K-theory of the fixed-point algebra $\mathcal{A}_p(\Omega)^G$ (which corresponds to the part of the equivariant K-theory attached to the trivial representation).

2.6 The equivariant Chern character

We can study the equivariant gap-labeling conjecture using the equivariant Chern character, just as the usual Chern character is used to study the usual gap-labeling. The equivariant Chern character (say for the action of a finite group G on a G-CW complex like the AP complex) is defined in [20], and gives a ring homomorphism (that becomes an isomorphism after tensoring the domain with $\mathbb{Q}$) $\mathrm{ch}_G : K_G^*(X) \to H_G^*(X; R(-) \otimes \mathbb{Q})$. Here the right-hand side is *Bredon cohomology* [12] of the functor $R(-) \otimes \mathbb{Q} : \mathrm{Or}(G)^{\mathrm{op}} \to \mathbb{Q}$-Mod, where $\mathrm{Or}(G)$ is the orbit category with objects the transitive G-spaces and morphisms the G-equivariant maps, defined by sending G/H to $R(H) \otimes \mathbb{Q}$. In the special case where X has a single orbit type, say $X = (G/H) \times Y$ with G acting trivially on Y, we have

$K_G^*(X) = R(H) \otimes K^*(Y)$, and the equivariant Chern character is just the usual Chern character for Y, tensored with $R(H)$.

Now to prove the equivariant gap-labeling conjecture for a particular tiling space Ω, one can try to replicate the strategy based on the diagram (4). The difference would be that we use equivariant K-theory and cohomology, and the map τ to $\mathbb{R}$ is now replaced by the map τ^g induced by τ on equivariant K-theory/cohomology, along with evaluation of (virtual) characters at an element $g \in G$. For simplicity let's take G to be abelian. Fix $g \in G$ and apply the Segal localization theorem with $\mathfrak{p} = \mathfrak{p}_g$ the prime ideal of $R(G)$ consisting of virtual representations whose characters vanish at g. The support of $\mathfrak{p}$ is the cyclic subgroup $H = \langle g \rangle$, so after localizing at $\mathfrak{p}$ (in other words, inverting characters that do not vanish at g), we can replace Ω by Ω^g. So diagram (4) becomes

$$\begin{array}{ccc} K_0^G(\mathcal{A}_p(\Omega))_{\mathfrak{p}} & \xrightarrow{\mathrm{ch}_d^G \circ \chi^{-1}} & \check{H}_G^d(\Omega^g; R(-)_{\mathfrak{p}} \otimes \mathbb{Q}) \\ \downarrow{\scriptstyle \tau^g} & & \downarrow{\scriptstyle C_\mu^g} \\ \mathbb{C} & \xrightarrow{\qquad\qquad \mathrm{Id} \qquad\qquad} & \mathbb{C} \end{array} \tag{6}$$

This reduces to (4) if $g = 1$, but in general, if $g \neq 1$, we need to use $\mathbb{C}$ as the target instead of $\mathbb{R}$ since G can have characters with nonreal values at g. (This already happens for G cyclic of order > 2.) The restriction of τ^g to $K_0^G(C(\Omega)) = K_G^0(\Omega)$ via the inclusion of $C(\Omega)$ into $\mathcal{A}_p(\Omega)$ can be computed via a G-invariant transversal N to the $\mathbb{R}$-orbits in Ω. Since N is totally disconnected (i.e., 0-dimensional), there are no denominators in the equivariant Chern character for N, which becomes an *integral* isomorphism. So just as in [1, Remark 9.7], we have $K_0^G(\mathcal{A}_p(\Omega)) \cong K_0^G(C(N) \rtimes \mathbb{Z}^d)$ and we want to relate this to $K_0^G(C(N)) = K_G^0(N) \cong \check{H}_G^0(N; R(-))$. The restriction of τ^g is $\mathbb{Z}^d$-invariant, so it factors through $H_G^0(N; R(-))_{\mathbb{Z}^d} \cong H_G^d(\Omega; R(-))$. So the equivariant gap-labeling conjecture reduces to the assertion that for each $g \in G$, the image of τ^g on $K_0^G(\mathcal{A}_p(\Omega))$ in (6) coincides (integrally!) with the image of $H_G^d(\Omega^g; R(-))$. Just as in the nonequivariant case, the conjecture follows from diagram (6) in dimensions up to 3 when there are no denominators in the Chern character.

Theorem 2.6 *The equivariant gap-labeling conjecture, Conjecture 2.3, holds when the dimension d is ≤ 3. However, it* ***fails*** *for a periodic lattice tiling with $\Omega = \mathbb{T}^4$ (the 4-torus) and G the cyclic group of order 2 interchanging the two factors in $\mathbb{T}^4 = \mathbb{T}^2 \times \mathbb{T}^2$.*

Proof We have already explained the proof when $d \leq 3$. It remains to show why the conjecture fails for $\Omega = \mathbb{T}^2 \times \mathbb{T}^2$ and G the cyclic group of order 2 interchanging the two factors. By Example 2.4, it suffices to show that the image of the trace on $H^4(\Omega/G; \mathbb{Z}) = H^4(\mathrm{SP}^2(\mathbb{T}^2); \mathbb{Z})$, $\mathrm{SP}^2(\mathbb{T}^2)$ the second symmetric product of $\mathbb{T}^2$, is strictly smaller than the image of the trace on the fixed-point algebra $\mathcal{A}_p(\Omega)^G$. Via the analogue of (4), the issue is to show that the top-degree Chern character on $K^0(\mathrm{SP}^2(\mathbb{T}^2))$ has image which is strictly bigger than $H^4(\mathrm{SP}^2(\mathbb{T}^2); \mathbb{Z})$. For this we can apply [21, (6.30) and (7.1)], which computes the structure of the integral cohomology ring of $\mathrm{SP}^2(\mathbb{T}^2)$. (In the language of [21], this is the case where X is a curve of genus $g = 1$ and $n = 2$.) Macdonald shows that the cohomology ring

is torsion-free and is generated over $\mathbb{Z}$ by generators ξ_1, ξ_2 of degree 1 (which anticommute with each other and with themselves) and a generator η of degree 2, commuting with ξ_1 and ξ_2, with the one relation $\eta^2 = \eta\xi_1\xi_2 =$ generator of $H^4(\mathrm{SP}^2(\mathbb{T}^2);\mathbb{Z})$. If L is the complex line bundle with $c_1(L) = \eta$, then $\mathrm{ch}([L]) = 1 + \eta + \frac{1}{2}\eta^2$, which is *not* integral, and its projection into $H^4(\mathrm{SP}^2(\mathbb{T}^2);\mathbb{Q})$ generates an infinite cyclic group which properly contains $H^4(\mathrm{SP}^2(\mathbb{T}^2);\mathbb{Z})$ (with index 2). □

An example of an aperiodic tiling space with a $\mathbb{Z}_2$-action which is also a counterexample is described in Section 4.3. We then have the following corollary as an immediate consequence of this counterexample.

Corollary 2.7 *The gap-labeling conjecture* ***fails*** *for rotational tiling spaces when the dimension d is ≥ 4.*

While patch frequencies are well defined for rotational tiling spaces, thus one can compute the associated frequency module, it is unclear if there is a Schrödinger operator whose gaps in the spectra are labeled by the K_0-group of the associated C^*-algebra.

Remark 2.8 We should mention that there seem to be different ways to formulate the equivariant conjecture, and at first sight they don't seem to be the same. Let us illustrate this with the case of Example 2.4, a $\mathbb{Z}_2$-action. If X is, say, a Cantor set with a G-action, where $G = \mathbb{Z}_2$, equipped with a G-invariant measure μ, then $A = C(X)$ is generated by projections χ_Y (corresponding to clopen subsets $Y \subseteq X$), whose classes also generate $K^0(X)$, and the trace τ defined by μ sends χ_Y to $\mu(Y)$. In Example 2.4, we said that the equivariant conjecture amounts to looking at τ on A along with the induced trace on $A^G = C(X/G)$. If $X^G = Z$, then X is the union of Z and the free G-space $X \smallsetminus Z$, so the (unrenormalized) trace on $C(X/G)$ has total mass $\mu(Z) + \frac{1}{2}\mu(X \smallsetminus Z)$. The formulation using (6) amounts to looking instead at the pair consisting of μ on X and μ restricted to $X^G = Z$. Knowing these, one can compute $\mu(X \smallsetminus Z)$, and thus get the same set of invariants. However, because of the factor of 2 that appears in the calculation, one has to be careful in looking at the *integral* range of the trace. The reduction to fixed-point sets in (6) appeals to the Segal localization theorem, which requires inverting elements of $R(G)$ not in the augmentation ideal I. This involves inverting 2, so there is no contradiction in relating the two versions of the equivariant conjecture, the one using the quotient space and the one using the fixed-point set.

3 Cubical substitutions and the cup product

In this section we describe our approach to compute the cohomology rings of tiling spaces given by cubical substitutions. With an eye of doing computations in dimension four in Section 4, we work out examples in dimension two where we explicitly compute the cohomology ring structure and other invariants. We will first start describing in detail how the type of collaring scheme used to compute the cohomology groups of cubical substitutions and the ring structure is computed by following [15]. Table 1 summarizes the properties of the different examples.[1]

[1]The table lists groups as $\mathbb{Z}[1/16]$, $\mathbb{Z}[1/8]$, $\mathbb{Z}[1/4]$ and $\mathbb{Z}[1/2]$, even though these can all be written as $\mathbb{Z}[1/2]$. Following [27, §3.5], we write them like this to emphasize their different scalings under the substitution rule.

example	cohomology groups	cup product	frequency module
Example 3.2	$\check{H}^2(\Omega) = \mathbb{Z}[1/25] \oplus \mathbb{Z}[1/15] \oplus \mathbb{Z}[1/5] \oplus \mathbb{Z}[1/3]$, $\check{H}^1(\Omega) = \mathbb{Z}[1/5]^2 \oplus \mathbb{Z}[1/3] \oplus \mathbb{Z}$, $\check{H}^0(\Omega) = \mathbb{Z}$	surjective	$\frac{1}{6}\mathbb{Z}[1/5]$
Example 3.3	$\check{H}^2(\Omega) = \mathbb{Z}[1/16] \oplus \mathbb{Z}[1/8] \oplus \mathbb{Z}[1/4] \oplus \mathbb{Z}[1/2]$, $\check{H}^1(\Omega) = \mathbb{Z}[1/4]^2 \oplus \mathbb{Z}[1/2] \oplus \mathbb{Z}$, $\check{H}^0(\Omega) = \mathbb{Z}$	surjective	$\frac{1}{3}\mathbb{Z}[1/2]$
Example 3.4	see Example 3.4 for details	nonsurjective	$\frac{1}{84}\mathbb{Z}[1/5]$
Example 3.5	$\check{H}^2(\Omega) = \mathbb{Z}[1/16] \oplus \mathbb{Z}[1/8] \oplus \mathbb{Z}[1/4] \oplus \mathbb{Z}[1/2]$, $\check{H}^1(\Omega) = \mathbb{Z}[1/4]^2 \oplus \mathbb{Z}[1/2] \oplus \mathbb{Z}$, $\check{H}^0(\Omega) = \mathbb{Z}$	nonsurjective	$\frac{1}{3}\mathbb{Z}[1/2]$

Table 1: Summary of Examples 3.2–3.5.

The four examples in this section are meant as two pairs of examples to be compared. Of particular note are Examples 3.3 and 3.5. These are two tiling spaces with isomorphic cohomology groups over $\mathbb{Z}$ and the same frequency module. Without the cup product, they would be indistinguishable. However, the cup product $\check{H}^1 \times \check{H}^1 \to \check{H}^2$ in one case is surjective, whereas it is not in the other. This is the first example we are aware of where the cup product is the only distinguishing feature between two tiling spaces. Examples 3.2 and 3.4 have the feature that they have indistinguishable cohomology spaces over $\mathbb{Q}$, but can be distinguished either through their cohomology groups over $\mathbb{Z}$ or through the ring structure.

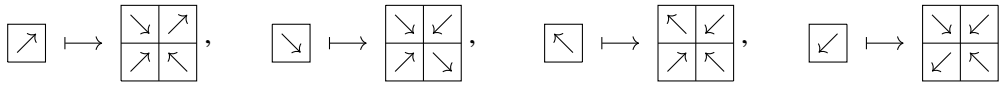
3.1 Computational setup

Let us first establish the notation and conventions.

Recall that a cubical substitution rule ς on m prototiles with (uniform) expansion matrix $\lambda = \lambda \cdot \mathrm{Id}_d$ has the set of prototiles $[0,1]^d \times \{0,\dots,m-1\}$. (I.e., the set of colors is $\mathrm{Col} = \{0,\dots,m-1\}$.) We will name each prototile by its second coordinate, its *label*. In addition, recall that each tile is a labeled region $t + [0,1]^d \subseteq \mathbb{R}^d$. We interpret this as being formed by anchoring the point $[0]^d \in [0,1]^d$, its *puncture*, in the corresponding prototile to $t \in \mathbb{R}^d$. We follow the convention of [15] where $[x]$ denotes a (degenerate) interval consisting of the single point x.

We decompose the substitution rule ς into the following two steps: the first, denoted by λ, inflates each tile by λ; the second, denoted by σ, subdivides the resulting region into tiles. If the initial tile is of the form (t,k) with $t \in \mathbb{Z}^d$, the resulting substituted patch will have each of the punctures of each of its tiles attached to $\mathbb{Z}^d$. Thus, let us put the lexicographic order on $\{0,\dots,\lambda-1\}^d$, and using this order we abbreviate the substitution rule as $\varsigma(k) = (k_0,\dots,k_{\lambda^d-1})$, where $0 \le k_i \le m-1$ is a prototile with its puncture attached at the coordinate that is the i-th entry in $(0,\dots,\lambda-1)^d$. We illustrate this convention through a well-known example.

Example 3.1 (chair tiling) The standard chair tiling can be decomposed into squares, yielding the MLD-equivalent cubical substitution rule in $d = 2$, as illustrated in



Labeling the prototiles $\{0, 1, 2, 3\} = \{\boxed{\nearrow}, \boxed{\searrow}, \boxed{\nwarrow}, \boxed{\swarrow}\}$, we abbreviate this substitution rule as

$$\varsigma(0) = (0, 1, 2, 0), \quad \varsigma(1) = (0, 1, 1, 3), \quad \varsigma(2) = (0, 2, 2, 3), \quad \varsigma(3) = (3, 1, 2, 3),$$

or further abbreviate it as a list in Sage as

```
[[0,1,2,0],[0,1,1,3],[0,2,2,3],[3,1,2,3]]
```

For arbitrary dimension $d > 1$, the d-dimensional chair substitution rule is, in Sage,

```
[[(2**d-1)-j if i==(2**d-1)-j else j for j in list(range(2**d))]
                                                    for i in list(range(2**d))]
```

One immediately encounters an obstacle that is very difficult to overcome when trying to construct the AP-inverse limit and compute its cohomology. The top-dimensional cells in the complex are collared prototiles, or, for cubical substitutions, patches of size 3^d, and the codimension-1 cells are obtained from adjacency information of pairs of collared tiles. The latter requires us to find all possible patches of size 4^d in $\mathcal{T}$. In higher dimensions, e.g., $d = 4$, this becomes unwieldy, since they are proportional to m^{4d}. In addition, it may require multiple substitutions before the supertiles of that level observe all such patches of size 4^d. In higher dimensions, these severely restrict the substitution rules we can check. Thus, rather than the standard collaring procedure yielding the AP-complex and inverse limit, let us consider an alternative, the dual complex, introduced in [13], that is a variant of the Barge–Diamond (BD) complex from [5] for $d = 1$ and [6] for $d > 1$.

For motivation, let us first consider an alternate point of view of how the AP-complex is formed. Suppose that we are given the set of all collared prototiles, viewed as patches in $\mathcal{T}$. They form the set of top-dimensional cells. The set of codimension-1 cells consists of unions of pairs of neighboring patches, viewed as intersections of their associated cylinder sets. Similarly, the set of codimension-k cells consists of unions of patches that share a common intersection. That is, the AP-complex is the Čech complex associated to collared prototiles.

For cubical complexes, the top-dimensional cells are patches (up to translation) of size 3^d, and the codimension-k cells are unions of 2^k patches that result in patches of $\mathcal{T}$ and share exactly $2^k 3^{d-k}$ tiles (so that the center tiles neighbor along a codimension-k cell; analogously, so that the differences are patches of size 1 along k directions, 3^{d-k} along the remaining $d-k$ directions). Repeating this process for each level of supertiles yields the AP-inverse limit. From the Čech perspective, the fact that this inverse limit is homeomorphic to $\Omega_{\mathcal{T}}$ is because the basic open sets of each (patches of collared supertiles for the AP point of view compared to arbitrary patches) are mutual refinements.

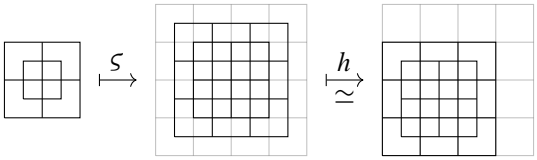


Figure 1: A substitution rule with even expansion applied to a patch of size 2^d representing a top-dimensional cell (center tile shaded) returns a patch (light gray lines) whose center tiles (shaded) has collars (black lines) that are not actual patches, therefore is not cellular. One needs to compose the substitution with a homotopy h to produce a cellular map.

Rather than working with patches of size 3^d, one can use patches of size 2^d, then carry through the Čech construction. The resulting complex, due to [13], is the *dual complex*. More precisely, at the initial level of the inverse limit, the codimension-k cells are unions of 2^k patches that share exactly $1^k 2^{d-k}$ tiles. To see that this is collaring of sorts, if a d-cell has the puncture of its first tile (in lexicographic order) situated at the origin, one pretends that the "center" tile (the actual d-cell) is the unit cube with its puncture at $[1/2]^d$.

To construct the complex at the supertile level, we first note that if λ is even, the substitution rule is not a cellular map (Figure 1). Assuming that each patch of size 2^d has the puncture of its first tile located at the origin, we define a supertile to be the collection of all subpatches of size 2^d (they will overlap) formed from tiles of the actual supertile with punctures located at $(0,\dots,\lambda)^d$. The substitution rule then takes a supertile of the dual complex and sends it to all subpatches of size 2^d. Higher-level supertiles are analogous.

We observe that if we focus on the "center" tiles (and treat the surrounding patches of size 2^d as decorations), this map is the ordinary substitution rule followed by a fixed homotopy, therefore the maps of the resulting inverse limit are stationary up to homotopy. Furthermore, the supertiles of this construction and the supertiles of the Anderson–Putnam construction are cofinal refinements of each other. This inverse limit is thus shape equivalent to $\Omega_{\mathcal{T}}$, and has the same Čech cohomology groups as those of $\Omega_{\mathcal{T}}$ [22, Theorem 16].[2]

For all of the computations that are performed directly using Sage, we will use the dual complex and the resulting inverse limit.

3.2 Distinguishing tiling spaces through cup product

As an application of the ring structure of cohomology on cubical substitutions, let us describe pairs of tiling spaces in $d = 2$ whose cohomology groups coincide, but the ring structure differs. We consider $d = 2$, since the required computations are much easier to visualize and interpret. These examples are inspired by [31, Example 6.7] due to the induced substitution matrix on $\check{H}^1$ containing two eigenvalues that do not come from the expansion.

We will consider two classes of examples. In all of the complexes that we construct, we take as positive left-to-right and down-to-up orientations of 1-cells, and right-handed orientation of 2-cells.

[2]We are indebted to the referee for this observation.

Example 3.2 (expansion 5 product) Let us consider the following two one-dimensional substitutions on two and three prototiles, respectively:

ς_1=`[[0,0,0,0,1],[1,0,0,0,0]]`

ς_2=`[[0,2,1,2,0],[0,1,1,1,0],[0,2,2,2,0]]`

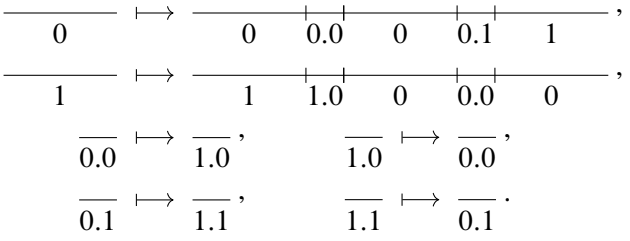
One easily checks that both substitutions are primitive and recognizable.

To compute the first cohomology group, we use the Barge–Diamond filtration from [5], where one successively blows up lower-dimensional skeleta to become d-dimensional (here $d = 1$) complexes. We will do this in detail for ς_1, then just state the main parts of the computation for ς_2.

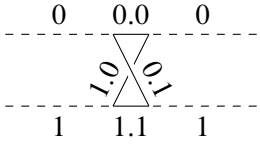
Let $\epsilon > 0$. We replace the prototiles themselves, assumed to be of length 1, by tiles of length $1-2\epsilon$, denoted using the original symbols, and replace the vertices that join two neighboring tiles by new tiles of length 2ϵ (called *vertex flaps* in [6]; more generally, k-flaps for new tiles of size $1-2\epsilon$ in k directions, 2ϵ in the remaining $d-k$ directions, which are blow ups of k-cells in the tiling). These new prototiles of length 2ϵ arise from the set of patches of length 2, denoted by 0.0, 0.1, 1.0, and 1.1 for this example. This new tiling is no longer self-similar, thus let us construct a homotopy that maintains the same self-similar structure as the original tiling.

To maintain the same combinatorial structure as the original substitution rule, we will require that the 0-flaps be nonexpanding, and the 1-flaps (the slightly shrunken prototiles) be fully expanding (more generally, for cubical substitutions, k-flaps are expanding in k directions, nonexpanding in the remaining $d-k$ directions). More precisely, if λ is the expansion of the original substitution rule, in dimension 1, the 0-flaps are substituted, via expansion by λ then a homotopy, to prototiles of size 2ϵ (again 0-flaps), and the 1-flaps are sent, via expansion by λ then a homotopy, to patches of size $\lambda - 2\epsilon$ (a combination of 0- and 1-flaps).

For ς_1, we have that the induced substitution rule, still denoted by ς_1, is



The Barge–Diamond filtration is a filtration of complexes $S_0 \subseteq S_1 \subseteq \cdots \subseteq S_d$. The complex S_0 is formed by taking the 0-flaps and performing identifications similar to the Anderson–Putnam construction, S_1 is formed by taking S_0, then adding in the 1-flaps (then possibly performing identifications if there are 2-flaps), etc. For ς_1, S_0, drawn in solid lines, is a loop



where the dashed lines indicate the 1-flaps that yield, for example, identification of the left endpoints of 0.0 and 0.1. The complex S_1 is then the same picture, together with the dashed lines turned solid. We define $\Xi_i = \varprojlim(S_i, \varsigma_1)$. Then $\Xi_1 = \Omega_{\varsigma_1}$.

One can check that $\varsigma_1 : S_0 \to S_0$ is an orientation-reversing self-homeomorphism, thus $\check{H}^1(\Xi_0) = \mathbb{Z}$. For Ξ_1, we use the long exact sequence

$$0 \longrightarrow \check{H}^0(\Xi_1, \Xi_0) \longrightarrow \check{H}^0(\Xi_1) \longrightarrow \check{H}^0(\Xi_0) \xrightarrow{\delta} \check{H}^1(\Xi_1, \Xi_0) \longrightarrow \check{H}^1(\Xi_1) \longrightarrow \check{H}^1(\Xi_0) \longrightarrow 0$$

in relative cohomology of a pair (Ξ_1, Ξ_0). Since S_0 has a single component, over reduced cohomology, $\check{H}^1(\Xi_1, \Xi_0) \to \check{H}^1(\Xi_1) \to \check{H}^1(\Xi_0) \to 0$ becomes a short exact sequence.

The space S_1/S_0 is a wedge of two circles, with its inverse limit computed exactly from the original substitution matrix. Thus

$$\check{H}^1(\Xi_1, \Xi_0) = \varinjlim\left(\mathbb{Z}^2, \begin{pmatrix} 4 & 1 \\ 4 & 1 \end{pmatrix}\right) = \mathbb{Z}[1/5].$$

Since $\check{H}^1(\Xi_0) = \mathbb{Z}$ is free, the bottom short exact sequence splits, yielding

$$\check{H}^1(\Omega_{\varsigma_1}) = \mathbb{Z}[1/5] \oplus \mathbb{Z}, \quad \check{H}^0(\Omega_{\varsigma_1}) = \mathbb{Z}.$$

The calculation for ς_2 is easier. It is a proper substitution, thus forces the border (see Remark 2.1), and

$$\check{H}^1(\Omega_{\varsigma_2}) = \varinjlim\left(\mathbb{Z}^3, \begin{pmatrix} 2 & 1 & 2 \\ 2 & 3 & 0 \\ 2 & 0 & 3 \end{pmatrix}\right).$$

The matrix has eigenvectors $\boldsymbol{e}_5 = (1, 1, 1)$ and $\boldsymbol{e}_3 = (0, 2, -1)$ with the subscripts their respective eigenvalues. Direct computation then shows that the direct limit splits as a direct sum, giving us

$$\check{H}^1(\Omega_{\varsigma_2}) = \mathbb{Z}[1/5] \oplus \mathbb{Z}[1/3] \quad \text{and} \quad \check{H}^0(\Omega_{\varsigma_2}) = \mathbb{Z}.$$

Let us also compute the induced substitution matrices on $\check{H}^1(\Omega_{\varsigma_1})$ using the dual complex. It is computed via direct limit of the matrix

$$\sigma_1^1 = \begin{pmatrix} 1 & 1 & 0 \\ 3 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

with eigenvalues of $3, -1$. The eigenvalue of -1 matches the orientation reversal of ς_1 on its respective S_0.

We finally consider the product tiling space $\Omega_{\varsigma_1} \times \Omega_{\varsigma_2}$. A direct application of the Künneth theorem yields that

$$\begin{aligned} \check{H}^2(\Omega_{\varsigma_1} \times \Omega_{\varsigma_2}) &= \mathbb{Z}[1/25] \oplus \mathbb{Z}[1/15] \oplus \mathbb{Z}[1/5] \oplus \mathbb{Z}[1/3], \\ \check{H}^1(\Omega_{\varsigma_1} \times \Omega_{\varsigma_2}) &= \mathbb{Z}[1/5]^2 \oplus \mathbb{Z}[1/3] \oplus \mathbb{Z}, \\ \check{H}^0(\Omega_{\varsigma_1} \times \Omega_{\varsigma_2}) &= \mathbb{Z} \end{aligned}$$

with $\check{H}^2(\Omega_{\varsigma_1}\times\Omega_{\varsigma_2})$ having eigenvalues 25, 15, -5, -3 and $\check{H}^1(\Omega_{\varsigma_1}\times\Omega_{\varsigma_2})$ having eigenvalues 5_2, 5_1, 3_2, -1_1, with the subscripts indicating the 1-dimensional substitution rule that gives rise to the corresponding eigenvalue.

Lifting each of the cohomology classes in $\check{H}^1(\Omega_{\varsigma_1}\times\Omega_{\varsigma_2})$ to 1-cochains, applying the cubical cup-product formula given in [15], then projecting to cohomology classes in $\check{H}^2(\Omega_{\varsigma_1}\times\Omega_{\varsigma_2})$ gives that, in the order given in the list of eigenvalues, the cup product is the resulting bilinear form

$$\check{H}^1\times\check{H}^1\to\check{H}^2:(a,b)\mapsto B(a,b):=\sum_{*\in\{25,15,-5,-3\}} B_*(a,b)\cdot e_*,$$

where $e_n\in\check{H}^2$ is a generator with eigenvalue n, and B_n is the bilinear form given by the matrices

$$B_{25}:=\begin{pmatrix}0&-1&0&0\\1&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad B_{15}:=\begin{pmatrix}0&0&0&0\\0&0&1&0\\0&-1&0&0\\0&0&0&0\end{pmatrix},\quad B_{-5}:=\begin{pmatrix}0&0&0&-1\\0&0&0&0\\0&0&0&0\\1&0&0&0\end{pmatrix},\quad B_{-3}:=\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&-1\\0&0&1&0\end{pmatrix}.$$

The important observation is that, as expected, there are always cohomology classes in $\check{H}^1(\Omega_{\varsigma_1}\times\Omega_{\varsigma_2};\mathbb{Z})$ cupping to any of the cohomology classes in $\check{H}^2(\Omega_{\varsigma_1}\times\Omega_{\varsigma_2};\mathbb{Z})$.

Lastly, let us compute the frequency module. The substitution matrix on the 2-cells of the dual complex is the matrix $\sigma^2=\sigma_{A1}$ found in the Appendix, which has Perron–Frobenius eigenvector

$$(95,19,25,5,25,5,5,1,76,20,20,4,38,10,10,2,38,10,10,2,\\ 76,20,20,4,19,5,5,1,38,76,10,20,10,20,2,4).$$

Its sum is 750, giving us that the frequency module is $\frac{1}{6}\mathbb{Z}[1/5]$. For illustration, we also compute the frequency modules for each of the one-dimensional substitutions. The substitution matrices on the 1-cells of the corresponding dual complexes are

$$\sigma_1^1=\begin{pmatrix}3&3&4&3\\1&1&0&1\\1&0&1&1\\0&1&0&0\end{pmatrix}\quad\text{and}\quad\sigma_2^1=\begin{pmatrix}1&1&1&1&1&1&1&1&1\\0&0&0&1&0&0&1&1&0\\1&1&1&0&1&1&0&0&1\\0&0&0&2&0&0&2&2&0\\1&1&1&0&0&0&0&0&0\\0&0&0&0&2&2&0&0&2\\0&0&0&1&0&0&1&1&0\\1&1&1&0&0&0&0&0&0\\1&1&1&0&1&1&0&0&1\end{pmatrix}$$

with Perron–Frobenius eigenvectors $(19,5,5,1)$ and $(5,1,4,2,2,4,1,2,4)$ that are duals of lifts of the generators of $\check{H}^1(\Omega_{\varsigma_i})$ with sums 30 and 25, respectively, giving frequency modules $\frac{1}{6}\mathbb{Z}[1/5]$ and $\mathbb{Z}[1/5]$, respectively. Since the Ruelle–Sullivan map is a ring homomorphism, $\frac{1}{6}\mathbb{Z}[1/5]\cdot\mathbb{Z}[1/5]=\frac{1}{6}\mathbb{Z}[1/5]$.

Example 3.3 (expansion 4 product) Let us consider the following two one-dimensional substitutions on two prototiles:

```
ς1=[[0,0,0,1],[0,1,1,1]]
ς2=[[0,1,1,0],[1,0,0,1]]
```

One, again, easily checks that both substitutions are primitive and recognizable. In particular, ς_2 is the (square of the) Thue–Morse substitution, whose cohomology groups were computed in [5], being

$$\check{H}^1(\Omega_{\varsigma_2}) = \mathbb{Z}[1/4] \oplus \mathbb{Z} \quad \text{and} \quad \check{H}^0(\Omega_{\varsigma_2}) = \mathbb{Z}.$$

The substitution ς_1 is proper, thus forces the border (see Remark 2.1), and

$$\check{H}^1(\Omega_{\varsigma_1}) = \varinjlim\left(\mathbb{Z}^2, \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}\right).$$

The matrix has eigenvectors $\boldsymbol{e}_4 = \binom{1}{1}$ and $\boldsymbol{e}_2 = \binom{1}{-1}$ with the subscripts their respective eigenvalues. Direct computation then shows that the direct limit splits as a direct sum, giving us

$$\check{H}^1(\Omega_{\varsigma_1}) = \mathbb{Z}[1/4] \oplus \mathbb{Z}[1/2] \quad \text{and} \quad \check{H}^0(\Omega_{\varsigma_1}) = \mathbb{Z}.$$

The Künneth theorem yields that, for the product $\Omega_{\varsigma_1} \times \Omega_{\varsigma_2}$,

$$\begin{aligned}
\check{H}^2(\Omega_{\varsigma_1} \times \Omega_{\varsigma_2}) &= \mathbb{Z}[1/16] \oplus \mathbb{Z}[1/8] \oplus \mathbb{Z}[1/4] \oplus \mathbb{Z}[1/2], \\
\check{H}^1(\Omega_{\varsigma_1} \times \Omega_{\varsigma_2}) &= \mathbb{Z}[1/4]^2 \oplus \mathbb{Z}[1/2] \oplus \mathbb{Z}, \\
\check{H}^0(\Omega_{\varsigma_1} \times \Omega_{\varsigma_2}) &= \mathbb{Z}
\end{aligned}$$

with $\check{H}^2(\Omega_{\varsigma_1} \times \Omega_{\varsigma_2})$ having eigenvalues 16, 8, 4, 2 and $\check{H}^1(\Omega_{\varsigma_1} \times \Omega_{\varsigma_2})$ having eigenvalues 4_2, 4_1, 2_1, 1_2, with the subscripts indicating the 1-dimensional substitution rule that gives rise to the corresponding eigenvalue.

The same procedure as the previous example gives that, in the order given in the list of eigenvalues, the cup product is the resulting bilinear form

$$\check{H}^1 \times \check{H}^1 \to \check{H}^2 : (a, b) \mapsto B(a, b) := \sum_{*\in\{16,8,4,2\}} B_*(a, b) \cdot e_*,$$

where $e_n \in \check{H}^2$ is a generator with eigenvalue n, and B_n is the bilinear form given by the matrices

$$B_{16} := \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B_8 := \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B_4 := \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad B_2 := \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}.$$

Again, we observe that, as expected, there are always cohomology classes in $\check{H}^1(\Omega_{\varsigma_1} \times \Omega_{\varsigma_2})$ cupping to any of the cohomology classes in $\check{H}^2(\Omega_{\varsigma_1} \times \Omega_{\varsigma_2})$.

Lastly, let us compute the frequency module. The substitution matrix on the 2-cells of the dual complex is $\sigma^2 = \sigma_{A2}$ found in the Appendix which has Perron–Frobenius eigenvector

$$(1, 2, 1, 2, 1, 2, 1, 2, 2, 2, 2, 2, 1, 1, 1, 1).$$

Its sum is 24, giving us that the frequency module is $\frac{1}{3}\mathbb{Z}[1/2]$. For illustration, we also compute the frequency modules for each of the one-dimensional substitutions. The substitution matrices on the 1-cells

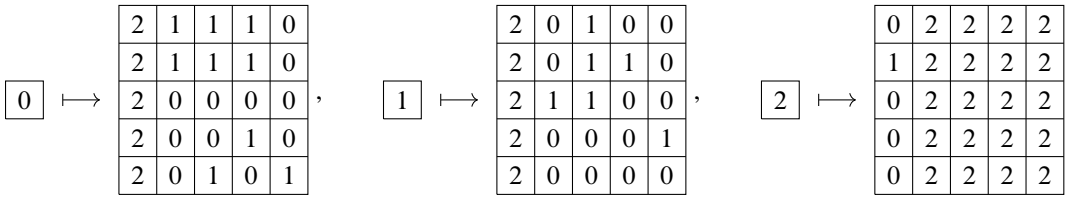


Figure 2: Substitution rule for an expansion 5 two-dimensional cubical substitution that has the cohomology groups of a product, but does not have the ring structure of a product.

of the corresponding dual complexes are

$$\sigma_1^1 = \begin{pmatrix} 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 2 & 2 \end{pmatrix} \quad \text{and} \quad \sigma_2^1 = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

with Perron–Frobenius eigenvectors that are duals of lifts of the generators of $\check{H}^1(\Omega_{\varsigma_i})$ are $(1,1,1,1)$ and $(1,2,2,1)$ with sums 4 and 6, respectively, giving frequency modules $\mathbb{Z}[1/2]$ and $\frac{1}{3}\mathbb{Z}[1/2]$, respectively. As expected, since the Ruelle–Sullivan map is a ring homomorphism, $\mathbb{Z}[1/2]\cdot\frac{1}{3}\mathbb{Z}[1/2] = \frac{1}{3}\mathbb{Z}[1/2]$.

Example 3.4 (expansion 5 nonproduct) We consider the following two-dimensional substitution on three prototiles (see Figure 2):

```
ς=[[2,2,2,2,2,0,0,0,1,1,1,0,0,1,1,0,1,0,1,1,1,0,0,0,0],
   [2,2,2,2,2,0,0,1,0,0,0,0,1,1,1,0,0,0,1,0,0,1,0,0,0],
   [0,0,0,1,0,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2]]
```

This does not force the border (e.g., the horizontal 1-cell containing the prototiles 0, 1 on its bottom and prototiles 0, 1 on its top stays branched regardless of the power of the substitution), but is primitive and recognizable.

Using the dual complex, $\check{H}^2(\Omega_\varsigma)$ and $\check{H}^1(\Omega_\varsigma)$ are computed via direct limits under the matrices

$$\sigma^2 = \begin{pmatrix} 1 & 0 & 3 & 1 & 2 & 1 & 1 & 2 & 0 & 1 & 1 & 0 & 2 & 1 & 2 \\ 1 & 0 & 3 & 1 & 2 & 1 & 1 & 2 & 0 & 1 & 1 & 0 & 2 & 1 & 2 \\ 2 & 1 & 3 & 1 & 3 & 2 & 2 & 2 & 0 & 1 & 1 & 20 & 2 & 1 & 2 \\ 5 & 4 & 6 & 2 & 5 & 5 & 5 & 1 & 2 & 2 & 1 & 40 & 3 & 1 & 2 \\ 1 & 0 & 3 & 1 & 2 & 1 & 1 & 2 & 0 & 1 & 1 & 0 & 2 & 1 & 2 \\ 6 & 3 & 9 & 3 & 5 & 5 & 5 & 1 & 3 & 2 & 2 & 0 & 4 & 1 & 2 \\ -3 & -1 & -3 & -1 & -2 & -2 & -2 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 0 \\ -3 & -1 & -3 & -1 & -2 & -2 & -2 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 3 & 1 & 2 & 1 & 1 & 2 & 0 & 1 & 1 & 0 & 2 & 1 & 2 \\ 3 & 1 & 3 & 1 & 2 & 2 & 2 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 6 & 2 & 6 & 2 & 4 & 4 & 4 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 15 & 0 & 0 & 0 \\ 3 & 2 & 6 & 2 & 3 & 3 & 3 & 1 & 2 & 2 & 1 & 0 & 3 & 1 & 2 \\ 6 & 3 & 9 & 3 & 5 & 5 & 5 & 1 & 3 & 2 & 2 & 0 & 4 & 1 & 2 \\ 3 & 1 & 3 & 1 & 2 & 2 & 2 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad \sigma^1 = \begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 4 & 1 & 4 \\ 0 & 0 & 1 & 3 \end{pmatrix}$$

with eigenvectors

$$\begin{aligned}
\boldsymbol{e}^1_{5_1} &= (1,0,0,0), & \boldsymbol{e}^2_{25} &= (1,1,2,4,1,3,-1,-1,1,1,2,1,2,3,1),\\
\boldsymbol{e}^1_{5_2} &= (0,1,2,1), & \boldsymbol{e}^2_{15} &= (1,1,0,0,1,3,-1,-1,1,1,2,-1,2,3,1),\\
\boldsymbol{e}^1_{3} &= (0,1,0,-1), & \boldsymbol{e}^2_{-5} &= (1,1,-4,-8,1,3,-1,-1,1,1,2,1,2,3,1),\\
\boldsymbol{e}^1_{-1} &= (0,1,-4,1), & \boldsymbol{e}^2_{-3} &= \begin{pmatrix} 101,101,90,-72,101,-201,151,151,\\ 101,-151,-302,-11,-50,-201,-151 \end{pmatrix},
\end{aligned}$$

where the subscripts indicate the corresponding eigenvalues. The cup product is the bilinear form

$$\check{H}^1 \times \check{H}^1 \to \check{H}^2 : (a,b) \mapsto B(a,b) := \sum_{*\in\{25,15,-5,-3\}} B_*(a,b)\cdot \boldsymbol{e}^2_*,$$

where $\boldsymbol{e}^2_n \in \check{H}^2$ is a generator with eigenvalue n, and B_n is the bilinear form given by the matrices

$$B_{25} := \begin{pmatrix} 0 & -1 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B_{15} := \begin{pmatrix} 0 & 0 & -1 & 0\\ 0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B_{-5} := \begin{pmatrix} 0 & 0 & 0 & -1\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad B_{-3} := \begin{pmatrix} 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The last component of the bilinear form shows that there are no cohomology classes in $\check{H}^1(\Omega_\varsigma)$ cupping to $\boldsymbol{e}^2_{-3}$ in $\check{H}^2(\Omega_\varsigma)$, contrasting the example coming from the product. This coincides with the fact that, up to a coboundary, the lifts of the eigenvectors $\boldsymbol{e}^1_1$ and $\boldsymbol{e}^1_{-1}$ to 1-cochains are

$$\widetilde{\boldsymbol{e}^1_3} = (0,0,0,0,0,1,1,0,0,1,0,0,-1,1), \quad \widetilde{\boldsymbol{e}^1_{-1}} = (0,0,0,0,0,1,1,-4,-4,1,0,0,1,1)$$

with the first five entries vertical 1-cells, the remaining horizontal 1-cells, therefore cannot cup nontrivially (to $\boldsymbol{e}^2_{-3}$).

At this point, we have yet to calculate the actual cohomology groups of Ω_ς over $\mathbb{Z}$. We follow the two-dimensional analogue of the Barge–Diamond filtration presented in [6]. For the sake of brevity, and since the calculations are in Example 3.5, here we skip most of the details and just state the conclusion.

The long exact sequence in relative cohomology of a pair (Ξ_1, Ξ_0) yields the short exact sequence

$$0 \longrightarrow \check{H}^1(\Xi_1,\Xi_0) \longrightarrow \check{H}^1(\Xi_1) \longrightarrow \check{H}^1(\Xi_0) \longrightarrow 0,$$

where

$$\check{H}^1(\Xi_1,\Xi_0) = \varinjlim\left(\mathbb{Z}^3, \begin{pmatrix} 5 & 0 & 0\\ 0 & 4 & 1\\ 0 & 1 & 4 \end{pmatrix}\right),$$

and of (Ξ_2, Ξ_1) yields

$$0 \longrightarrow \check{H}^1(\Xi_2,\Xi_1) \longrightarrow \check{H}^1(\Xi_2) \longrightarrow \check{H}^1(\Xi_1) \xrightarrow{\delta^1_1} \check{H}^2(\Xi_2,\Xi_1) \longrightarrow \check{H}^2(\Xi_2) \longrightarrow \check{H}^2(\Xi_1) \longrightarrow 0,$$

where

$$\check{H}^2(\Xi_2, \Xi_1) = \varinjlim\left(\mathbb{Z}^3, \begin{pmatrix} 11 & 9 & 5 \\ 14 & 6 & 5 \\ 4 & 1 & 20 \end{pmatrix}\right).$$

The eigenvalues of these matrices are 5, 3 and 25, 15, -3, respectively. A short calculation gives that $\check{H}^1(\Xi_0) = \mathbb{Z}$. Thus $\check{H}^1(\Xi_1, \Xi_0)$ is a direct summand of $\check{H}^1(\Xi_1)$. By Schur's lemma, $\delta_1^1 = 0$. Furthermore, since Ξ_2 is a wedge of 2-spheres, $\check{H}^1(\Xi_2, \Xi_1) = 0$. Thus $\check{H}^1(\Omega_S) = \check{H}^1(\Xi_2) = \check{H}^1(\Xi_1)$, which contains $\check{H}^1(\Xi_1, \Xi_0)$ as a direct summand.

However, the direct limit of the matrix

$$\sigma = \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}$$

does not split as a direct sum $\mathbb{Z}[1/5] \oplus \mathbb{Z}[1/3]$![3] To see this, consider any

$$\iota : \mathbb{Z}[1/5] \to \varinjlim(\mathbb{Z}^2, \sigma).$$

Since $1 \in \mathbb{Z}[1/5]$ is infinitely divisible by 5, its image must be as well, thus this map must be of the form $1 \mapsto (c, c)$, an eigenvector with eigenvalue 5, with $c \in \mathbb{Z}$. By writing

$$\begin{array}{ccc} \mathbb{Z} & \xrightarrow{\iota_0} & \mathbb{Z}^2 \\ \downarrow{\scriptstyle \cdot 5} & & \downarrow{\scriptstyle \sigma} \\ \mathbb{Z} & \xrightarrow{\iota_1} & \mathbb{Z}^2 \\ \downarrow{\scriptstyle \cdot 5} & & \downarrow{\scriptstyle \sigma} \\ \vdots & & \vdots \\ \downarrow & & \downarrow \\ \mathbb{Z}[1/5] & \xrightarrow{\iota} & \varinjlim(\mathbb{Z}^2, \sigma) \end{array}$$

where each $\iota_n : \mathbb{Z} \to \mathbb{Z}^2$ is $1 \mapsto (c, c)$,

$$\operatorname{coker} \iota = \varinjlim \operatorname{coker} \iota_n.$$

Since $\mathbb{Z} \leq \operatorname{coker} \iota_n$, without loss of generality, let us assume $c = 1$ and $\operatorname{coker} \iota_n = \mathbb{Z}$, with $\mathbb{Z}^2 \to \operatorname{coker} \iota_n$ given by $(a, b) \mapsto a - b$. Choosing a right splitting $\operatorname{coker} \iota_n \to \mathbb{Z}^2$ to be $1 \mapsto (1, 0)$ gives that the induced map

$$\sigma_* : \operatorname{coker} \iota_n \to \operatorname{coker} \iota_{n+1}$$

[3] If one has the matrix $\sigma = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right)$ instead, by [19, Exercise 7.5.2 (a)], the direct limit does not split as a direct sum of the eigenspaces. However, it is still a direct sum $\mathbb{Z}[1/3] \oplus \mathbb{Z}$ with $\mathbb{Z}[1/3]$ generated by the eigenvector with eigenvalue 3, since $(\varinjlim(\mathbb{Z}^2, \sigma))/\mathbb{Z}[1/3] = \mathbb{Z}$, and the short exact sequence

$$0 \longrightarrow \mathbb{Z}[1/3] \longrightarrow \varinjlim(\mathbb{Z}^2, \sigma) \longrightarrow \mathbb{Z} \longrightarrow 0$$

still splits. Note that $\mathbb{Z}$ *is not generated by the eigenvector with eigenvalue* 1!

is multiplication by 3, and

$$\operatorname{coker}\iota = \mathbb{Z}[1/3].$$

To see that this right splitting does not induce a splitting in the limit, take $1/3 \in \mathbb{Z}[1/3]$. This maps to $(1/3, 0)$. By definition, if it belongs to the direct limit, there must be some sufficiently large $k \in \mathbb{N}$ so that $\sigma^k(1/3, 0) \in \mathbb{Z}^2$, but

$$\sigma^k\begin{pmatrix}\frac{1}{3}\\0\end{pmatrix} = \tfrac{1}{2}\sigma^k\left(\begin{pmatrix}\frac{1}{3}\\\frac{1}{3}\end{pmatrix} + \begin{pmatrix}\frac{1}{3}\\-\frac{1}{3}\end{pmatrix}\right) = \tfrac{1}{2\cdot 3}\left(5^k\begin{pmatrix}1\\1\end{pmatrix} + 3^k\begin{pmatrix}1\\-1\end{pmatrix}\right) \notin \mathbb{Z}^2.$$

Given any right splitting, since the image of $1/3^n$ for sufficiently large $n \in \mathbb{N}$ under the splitting can be decomposed like this to not land in $\mathbb{Z}^2$ under any σ^k, any short exact sequence of the form

$$0 \longrightarrow \mathbb{Z}[1/5] \xrightarrow{\iota} \varinjlim(\mathbb{Z}^2, \sigma) \longrightarrow \mathbb{Z}[1/3] \longrightarrow 0$$

cannot be split.

Finally, applying the same argument to the map

$$\mathbb{Z}[1/3] \to \varinjlim(\mathbb{Z}^2, \sigma)$$

shows that the direct limit does not split as $\mathbb{Z}[1/5] \oplus \mathbb{Z}[1/3]$.

Therefore, this is an example where the cohomology groups over $\mathbb{Q}$ are indistinguishable from those of a product, and cup product can be used to differentiate the spaces there, but the cohomology groups over $\mathbb{Z}$ *also already differ from that of a product*!

Lastly, let us compute the frequency module. The substitution matrix on the 2-cells of the dual complex is $\sigma = \sigma_{A3}$ found in the Appendix which has Perron–Frobenius eigenvector

$$(1913, 962, 5250, 1750, 2213, 1775, 1775, 1750, 1113, 612, 912, 813,\\ 4575, 1750, 1750, 675, 17500, 1725, 675, 813, 375, 1350, 474).$$

Its sum is 52500, giving us that the frequency module is $\frac{1}{84}\mathbb{Z}[1/5]$, which is different from the frequency module of the product.

For illustration, we also compute the frequency modules for substitution on the vertical and horizontal 1-cells of the dual complex. The substitution matrices are

$$\sigma_v^1 = \begin{pmatrix}0&0&3&0&0\\0&0&1&0&0\\5&5&0&5&5\\0&0&1&0&0\\0&0&0&0&0\end{pmatrix} \quad\text{and}\quad \sigma_h^1 = \begin{pmatrix}0&0&0&4&3&0&0&0&3\\2&2&2&0&0&0&0&0&0\\0&0&0&0&1&1&1&1&1\\1&1&0&0&0&0&0&0&0\\1&1&2&0&0&0&0&0&0\\1&1&1&1&1&0&0&1&1\\0&0&0&0&0&0&0&0&0\\0&0&0&0&0&4&4&3&0\\0&0&0&0&0&0&0&0&0\end{pmatrix}.$$

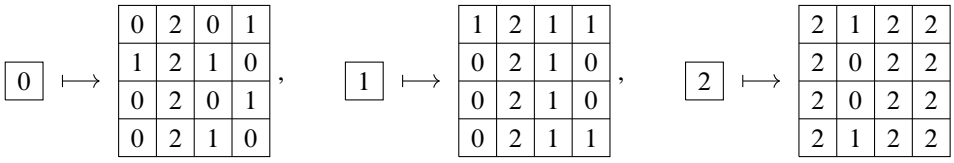


Figure 3: Substitution rule for an expansion 4 two-dimensional cubical substitution that has the cohomology groups of a product, but does not have the ring structure of a product.

The Perron–Frobenius eigenvectors that are duals of lifts of the generators of $\check{H}^1(\Omega_\varsigma)$ of eigenvalue 5 are $(3,1,5,1,0)$ and $(4,6,5,2,4,7,0,14,0)$. They have sums 10 and 42, respectively, giving frequency modules $\frac{1}{2}\mathbb{Z}[1/5]$ and $\frac{1}{42}\mathbb{Z}[1/5]$, respectively.

Example 3.5 (expansion 4 nonproduct) In the final example of this section, we consider the following two-dimensional substitution on three prototiles (see Figure 3):

```
ς=[[0,0,1,0,2,2,2,2,1,0,1,0,0,1,0,1],[0,0,0,1,2,2,2,2,1,1,1,1,1,0,0,1],
                                      [2,2,2,2,1,0,0,1,2,2,2,2,2,2,2,2]]
```

This, again, does not force the (see Remark 2.1) border, but is primitive and recognizable.

Using the dual complex, $\check{H}^2(\Omega_\varsigma)$ and $\check{H}^1(\Omega_\varsigma)$ are computed via direct limits under the matrices

$$\sigma^2 = \begin{pmatrix} 1 & 2 & 0 & 0 & 0 & 2 & 1 & 3 & 0 & 0 & 0 & 2 & 0 \\ 4 & 4 & 2 & -1 & -1 & 4 & -1 & -1 & 6 & 8 & 0 & 0 & 0 \\ 4 & 2 & 5 & 0 & 0 & 3 & 0 & 0 & 3 & 4 & 0 & 0 & 1 \\ 2 & 2 & 0 & 0 & 0 & 1 & 0 & 3 & 0 & 0 & 1 & 3 & 0 \\ 4 & 2 & 4 & 1 & 1 & 2 & 0 & 2 & 0 & 0 & 1 & 3 & 0 \\ 3 & 2 & 4 & 1 & 1 & 3 & 1 & 2 & 0 & 0 & 0 & 2 & 0 \\ 8 & 4 & 8 & 1 & 2 & 4 & 1 & 5 & 0 & 0 & 2 & 5 & 1 \\ -2 & 0 & -4 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 2 & 2 & 1 & 0 & 0 & 2 & -1 & -1 & 3 & 4 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 8 & 0 & 0 & 0 \\ 4 & 2 & 4 & 1 & 1 & 2 & 0 & 2 & 0 & 0 & 1 & 3 & 0 \\ 6 & 2 & 8 & 1 & 2 & 3 & 1 & 2 & 0 & 0 & 1 & 2 & 1 \\ 2 & 0 & 4 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \sigma^1 = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 1 & 3 & -1 & -1 \\ 1 & -1 & 2 & 0 \\ 1 & -1 & 0 & 2 \end{pmatrix},$$

with eigenvectors

$$\begin{aligned} \boldsymbol{e}^1_{4_1} &= (2,0,1,1), & \boldsymbol{e}^2_{16} &= (1,2,2,1,2,2,4,-1,1,1,2,3,1), \\ \boldsymbol{e}^1_{4_2} &= (0,2,-1,-1), & \boldsymbol{e}^2_{8} &= (1,-2,0,1,2,2,4,-1,-1,-1,2,3,1), \\ \boldsymbol{e}^1_{2} &= (0,0,1,-1), & \boldsymbol{e}^2_{4} &= (1,-4,-1,1,2,2,4,-1,-2,1,2,3,1), \\ \boldsymbol{e}^1_{1} &= (0,1,1,1), & \boldsymbol{e}^2_{2} &= (1,0,-1,1,0,0,0,1,0,0,0,-1,-1), \end{aligned}$$

where the subscripts indicate the corresponding eigenvalues. The cup product is the bilinear form

$$\check{H}^1 \times \check{H}^1 \to \check{H}^2 : (a,b) \mapsto B(a,b) := \sum_{*\in\{16,8,4,2\}} B_*(a,b)\cdot \boldsymbol{e}_*^2,$$

where $\boldsymbol{e}_n^2 \in \check{H}^2$ is a generator with eigenvalue n, and B_n is the bilinear form given by the matrices

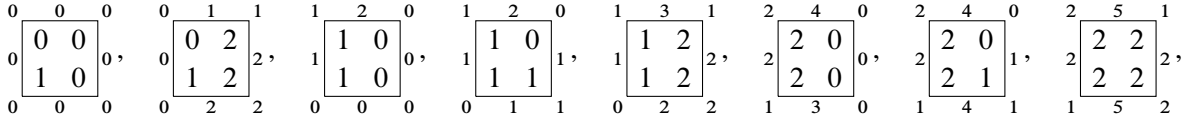
$$B_{16} := \begin{pmatrix} 0 & 2 & 0 & 0 \\ -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B_8 := \begin{pmatrix} 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B_4 := \begin{pmatrix} 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{pmatrix}, \quad B_2 := \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Like the previous example, the last component of the bilinear form shows that there are no cohomology classes in $\check{H}^1(\Omega_\varsigma)$ cupping to $\boldsymbol{e}_2^2$ in $\check{H}^2(\Omega_\varsigma)$, contrasting the example coming from the product. This coincides with the fact that, up to a coboundary, the lifts of the eigenvectors $\boldsymbol{e}_2^1$ and $\boldsymbol{e}_1^1$ to 1-cochains are

$$\widetilde{\boldsymbol{e}_2^1} = (0,0,0,0,0,0,0,0,0,1,1,0,1,1,-1), \quad \widetilde{\boldsymbol{e}_1^1} = (0,0,0,0,0,-1,0,0,1,1,-1,1,1,1)$$

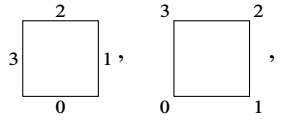
with the first five entries vertical 1-cells, the remaining horizontal 1-cells, therefore cannot cup nontrivially (to $\boldsymbol{e}_2^2$).

Let us now calculate the actual cohomology groups of Ω_ς over $\mathbb{Z}$ following the two-dimensional analogue of the Barge–Diamond filtration. Let us begin by drawing S_0 restricted to the eventual range



where the boundary identifications are *only along the respective positions*, e.g., none of the bottom horizontal 1-cells are identified to any of the top horizontal 1-cells, despite the numbering, and none of the bottom right 0-cells are identified to any of the top right 0-cells, despite the numbering.

We arrange the 1-cells and the 0-cells on S_0 in the order



then by the numbering in S_0. This gives us the coboundary matrices

$$\partial_0^1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 \end{pmatrix},$$

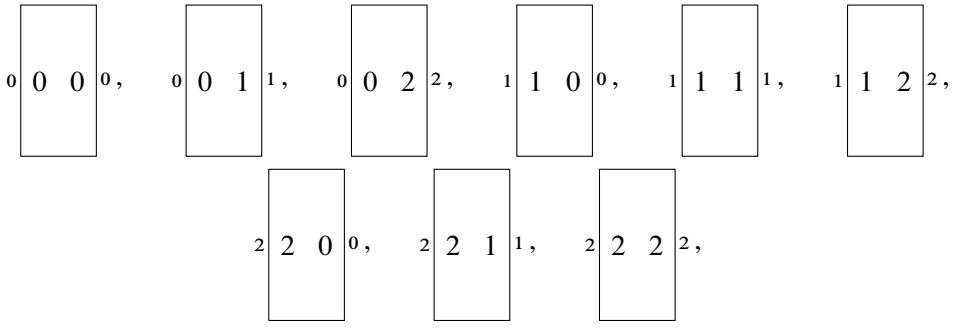
$$\partial_0^0 = \begin{pmatrix} -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Direct calculation shows that

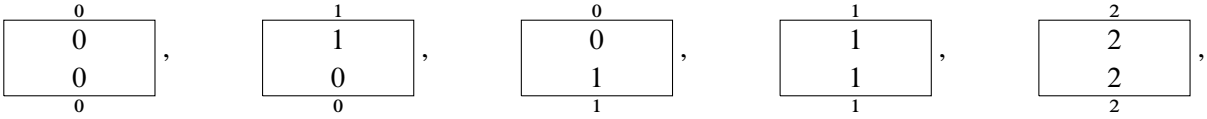
$$\check{H}^2(\Xi_0) = 0, \quad \check{H}^1(\Xi_0) = \mathbb{Z} = \langle \overline{(1,0,0,0,-1,0,0,1,0,0,-1,0,-1,0,0,1,1,0)} \rangle, \quad \check{H}^0(\Xi_0) = \mathbb{Z},$$

where the vector is a 1-cochain in S_0.

The extra cells that we add on to S_0 to form S_1 are the same as the previous example, being the vertical 1-flaps



where the top and the bottom horizontal 1-cells are identified to their appropriate bottom and top horizontal 1-cells in S_0, and the horizontal 1-flaps



where the left and the right vertical 1-cells are identified to their appropriate right and left vertical 1-cells in S_0. The space S_1/S_0 has three generating loops up to homotopy, one from the vertical 1-flaps, one

from the first four horizontal 1-flaps, and one from the last horizontal 1-flap. We then compute that

$$\check{H}^1(\Xi_1, \Xi_0) = \varinjlim\left(\mathbb{Z}^3, \begin{pmatrix} 4 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 1 & 3 \end{pmatrix}\right)$$

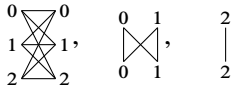
with the order of the entries of the matrix given as described. The matrix has eigenvectors

$$\boldsymbol{e}_{4_1} = (1, 0, 0), \quad \boldsymbol{e}_{4_2} = (0, 1, 1), \quad \boldsymbol{e}_2 = (0, 1, -1),$$

with the subscripts their respective eigenvalues. Direct computation then shows that the direct limit splits as a direct sum, giving us

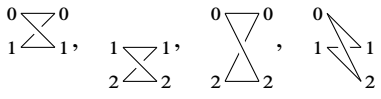
$$\check{H}^1(\Xi_1, \Xi_0) = \mathbb{Z}[1/4]^2 \oplus \mathbb{Z}[1/2].$$

The group $\check{H}^2(\Xi_1, \Xi_0)$ requires more effort. We observe that the generating 2-spheres of S_1/S_0, up to homotopy, occur when 1-flaps form a "tube", e.g., the first four of the horizontal 1-flaps. Let us represent all such tubes by graphs that represents the horizontal and vertical cross sections of vertical and horizontal 1-flaps, respectively, with the vertices the 1-cells in S_1 not attached to S_0, and the edges the 1-flaps themselves. In this example, they are, drawn in the orientation matching the cross sections of the respective 1-flaps,

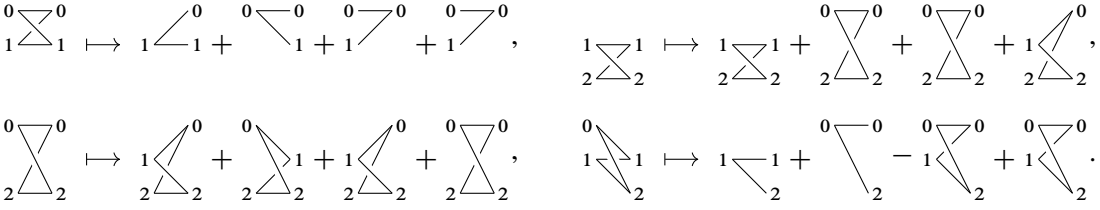


with labels of the vertices corresponding to the labels of the 1-cells marked in the set of 1-flaps.

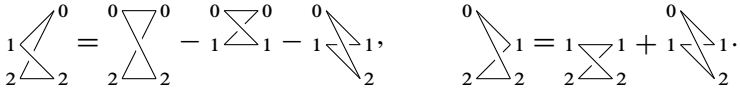
The first graph corresponding to vertical 1-flaps has a cycle basis consisting of four elements, therefore four associated generating 2-spheres in S_1/S_0, which we pick to be



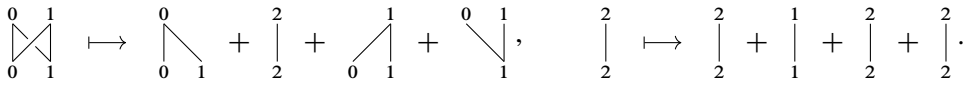
oriented so that the top edge is left-to-right. In fact, let us orient for all other cycles the same way. The substitution rule then dictates that



The cycles not in our basis decompose as



A similar calculation for the second and the third graphs gives us



Observing that the substitution rule is trivial on the second and the third graphs indicates that we can calculate $\check{H}^2(\Xi_1, \Xi_0)$ using only the vertical 1-flaps, which is

$$\check{H}^2(\Xi_1, \Xi_0) = \varinjlim \left(\mathbb{Z}^4, \begin{pmatrix} 0 & 0 & 0 & 0 \\ -1 & 1 & 3 & -1 \\ -2 & 1 & 3 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \right),$$

where the entries are ordered in the cycle basis we selected. Finally,

$$\check{H}^2(\Xi_1, \Xi_0) = \mathbb{Z}[1/4] = \langle (0, 1, 1, 0) \rangle.$$

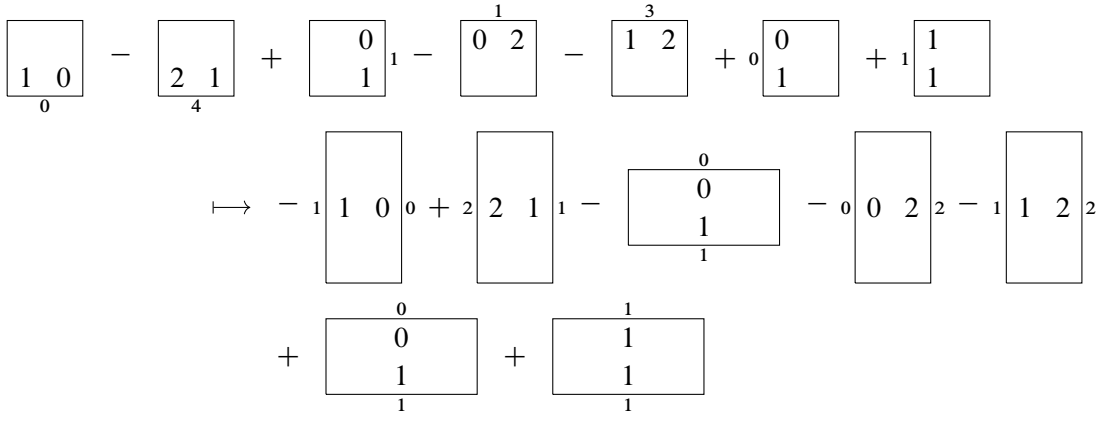
Thus the long exact sequence in relative cohomology of a pair (Ξ_1, Ξ_0) reads

$$\begin{aligned} 0 \longrightarrow \check{H}^0(\Xi_1, \Xi_0) \longrightarrow \check{H}^0(\Xi_1) \xrightarrow[\delta_0^0]{} \mathbb{Z} \longrightarrow \\ \mathbb{Z}[1/4]^2 \oplus \mathbb{Z}[1/2] \longrightarrow \check{H}^1(\Xi_1) \xrightarrow[\delta_0^1]{} \mathbb{Z} \longrightarrow \\ \mathbb{Z}[1/4] \longrightarrow \check{H}^2(\Xi_1) \longrightarrow 0 \longrightarrow 0. \end{aligned}$$

Recalling that

$$\check{H}^1(\Xi_0) = \mathbb{Z} = \langle \overline{(1, 0, 0, 0, -1, 0, 0, 1, 0, 0, -1, 0, -1, 0, 0, 1, 1, 0)} \rangle$$

we check the image of the generator under δ_0^1, which is



and is contractible, so $\delta_0^1 = 0$. This is the same conclusion reached by Schur's lemma. Thus, working over reduced cohomology,

$$\check{H}^2(\Xi_1) = \mathbb{Z}[1/4], \quad \check{H}^1(\Xi_1) = \mathbb{Z}[1/4]^2 \oplus \mathbb{Z}[1/2] \oplus \mathbb{Z}.$$

Towards the relative cohomology of a pair (Ξ_2, Ξ_1), we have that

$$\check{H}^2(\Xi_2, \Xi_1) = \varinjlim \left(\mathbb{Z}^3, \begin{pmatrix} 7 & 5 & 4 \\ 5 & 7 & 4 \\ 2 & 2 & 12 \end{pmatrix} \right)$$

with eigenvectors

$$\boldsymbol{e}_{16} = (1, 1, 1), \quad \boldsymbol{e}_8 = (1, 1, -1), \quad \boldsymbol{e}_2 = (1, -1, 0).$$

A straightforward calculation and the observation that S_2/S_1 is a wedge of 2-spheres give us

$$\check{H}^2(\Xi_2, \Xi_1) = \mathbb{Z}[1/16] \oplus \mathbb{Z}[1/8] \oplus \mathbb{Z}[1/2], \quad \check{H}^1(\Xi_2, \Xi_1) = 0.$$

The long exact sequence in relative cohomology of a pair (Ξ_2, Ξ_1) then reads

$$\begin{array}{l} 0 \longrightarrow \check{H}^0(\Xi_2, \Xi_1) \longrightarrow \check{H}^0(\Xi_2) \xrightarrow[\delta_1^0]{} \mathbb{Z} \\ \quad \hookrightarrow 0 \longrightarrow \check{H}^1(\Xi_2) \xrightarrow[\delta_1^1]{} \mathbb{Z}[1/4]^2 \oplus \mathbb{Z}[1/2] \oplus \mathbb{Z} \\ \quad\quad \hookrightarrow \mathbb{Z}[1/16] \oplus \mathbb{Z}[1/8] \oplus \mathbb{Z}[1/2] \longrightarrow \check{H}^2(\Xi_2) \longrightarrow \mathbb{Z}[1/4] \longrightarrow 0. \end{array}$$

By Schur's lemma, to show that $\delta_1^1 = 0$, it suffices to check that $\delta_1^1(0, 1, -1) = 0$, where $(0, 1, -1)$ is the generator of $\mathbb{Z}[1/2] \leq \check{H}^1(\Xi_1)$, which is straightforward. In fact, δ_1^1 applied to each of the generating loops in S_1/S_0 (then lifted to a 1-cochain in S_1, modulo homotopy) yields as many 2-flaps of the same type on one side of a loop as the other.

In the very last step, we observe that $\check{H}^2(\Xi_2, \Xi_1)$ and $\check{H}^2(\Xi_1)$ have every element 2-divisible. We first show that $\check{H}^2(\Xi_2)$ is 2-divisible as well. Consider any $c \in \check{H}^2(\Xi_2)$. Let $d \in \check{H}^2(\Xi_1)$ so that $c \mapsto d$. By assumption, $d/2 \in \check{H}^2(\Xi_1)$. Let $c' \in \check{H}^2(\Xi_2)$ so that $c' \mapsto d/2$. Then $c - 2c' \mapsto 0$, and $c - 2c' \in \check{H}^2(\Xi_2, \Xi_1)$. Again, by assumption, $(c - 2c')/2 \in \check{H}^2(\Xi_2, \Xi_1) \leq \check{H}^2(\Xi_2)$. Then $2((c - 2c')/2 + c') = c$, and $\check{H}^2(\Xi_2)$ is 2-divisible. To construct a splitting map $\check{H}^2(\Xi_1) \to \check{H}^2(\Xi_2)$, take any $c \in \check{H}^2(\Xi_2)$ so that $c \mapsto 1$. The map is then defined by $a/2^n \mapsto ac/2^n$, where $a/2^n \in \check{H}^2(\Xi_1)$, and $c/2^n$ exists since $\check{H}^2(\Xi_2)$ is 2-divisible.

Thus the bottom short exact sequence splits, giving us that, over $\mathbb{Z}$,

$$\check{H}^2(\Omega_S) = \mathbb{Z}[1/16] \oplus \mathbb{Z}[1/8] \oplus \mathbb{Z}[1/4] \oplus \mathbb{Z}[1/2], \quad \check{H}^1(\Omega_S) = \mathbb{Z}[1/4]^2 \oplus \mathbb{Z}[1/2] \oplus \mathbb{Z}, \quad \check{H}^0(\Omega_S) = \mathbb{Z},$$

which are exactly the cohomology groups of a product. Therefore, this is an example over $\mathbb{Z}$ that the cup-product structure discerns from a product space.

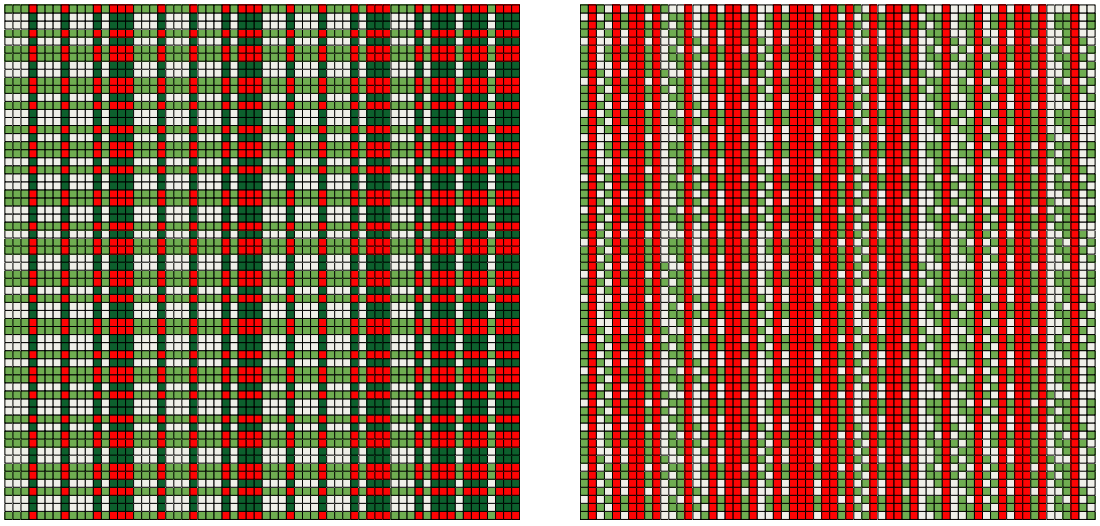


Figure 4: Patches from Examples 3.3 and 3.5, respectively. That their tiling spaces are not homeomorphic can only be detected through the cup product.

Lastly, let us compute the frequency module. The substitution matrix on the 2-cells of the dual complex is $\sigma = \sigma_{A4}$ which is found in the Appendix and has Perron–Frobenius eigenvector

$$(2512, 2048, 2048, 2304, 55, 112, 2352, 121, 696, 2352, 1792, 4096,$$
$$80, 313, 791, 712, 448, 216, 976, 384, 168).$$

Its sum is 24576, giving us that the frequency module is $\frac{1}{3}\mathbb{Z}[1/2]$, making it indistinguishable from the frequency module of the product.

For illustration, we also compute the frequency modules for substitution on the vertical and horizontal 1-cells of the dual complex. There are two components to the vertical 1-cells. The substitution matrices are then

$$\sigma^1_{v_1} = (4), \quad \sigma^1_{v_2} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 2 & 2 & 2 & 2 \end{pmatrix}, \quad \sigma^1_h = \begin{pmatrix} 2 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

By replacing $\boldsymbol{e}^1_{4_1}$ with $\boldsymbol{e}^1_{4_1} + \boldsymbol{e}^1_{4_2} = (2, 2, 0, 0)$, the Perron–Frobenius eigenvectors that are duals of lifts of the generators of $\check{H}^1(\Omega_\varsigma)$ of eigenvalue 4 are (1), $(1, 1, 0, 2)$, and $(2, 3, 1, 0, 1, 2, 0, 1, 2)$ with the sum of the first two the lift of the new $\boldsymbol{e}^1_{4_1}$. They have sums 1, 4, and 12, respectively, giving frequency modules $\mathbb{Z}[1/2]$, $\mathbb{Z}[1/2]$, and $\frac{1}{3}\mathbb{Z}[1/2]$, respectively.

The patches from Examples 3.3 and 3.5 are shown in Figure 4.

4 Breakdown of the Chern character in dimension four

This section contains what we consider the main result of this paper: a cubical substitution in dimension four where the Chern character cannot factor as an integral isomorphism. The case of dimension four is very special, since the complex K-theory and Chern character can be completely determined from the cohomology ring [26], even though the K-theory ring and cohomology ring may not be isomorphic. An immediate consequence of [26] is that if Ω is a 4-dimensional complex and the Chern character on $K^0(\Omega)$ does *not* factor through $H^4(\Omega;\mathbb{Z})$, then there must be an element $[c]$ of $H^2(\Omega;\mathbb{Z})$ whose cup square $[c]^2 = [c] \smile [c]$ is not divisible by 2 in $H^4(\Omega;\mathbb{Z})$.

4.1 The example

Let us consider the following four-dimensional expansion 3 substitution on eight prototiles:

```
ς=[[0,1,0,2,3,2,0,1,0,2,3,2,0,1,0,2,3,2,0,1,0,2,3,2,0,1,0,2,3,
    2,0,1,0,0,1,0,0,1,0,2,3,2,0,1,0,2,3,2,0,1,0,2,3,2,0,1,0,2,
                                                  3,2,0,1,0,2,3,2,0,1,0],
   [1,2,1,3,4,3,1,2,1,3,4,3,1,2,1,3,4,3,1,2,1,3,4,3,1,2,1,3,4,3,1,2,1,3,4,
    3,1,2,1,1,2,1,1,2,1,3,4,3,1,2,1,3,4,3,1,2,1,3,4,3,1,2,1,3,4,3,1,2,1,3,
                                                  4,3,1,2,1,3,4,3,1,2,1],
   [2,3,2,4,5,4,2,3,2,4,5,4,2,3,2,4,5,4,2,3,2,4,5,4,2,3,2,4,5,4,2,3,2,4,5,
    4,2,3,2,2,3,2,2,3,2,4,5,4,2,3,2,4,5,4,2,3,2,4,5,4,2,3,2,4,5,4,2,3,2,4,
                                                  5,4,2,3,2,4,5,4,2,3,2],
   [3,4,3,5,6,5,3,4,3,5,6,5,3,4,3,5,6,5,3,4,3,5,6,5,3,4,3,5,6,5,3,4,3,5,6,
    5,3,4,3,3,4,3,3,4,3,5,6,5,3,4,3,5,6,5,3,4,3,5,6,5,3,4,3,5,6,5,3,4,3,5,
                                                  6,5,3,4,3,5,6,5,3,4,3],
   [4,5,4,6,7,6,4,5,4,6,7,6,4,5,4,6,7,6,4,5,4,6,7,6,4,5,4,6,7,6,4,5,4,6,7,
    6,4,5,4,4,5,4,4,5,4,6,7,6,4,5,4,6,7,6,4,5,4,6,7,6,4,5,4,6,7,6,4,5,4,6,
                                                  7,6,4,5,4,6,7,6,4,5,4],
   [5,6,5,7,0,7,5,6,5,7,0,7,5,6,5,7,0,7,5,6,5,7,0,7,5,6,5,7,0,7,5,6,5,7,0,
    7,5,6,5,5,6,5,5,6,5,7,0,7,5,6,5,7,0,7,5,6,5,7,0,7,5,6,5,7,0,7,5,6,5,7,
                                                  0,7,5,6,5,7,0,7,5,6,5],
   [6,7,6,0,1,0,6,7,6,0,1,0,6,7,6,0,1,0,6,7,6,0,1,0,6,7,6,0,1,0,6,7,6,0,1,
    0,6,7,6,6,7,6,6,7,6,0,1,0,6,7,6,0,1,0,6,7,6,0,1,0,6,7,6,0,1,0,6,7,6,0,
                                                  1,0,6,7,6,0,1,0,6,7,6],
   [7,0,7,1,2,1,7,0,7,1,2,1,7,0,7,1,2,1,7,0,7,1,2,1,7,0,7,1,2,1,7,0,7,1,2,
    1,7,0,7,7,0,7,7,0,7,1,2,1,7,0,7,1,2,1,7,0,7,1,2,1,7,0,7,1,2,1,7,0,7,1,
                                                  2,1,7,0,7,1,2,1,7,0,7]]
```

The underlying idea is that in dimension four, we look for a substitution with a torsion term in $\check{H}^4(\Omega_\varsigma;\mathbb{Z})$. Furthermore, since the Chern character in dimension four involves a division by 2, for it to not be an isomorphism over $\mathbb{Z}$, we want $\check{H}^4(\Omega_\varsigma;\mathbb{Z})$ to contain even torsion, since odd torsion has multiplication by 2 invertible.

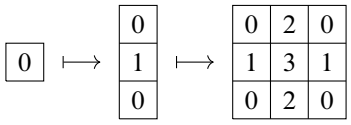


Figure 5: A two-dimensional analogue of the main example in Section 4.1. We begin with a one-dimensional checkerboard substitution on 8 prototiles (drawn vertically in the first map; only the first supertile is drawn, the rest are obtained by $+1 \bmod 8$), then extending it along the remaining coordinate (horizontally) by another (different) checkerboard (second map). The resulting tiling has rotational symmetry about axis the pattern is extended along (horizontal here).

This example is built from checkerboard patterns (Figure 5). It appears that although checkerboard patterns do not necessarily force the border (see Remark 2.1), the level of supertiles required to exhaust all 1-collared[4] prototiles to construct the dual complex is not too large. Since we want even torsion, we create a checkerboard pattern on $[0,3]^3$, then extend in the last dimension with yet another checkerboard pattern that depends on the prototile in the pattern on $[0,3]^3$, taking the result to be the supertiles, so as to introduce four-fold rotational symmetry with odd expansion. For this particular example, the dual complex has 1120 4-cells, 1232 3-cells, 480 2-cells, 88 1-cells, and 8 0-cells. One can get by with fewer prototiles or simpler patterns (see the next subsection), but this is the first example we discovered where everything we desire is evident.

We only focus on $\check{H}^2(\Omega_\zeta;\mathbb{Z})$ and $\check{H}^4(\Omega_\zeta;\mathbb{Z})$, since we want to identify a class $[c]\in\check{H}^2(\Omega_\zeta;\mathbb{Z})$ so that $[c]^2=[c]\smile[c]\in\check{H}^4(\Omega_\zeta;\mathbb{Z})$ is not (uniquely) divisible by 2 (see (3)). $\check{H}^2(\Omega_\zeta;\mathbb{Z})$ is the direct limit computed via the matrix $\sigma^2=\sigma^2_{A5}$, which is found in the Appendix, and $\check{H}^4(\Omega_\zeta;\mathbb{Z})$ is computed via a matrix too large to be listed here (354×354), but it suffices to know that for the dual complex, $\check{H}^4\cong\mathbb{Z}_4^3\oplus F$ for some torsion-free group F. Since we only care about the even torsion terms, there being $\mathbb{Z}_4^3$ in $\check{H}^4$ of the dual complex, we write the substitution action on the dual complex as

$$\sigma^4=\left(\begin{array}{c|c}\sigma^4_{\text{torsion}} & *\\ \hline 0 & \sigma^4_{\text{torsion-free}}\end{array}\right)=\left(\begin{array}{ccc|c}1&0&0&\\0&2&3&*\\0&3&2&\\ \hline \multicolumn{3}{c|}{0} & \sigma^4_{\text{torsion-free}}\end{array}\right), \tag{7}$$

where $\sigma^4_{\text{torsion}}$ is the action restricted to the torsion part of $\check{H}^4$.

Let us consider the eigenvector of σ^2, with eigenvalue 3,

$$v=(0,0,0,0,1,1,0,0,1,0,0,0,0,0,-1,1,1,0,0,0,0,0,0,0).$$

Lifting it to a 2-cochain, cupping it with itself via the cubical cup-product formula given in [15], then projecting to $\check{H}^4$ of the dual complex gives the class

$$w=v\smile v=(2,0,2,\dots)\in\mathbb{Z}_4^3\oplus F,$$

[4]Although the top-dimensional cells are "half-collars", the construction of the complex requires knowledge of how they are identified in the tiling, thus another "half-collar" of the half-collared prototiles, i.e., 1-collared prototiles.

which is an eigenvector with eigenvalue 9. By (7), the first copy of $\mathbb{Z}_4$ is preserved in the direct limit. Since the entry of w in this copy of $\mathbb{Z}_4$ is 2, $w = v \smile v$ is not (uniquely) divisible by 2, and the Chern character cannot be an isomorphism over $\mathbb{Z}$. Incidentally, this provides a negative answer to the question posed on the last line of [9].

Following the same procedure as before, the Perron–Frobenius eigenvector of the substitution matrix on the 4-cells of the dual complex sums to 29952, giving us that the frequency module for this cubical substitution tiling is $\frac{1}{3328}\mathbb{Z}[1/3]$.

4.1.1 Additional examples We have discovered numerous other examples with the same features as the example above. Here we mention two simple ones.

The first example consists on a cubical substitution on three prototiles and expansion 3. Although this substitution rule is easier to define, the cohomological computations are more labor-intensive. The substitution rule is

```
ς=[[0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,
    1,0,1,0,1,1,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,
                                                  0,1,0,1,0,1,0,1,0,1,0],
   [1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,
    2,1,2,1,2,2,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,
                                                  1,2,1,2,1,2,1,2,1,2,1],
   [2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,
    0,2,0,2,0,0,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,2,0,
                                                  2,0,2,0,2,0,2,0,2,0,2]]
```

and is the analogue of the two-dimensional substitution rule found in Figure 6, bottom. We leave the details to the reader.

The second example is a cubical substitution on two prototiles and expansion 3. The substitution rule is

```
ς=[[0,0,0,0,0,1,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,1,0,0,0,0,0,0,0,1,0,0,0,1,0,
    0,0,0,0,0,1,1,0,1,0,1,0,0,0,1,0,0,0,1,0,1,0,1,0,0,0,0,0,1,0,0,0,1,0,0,
                                                  0,1,0,0,0,0,0,1,0,1,0],
   [1,1,1,1,1,0,1,0,1,1,1,0,1,1,1,0,1,1,1,0,1,0,1,1,1,1,1,1,1,0,1,1,1,0,1,
    1,1,1,1,1,0,0,1,0,1,0,1,1,1,0,1,1,1,0,1,0,1,0,1,1,1,1,1,1,0,1,1,1,0,1,1,
                                                  1,0,1,1,1,1,1,1,0,1,0,1]]
```

Its construction is much more involved, which we elect to forgo, and does not have two-dimensional analogues. We again leave the computational details to the reader.

All of the previous examples had a torsion-free generator in $\check{H}^2$ squaring to twice a torsion generator in $\check{H}^4$. To demonstrate that $\mathbb{RP}^4$-like behavior can also occur, that is, a torsion generator in $\check{H}^2$ can square to twice a torsion generator in $\check{H}^4$, we first give some motivation.

It turns out that the first example we provided in this section is a four-dimensional analogue of the three-prototile version of the squiral substitution (encoded slightly differently than usual). In $d = 2$, the squiral tiling was shown in [4] to be a substitution tiling with expansion 3 on squares using two prototiles

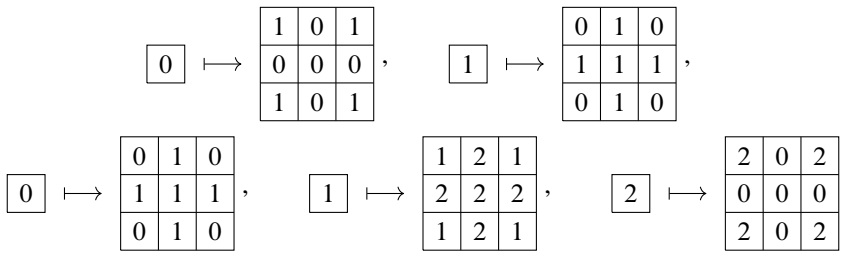


Figure 6: Top: the substitution rule that produces the squiral tiling. Bottom: the four-dimensional version of a three-prototile variant of the squiral substitution has the same features as the example in (7).

(Figure 6, top), and it was computed there that the tiling has singular continuous spectrum. If one writes down the induced substitution rule on collared prototiles (from the dual complex), one can impose a $\mathbb{Z}_2$-action that interchanges the two prototiles (Figure 7). It turns out that quotienting the squiral tiling space by this action results in another substitution tiling space, but with $\mathbb{Z}_2$-torsion in $\check{H}^2$ when there was no torsion in the original squiral tiling.

Motivated by this, let us consider

```
ς=[[(i+1)%2 if 1 not in t else i%2 for t in list(cartesian_product([range(3)]*d))] for i in range(2)]
```

which is a generalization of the squiral substitution to arbitrary d that is different from the one in the first example. For $d = 4$, the rule reads as

```
ς=[[1,0,1,0,0,0,1,0,1,0,0,0,0,0,0,0,0,1,0,1,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
                          0,0,0,0,0,0,0,0,0,0,0,1,0,1,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,1,0,1,0,0,0,1,0,1],
   [0,1,0,1,1,1,0,1,0,1,1,1,1,1,1,1,1,1,0,1,0,1,1,1,0,1,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
                                  1,1,1,1,1,1,0,1,0,1,1,1,0,1,0,1,1,1,1,1,1,1,1,1,0,1,0,1,1,1,0,1,0]]
```

We computed there to be 478 "half-collared" prototiles, and quotienting by the $\mathbb{Z}_2$-action resulted in 239 prototiles with an associated induced substitution rule, $\bar{\varsigma}$, that is too large to be written here. Fortunately, since the induced substitution on the half-collared prototiles of the squiral tiling forces the border (see Remark 2.1), as does the resulting induced substitution rule, thus it suffices to work with the uncollared AP-complex, which we denote by $\overline{\Gamma}$ (reserving Γ for the original substitution rule ς). It has 239 4-cells, 160 3-cells, 48 2-cells, 8 1-cells, and 1 0-cell.

We have that

$$\check{H}^2(\overline{\Gamma};\mathbb{Z}) = \mathbb{Z}_2 \oplus \mathbb{Z}^9, \quad \check{H}^4(\overline{\Gamma};\mathbb{Z}) = \mathbb{Z}_2^{14} \oplus \mathbb{Z}_4 \oplus \mathbb{Z}^{126}.$$

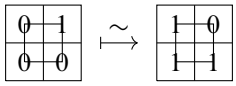


Figure 7: Two "half-collared" prototiles in the dual complex where quotienting by the $\mathbb{Z}_2$-action that interchanges the underlying prototiles identifies them.

The induced substitution matrices are

$$\sigma^2 = \left(\begin{array}{c|c} \sigma^2_{\text{torsion}} & * \\ \hline 0 & \sigma^2_{\text{torsion-free}} \end{array}\right) = \left(\begin{array}{c|c} 1 & * \\ \hline 0 & \sigma^2_{\text{torsion-free}} \end{array}\right), \quad \sigma^4 = \left(\begin{array}{c|c} \sigma^4_{\text{torsion}} & * \\ \hline 0 & \sigma^4_{\text{torsion-free}} \end{array}\right) = \left(\begin{array}{c|c} \mathrm{Id}_{15} & * \\ \hline 0 & \sigma^4_{\text{torsion-free}} \end{array}\right),$$

so the torsion terms stay in the limit and belong to the respective cohomology groups of $\Omega_{\bar{\varsigma}}$.

Let

$$v = (1,0,0,0,0,0,0,0,0,0) \in \check{H}^2(\overline{\Gamma};\mathbb{Z}).$$

Lifting it to a 2-cochain, cupping it with itself, then projecting to $\check{H}^4(\overline{\Gamma};\mathbb{Z})$ gives the class

$$w = v \smile v = (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,\dots) \in \check{H}^2(\overline{\Gamma};\mathbb{Z}),$$

where the single nonzero entry corresponds exactly to twice the generator of $\mathbb{Z}_4 \leq \check{H}^4(\overline{\Gamma};\mathbb{Z})$.

4.2 Gap labels

What does the example above say about the gap-labeling conjecture? What it *does not* say is that it is false: the example above relied heavily on the cohomology of the tiling space having nontrivial torsion terms, which were the source of the breakdown of the Chern character map. However, *the torsion is not detected by the Ruelle–Sullivan current* C_μ, a homomorphism to $\mathbb{R}$, meaning that the torsion is not detected in the frequency module. In addition, it is not clear that the image of the trace map on $K_0(\mathcal{A}_p(\Omega))$ is contained in the image of the trace map on $K_0(\mathcal{A}_{\mathrm{AF}}(\Omega))$, so we cannot compare $\tau\big(K_0(\mathcal{A}_p(\Omega))\big)$ with the frequency module, as we only know how to compute the image of the tracial state of the AF-algebra $\mathcal{A}_{\mathrm{AF}}(\Omega)$.

What it *does* say is that the gap in several papers relying on the Chern character factoring through integral cohomology to give (5) is real—our example shows it. We do not believe that a true counterexample to the gap-labeling conjecture—which would only exist in dimensions greater than three—will be found using cubical substitutions, and so the problem of finding a counterexample becomes much harder, as one needs to define a noncubical substitution rule in four dimensions (or higher) and compute its cohomology ring. This task seems out of reach at the moment.

4.3 An aperiodic counterexample to equivariant gap-labeling in dimension 4

Example 4.1 To provide an additional counterexample for Conjecture 2.3 that is a substitution tiling, we construct the simplest possible one that is not a solenoid by building off of Theorem 2.6. We aim for its complex to have the same 3-skeleton as T^4, but with an additional 4-cell attached in the same way as the original 4-cell in T^4.

To do so, consider the following four-dimensional substitution with expansion 3 on two prototiles:

```
ς=[[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
   0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
                                                    0,0,0,0,0,0,0,0,0,0],
  [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
   0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
                                                    0,0,0,0,0,0,0,0,0,0]]
```

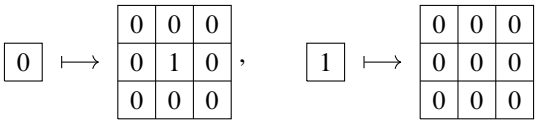


Figure 8: The four-dimensional version of this substitution rule is the same 3-skeleton as T^4, with the only difference being that there are two 4-cells (the two prototiles) attached to the 3-cells.

Its two-dimensional analogue is Figure 8.

This substitution is primitive and recognizable, and clearly forces the border (see Remark 2.1), thus we can use the *uncollared* AP-complex instead, which is the AP-construction without using collared prototiles. By the same theorem in [2], its inverse limit is homeomorphic to the tiling space.

The 2^4 patch of all 0's indicates that the complex contains a copy of T^4. The prototile 1 being present in all possible positions in a 2^4 patch asserts that its associated 4-cell is attached to the 3-skeleton in the same exact way as the 4-cell associated to the prototile 0.

All of the coboundary maps are trivial, giving us that the cohomology groups we are interested in are

$$\check{H}^4(\Omega_\zeta;\mathbb{Z}) = \varinjlim\left(\mathbb{Z}^2, \begin{pmatrix} 80 & 1 \\ 81 & 0 \end{pmatrix}\right), \quad \check{H}^2(\Omega_\zeta;\mathbb{Z}) = \varinjlim(\mathbb{Z}^6, 9\cdot \mathrm{Id}),$$

where $\mathbb{Z}^2$ is simultaneously C^4 and H^4 of the AP-complex, and $\mathbb{Z}^6$ is simultaneously C^2 and H^2 of the AP-complex. Due to this, the cubical cup product on the cohomology ring coincides with the cubical cup product on the cochains, which is easy to describe. For example, the two generators $c_1 = [0,1]\times[0,1]\times[0]\times[0]$, $c_2 = [0]\times[0]\times[0,1]\times[0,1] \in C^2$ cup to $(1,1) \in C^4$, the sum of the duals of the two prototiles. That is, the cohomology ring structure of this AP-complex only differs from that of T^4 by the cup product always witnessing both 4-cochains whenever it is nontrivial.

One then follows the same exact argument provided in Theorem 2.6 and concludes that this substitution tiling space, upon quotienting by the same $\mathbb{Z}_2$-action, yields a counterexample to Conjecture 2.3. To be precise, let Γ be the AP complex of the substitution described above, and let $\rho : \Gamma \to \mathbb{T}^4$ be the factor map which collapses the 4 cells. There is an involution on Γ which is equivariant (under ρ) with the involution of $\mathbb{T}^4 = \mathbb{T}^2 \times \mathbb{T}^2$ used in Theorem 2.6. This involution is defined by exchanging the 2-cells c_1 and c_2 described above. The induced map on cohomology $\rho^* : H^4(\mathbb{T}^4;\mathbb{Z}) \to H^4(\Gamma;\mathbb{Z})$ sends the generator in $H^4(\mathbb{T}^4;\mathbb{Z})$ to $(1,1) \in H^4(\Gamma;\mathbb{Z})$. Let $S_4 = \varprojlim(\mathbb{T}^4, 3\cdot\mathrm{Id})$ be the four dimensional solenoid constructed with maps of expansion 3, and note that it is a factor of the tiling space Ω corresponding to the substitution above. Then the argument from Theorem 2.6 carries over to S_4 in the direct limit since the expansion is by 3, and it can be pulled back to $H^*(\Omega;\mathbb{Z})$ using the map induced by the factor map (because the expansions of both systems are by 3 and thus cannot be divided by 2 by the Chern character). Thus this example provides an aperiodic counterexample to Conjecture 2.3.

Let us remark on the mechanism that yields these counterexamples. In the original AP-complex, the nontrivial squares are of the form $(c_1+c_2)\smile(c_1+c_2) = c_1\smile c_1 + c_1\smile c_2 + c_2\smile c_1 + c_2\smile c_2 = 2c_1\smile c_2$, which always return twice the sum of the generators in H^4, and therefore still yields integrality of

the Chern character. In this $\mathbb{Z}_2$ quotient, c_1 and c_2 are identified, so $c_1 \smile c_2 = c_1 \smile c_1$ (in the quotient), and we no longer need to add cochains together to produce a nontrivial square in H^4. This breaks the integrality.

Are there other ($\mathbb{Z}_2$) actions that produce the same effect? There are a few obvious choices that do not appear to work.

- If the substitution is on two prototiles, let the $\mathbb{Z}_2$-action be swapping the two prototiles, assuming that the supertiles are related by $+1 \bmod 2$ (so that the action is well defined). This does not appear to work, because "horizontal" cells remain "horizontal", and "vertical" cells remain "vertical", thus to obtain a nontrivial square, one still needs to add together both "horizontal" and "vertical", resulting in a factor of two on the 4-cochain.
- If the substitution has its supertiles symmetric about, say, the axis along $(1,0,0,0)$, let the $\mathbb{Z}_2$-action be flipping the axis. For the same reason as above, this also does not appear to work.

Appendix Large matrices

Here we define some larger matrices used in this paper:

$$\sigma_{A1} = \begin{pmatrix}
3&3&3&3&4&4&3&3&3&3&4&3&3&3&4&3&3&3&4&3&3&3&4&3&3&3&4&3&3&3&3&3&4&4&3&3\\
0&0&0&0&0&0&0&0&0&0&0&0&0&3&3&4&3&0&0&0&0&0&0&0&0&0&3&3&4&3&3&0&3&0&4&0&3&0\\
1&1&1&1&0&0&1&1&1&1&0&1&1&1&0&1&1&1&0&1&1&1&0&1&1&1&0&1&1&1&1&1&0&0&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&1&0&0&0&0&0&0&0&0&0&1&1&0&1&1&0&1&0&0&0&1&0\\
1&1&0&0&1&1&1&1&1&0&1&1&1&0&1&1&1&0&1&1&1&0&1&1&1&0&1&1&1&1&0&0&1&1&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&1&0&0&0&0&0&0&0&0&0&1&0&1&1&1&0&0&0&1&0&1&0\\
0&0&1&1&0&0&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&0&0&1&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0&0\\
3&3&3&3&4&4&3&3&3&3&4&3&0&0&0&0&3&3&4&3&3&3&4&3&0&0&0&0&0&3&0&3&0&4&0&3\\
1&1&1&1&0&0&1&1&1&1&0&1&0&0&0&0&1&1&0&1&1&1&0&1&0&0&0&0&0&1&0&1&0&0&0&1\\
1&1&0&0&1&1&1&1&1&0&1&1&0&0&0&0&1&0&1&1&1&0&1&1&0&0&0&0&0&1&0&0&0&1&0&1\\
0&0&1&1&0&0&0&0&0&1&0&0&0&0&0&0&0&1&0&0&0&1&0&0&0&0&0&0&0&0&0&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&6&6&8&6&0&0&0&0&0&0&0&0&6&6&8&6&6&0&6&0&8&0&6&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&2&2&0&2&0&0&0&0&0&0&0&0&2&2&0&2&2&0&2&0&0&0&2&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&2&0&2&2&0&0&0&0&0&0&0&0&2&0&2&2&2&0&0&0&2&0&2&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&2&0&0&0&0&0&0&0&0&0&0&0&2&0&0&0&0&2&0&0&0&0&0\\
3&3&3&3&4&4&3&3&3&3&4&3&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&1&1&0&0&1&1&1&1&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&0&0&1&1&1&1&1&0&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&6&6&8&6&6&6&8&6&0&0&0&0&0&6&0&6&0&8&0&6\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&2&2&0&2&2&2&0&2&0&0&0&0&0&2&0&2&0&0&0&2\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&2&0&2&2&2&0&2&2&0&0&0&0&0&2&0&0&0&2&0&2\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&2&0&0&0&2&0&0&0&0&0&0&0&0&2&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&3&3&4&3&0&0&0&0&0&0&0&0&3&3&4&3&3&0&3&0&4&0&3&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&1&0&0&0&0&0&0&0&0&1&1&0&1&1&0&1&0&0&0&1&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&1&0&0&0&0&0&0&0&0&1&0&1&1&1&0&0&0&1&0&1&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&1&0&0&0&0&0&0\\
3&3&3&3&4&4&3&3&3&3&4&3&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
3&3&3&3&4&4&3&3&3&3&4&3&0&0&0&0&3&3&4&3&3&3&4&3&0&0&0&0&0&3&0&3&0&4&0&3\\
1&1&1&1&0&0&1&1&1&1&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&1&1&0&0&1&1&1&1&0&1&0&0&0&0&1&1&0&1&1&1&0&1&0&0&0&0&0&1&0&1&0&0&0&1\\
1&1&0&0&1&1&1&1&1&0&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&0&0&1&1&1&1&1&0&1&1&0&0&0&0&1&0&1&1&1&0&1&1&0&0&0&0&0&1&0&0&0&1&0&1\\
0&0&1&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&1&0&0&0&0&0&1&0&0&0&0&0&0&0&1&0&0&0&1&0&0&0&0&0&0&0&0&1&0&0&0&0&0
\end{pmatrix},$$

$$\sigma_{A2} = \begin{pmatrix}
2&0&2&0&0&0&0&0&2&2&0&0&2&2&0&0\\
2&4&2&4&0&0&0&0&2&2&0&0&2&2&0&0\\
1&0&1&0&1&0&1&0&1&1&1&1&1&1&1&1\\
1&2&1&2&1&2&1&2&1&1&1&1&1&1&1&1\\
1&0&1&0&1&0&1&0&1&1&1&1&1&1&1&1\\
1&2&1&2&1&2&1&2&1&1&1&1&1&1&1&1\\
0&0&0&0&2&0&2&0&0&0&2&2&0&0&2&2\\
0&0&0&0&2&4&2&4&0&0&2&2&0&0&2&2\\
2&2&2&2&0&0&0&0&4&4&0&0&2&2&0&0\\
1&1&1&1&1&1&1&1&2&2&2&2&1&1&1&1\\
1&1&1&1&1&1&1&1&2&2&2&2&1&1&1&1\\
0&0&0&0&2&2&2&2&0&0&4&4&0&0&2&2\\
2&2&2&2&0&0&0&0&0&0&0&0&2&2&0&0\\
1&1&1&1&1&1&1&1&0&0&0&0&1&1&1&1\\
1&1&1&1&1&1&1&1&0&0&0&0&1&1&1&1\\
0&0&0&0&2&2&2&2&0&0&0&0&0&0&2&2
\end{pmatrix},$$

$$\sigma_{A3} = \begin{pmatrix}
1&1&2&3&1&1&2&2&2&1&3&3&0&0&0&0&0&1&1&2&3&3&2\\
0&0&1&1&0&0&2&3&2&0&1&1&0&0&0&0&0&0&0&2&1&1&2\\
3&3&0&0&3&4&2&0&2&3&3&3&3&3&3&3&3&4&4&2&3&3&2\\
1&1&0&0&1&0&2&0&2&1&1&1&1&1&1&1&1&0&0&2&1&1&2\\
2&2&3&3&2&2&1&2&1&2&2&2&0&0&0&0&0&2&2&1&2&2&1\\
1&1&2&1&1&1&2&4&2&1&2&2&0&0&0&0&0&1&1&2&2&2&2\\
1&1&2&3&1&2&1&2&1&1&2&2&0&0&0&0&0&2&2&1&2&2&1\\
1&1&0&0&1&1&1&0&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
2&2&2&1&2&1&0&0&0&2&0&0&0&0&0&0&0&1&1&0&0&0&0\\
0&0&0&0&0&0&2&2&2&0&1&1&0&0&0&0&0&0&0&2&1&1&2\\
1&1&1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&1&1&1&0&0&1\\
1&1&1&0&1&0&1&1&1&1&1&1&0&0&0&0&0&0&0&1&1&1&1\\
2&2&2&2&2&2&3&3&3&2&3&3&0&0&0&0&3&2&2&3&3&3&3\\
1&1&1&1&1&1&1&1&1&1&1&1&0&0&0&0&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&0&0&0&0&1&1&1&1&1&1&1\\
1&1&1&1&1&1&0&0&0&1&0&0&0&0&0&0&0&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&20&20&20&20&15&0&0&0&0&0&0\\
2&2&2&2&2&2&1&1&1&2&1&1&0&0&0&0&0&2&2&1&1&1&1\\
1&1&1&1&1&1&0&0&0&1&0&0&0&0&0&0&0&1&1&0&0&0&0\\
1&1&1&0&1&0&1&1&1&1&1&1&0&0&0&0&0&0&0&1&1&1&1\\
0&0&0&0&0&0&1&1&1&0&1&1&0&0&0&0&0&0&0&1&1&1&1\\
2&2&2&2&2&2&0&0&0&2&0&0&0&0&0&0&0&2&2&0&0&0&0\\
0&0&0&2&0&2&0&0&0&0&0&0&0&0&0&0&0&2&2&0&0&0&0
\end{pmatrix},$$

$$\sigma_{A4} = \begin{pmatrix}
2&1&1&1&2&2&2&2&2&3&1&1&2&2&2&2&2&2&3&2&2\\
2&2&2&2&2&2&2&2&2&0&2&1&2&0&0&0&0&2&0&0&0\\
2&2&2&2&2&2&2&2&2&0&2&1&2&0&0&0&0&2&0&0&0\\
0&1&1&1&0&0&0&0&0&4&1&1&0&4&4&4&4&0&4&4&4\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&1\\
0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&1&0&1&0&0&1\\
3&1&1&1&1&1&3&1&1&2&1&1&1&1&1&1&1&1&2&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&1&0&0&1&0\\
2&0&0&0&3&2&1&3&1&0&0&0&2&1&1&0&1&1&0&1&0\\
3&1&1&1&1&1&2&1&1&3&1&1&1&1&1&1&1&1&2&1&1\\
0&3&3&3&0&0&0&0&0&0&3&1&0&0&0&0&0&0&0&0&0\\
0&4&4&4&0&0&0&0&0&0&4&8&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&0&1&1&0&0&0&1&0&0&0&0&1&0&0&0\\
0&0&0&0&1&0&0&1&1&0&0&0&0&2&2&1&1&1&0&1&1\\
2&0&0&0&3&3&2&3&2&0&0&0&3&0&0&0&0&2&0&0&0\\
0&0&0&0&0&0&0&0&0&1&2&0&0&0&2&2&2&1&1&1&1&2\\
0&0&0&0&0&0&1&0&0&0&1&0&0&1&1&1&1&2&0&1&2&1\\
0&0&0&0&0&0&1&1&0&1&0&0&0&1&0&0&0&0&1&0&0&0\\
0&1&1&1&0&0&1&0&0&0&1&1&0&0&0&0&0&0&0&1&0&0\\
0&0&0&0&0&0&0&0&0&0&1&0&0&0&1&1&1&1&0&1&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&1&1&1
\end{pmatrix},$$

$$\sigma_{A5}^2 = \begin{pmatrix}
-8&-7&-4&-8&-2&4&0&-2&-11&-4&0&0&0&0&0&0&9&0&0&0&0&0&0&0\\
1&2&14&16&1&-2&0&1&1&14&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&9&6&0&2&-2&0&0&6&0&0&0&-2&2&0&0&2&-4&2&-2&-2&2&0\\
0&0&-3&0&0&-2&2&0&0&-3&0&0&0&2&-2&0&0&-2&4&-2&2&2&-2&0\\
9&7&6&8&3&-2&0&2&11&-2&0&0&0&0&0&0&-9&0&0&0&0&0&0&0\\
0&0&3&3&0&1&2&0&0&3&0&0&0&2&-2&0&0&-2&4&-2&2&2&-2&0\\
-3&-3&6&6&0&-2&5&0&-3&6&0&0&0&2&-2&0&3&-2&4&-2&2&2&-2&0\\
-2&-2&8&8&1&0&0&2&-4&16&0&0&0&0&0&0&3&0&0&0&0&0&0&0\\
3&3&0&0&0&0&0&0&6&0&0&0&0&0&0&0&-3&0&0&0&0&0&0&0\\
0&0&3&3&0&0&0&0&0&6&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
-13&-11&10&8&-1&2&0&-1&-13&10&9&6&0&0&0&6&6&0&2&0&2&2&0&0\\
0&0&0&0&0&0&0&0&0&0&-3&0&0&0&0&-3&3&0&-2&0&-2&-2&0&0\\
0&0&0&0&0&8&-8&0&0&0&7&9&5&-8&8&7&-7&0&-6&0&2&2&0&0\\
0&0&0&0&0&0&0&0&0&0&3&3&0&3&0&3&-3&0&-2&0&-2&-2&0&0\\
-3&-3&3&3&0&2&-2&0&-3&3&1&3&2&-2&5&1&2&0&-2&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&3&3&0&0&0&6&-3&0&0&0&0&0&0&0\\
-13&-11&10&8&-1&2&0&-1&-13&10&0&0&0&0&0&0&15&0&0&0&0&0&0&0\\
0&0&0&0&0&-8&8&0&0&0&2&0&-2&8&-8&2&-2&-1&16&-4&4&4&-2&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&5&0&2&2&0&0\\
0&0&0&0&0&0&0&0&0&0&2&0&-2&0&0&2&-2&2&0&5&0&0&2&0\\
0&0&0&0&0&4&-4&0&0&0&-2&0&2&-4&4&-2&2&0&-4&0&1&0&0&0\\
0&0&0&0&0&-4&4&0&0&0&2&0&-2&4&-4&2&-2&0&12&0&4&5&0&0\\
0&0&0&0&0&16&-16&0&0&0&-2&0&2&-16&16&-2&2&6&-24&12&-4&-4&7&0\\
0&0&0&0&0&-4&4&0&0&0&0&0&0&4&-4&0&0&-2&8&-4&2&2&-2&1
\end{pmatrix}.$$

Acknowledgements

Liu was supported in part by NSF grant DMS-1937215. Treviño was supported by the NSF award DMS-2143133 Career. We are grateful to Michael Baake, Franz Gähler, and Lorenzo Sadun, for insightful conversations during the preparation of this manuscript, and for suggesting the quotient of the squiral substitution. Thanks to Lorenzo Sadun for pointing out a mistake in the first pair of examples, and to the anonymous referee for useful comments.

References

[1] **E Akkermans**, **Y Don**, **J Rosenberg**, **C L Schochet**, *Relating diffraction and spectral data of aperiodic tilings: towards a Bloch theorem*, J. Geom. Phys. 165 (2021) art. id. 104217 MR

[2] **J E Anderson**, **I F Putnam**, *Topological invariants for substitution tilings and their associated C^*-algebras*, Ergodic Theory Dynam. Systems 18:3 (1998) 509–537 MR

[3] **M F Atiyah**, **G B Segal**, *The index of elliptic operators, II*, Ann. of Math. (2) 87 (1968) 531–545 MR

[4] **M Baake**, **U Grimm**, *Squirals and beyond: substitution tilings with singular continuous spectrum*, Ergodic Theory Dynam. Systems 34:4 (2014) 1077–1102 MR

[5] **M Barge**, **B Diamond**, *Cohomology in one-dimensional substitution tiling spaces*, Proc. Amer. Math. Soc. 136:6 (2008) 2183–2191 MR

[6] **M Barge**, **B Diamond**, **J Hunton**, **L Sadun**, *Cohomology of substitution tiling spaces*, Ergodic Theory Dynam. Systems 30:6 (2010) 1607–1627 MR

[7] **J Bellissard**, *K-theory of C^*-algebras in solid state physics*, from "Statistical mechanics and field theory: mathematical aspects" (Groningen, 1985) (T C Dorlas, N M Hugenholtz, M Winnink, editors), Lecture Notes in Phys. 257, Springer (1986) 99–156 MR

[8] **J Bellissard**, **R Benedetti**, **J-M Gambaudo**, *Spaces of tilings, finite telescopic approximations and gap-labeling*, Comm. Math. Phys. 261:1 (2006) 1–41 MR

[9] **M T Benameur**, **V Mathai**, *Proof of the magnetic gap-labelling conjecture for principal solenoidal tori*, J. Funct. Anal. 278:3 (2020) art. id. 108323 MR

[10] **M-T Benameur**, **H Oyono-Oyono**, *Gap-labelling for quasi-crystals (proving a conjecture by J. Bellissard)*, from "Operator algebras and mathematical physics" (Constanţa, 2001) (J-M Combes, J Cuntz, G A Elliott, G Nenciu, H Siedentop, Ş Strătilă, editors), Theta, Bucharest (2003) 11–22 MR

[11] **B Blackadar**, *K-theory for operator algebras*, 2nd edition, Mathematical Sciences Research Institute Publications 5, Cambridge Univ. Press (1998) MR

[12] **G E Bredon**, *Equivariant cohomology theories*, Lecture Notes in Mathematics 34, Springer (1967) MR

[13] **F Gähler**, **J Hunton**, **G Maloney**, *A modified dual-tiling approach to the Anderson–Putnam method*, lecture slides (2013)

[14] **A Hatcher**, *Vector bundles and K–theory*, preprint (2017) Available at https://pi.math.cornell.edu/~hatcher/VBKT/VBpage.html

[15] **T Kaczynski**, **M Mrozek**, *The cubical cohomology ring: an algorithmic approach*, Found. Comput. Math. 13:5 (2013) 789–818 MR

[16] **J Kaminker**, **I Putnam**, *A proof of the gap labeling conjecture*, Michigan Math. J. 51:3 (2003) 537–546 MR

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

[18] **J Kellendonk**, **I F Putnam**, *The Ruelle–Sullivan map for actions of $\mathbb{R}^n$*, Math. Ann. 334:3 (2006) 693–711 MR

[19] **D Lind**, **B Marcus**, *An introduction to symbolic dynamics and coding*, 2nd edition, Cambridge Univ. Press (2021) MR

[20] **W Lück**, **B Oliver**, *Chern characters for the equivariant K-theory of proper G-CW-complexes*, from "Cohomological methods in homotopy theory" (Bellaterra, 1998) (J Aguadé, C Broto, C Casacuberta, editors), Progr. Math. 196, Birkhäuser, Basel (2001) 217–247 MR

[21] **I G Macdonald**, *Symmetric products of an algebraic curve*, Topology 1 (1962) 319–343 MR

[22] **S Mardešić**, **J Segal**, *Shapes of compacta and ANR-systems*, Fund. Math. 72:1 (1971) 41–59 MR

[23] **H Moustafa**, *PV cohomology of the pinwheel tilings, their integer group of coinvariants and gap-labeling*, Comm. Math. Phys. 298:2 (2010) 369–405 MR

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

[25] **J Rosenberg**, *The Künneth theorem in equivariant K-theory for actions of a cyclic group of order* 2, Algebr. Geom. Topol. 13:2 (2013) 1225–1241 MR

[26] **J Rosenberg**, *Complex K-theory of* 4*-complexes* (2024) arXiv 2409.16557 To appear in Indiana Univ. Math. J.

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

[28] **G Segal**, *Equivariant K-theory*, Inst. Hautes Études Sci. Publ. Math. :34 (1968) 129–151 MR

[29] **G Segal**, *The representation ring of a compact Lie group*, Inst. Hautes Études Sci. Publ. Math. :34 (1968) 113–128 MR

[30] **B Solomyak**, *Dynamics of self-similar tilings*, Ergodic Theory Dynam. Systems 17:3 (1997) 695–738 MR

[31] **B Solomyak**, **R Treviño**, *Spectral cocycle for substitution tilings*, Ergodic Theory Dynam. Systems 44:6 (2024) 1629–1672 MR

[32] **C Starling**, *K-theory of crossed products of tiling C^*-algebras by rotation groups*, Comm. Math. Phys. 334:1 (2015) 301–311 MR

[33] **M F Whittaker**, *C^*-algebras of tilings with infinite rotational symmetry*, J. Operator Theory 64:2 (2010) 299–319 MR

JIANLONG LIU jlliu@utexas.edu
Department of Mathematics, University of Texas at Austin, Austin, TX, United States

JONATHAN ROSENBERG jmr@umd.edu
Department of Mathematics, University of Maryland, College Park, MD, United States

RODRIGO TREVIÑO rodrigo@math.umd.edu
Department of Mathematics, University of Maryland, College Park, MD, United States

Received: October 11, 2024 Revised: March 26, 2025

msp

Algebraic & Geometric Topology **26**:3 (2026) 1195–1227
DOI: 10.2140/agt.2026.26.1195
Published: April 1, 2026

Relative bounded cohomology on groups with contracting elements

ZHENGUO HUANGFU AND RENXING WAN

Let G be a countable group acting properly on a metric space with contracting elements and $\{H_i : 1 \le i \le n\}$ be a finite collection of Morse subgroups in G. We prove that each H_i has infinite index in G if and only if the relative second bounded cohomology $H_b^2(G, \{H_i\}_{i=1}^n; \mathbb{R})$ is infinite-dimensional. In addition, we also prove that for any contracting element g, there exists $k > 0$ such that $H_b^2(G, \langle\langle g^k \rangle\rangle; \mathbb{R})$ is infinite-dimensional. Our results generalize a theorem of Pagliantini–Rolli for finite-rank free groups and yield new results on the (relative) second bounded cohomology of groups.

1 Introduction

1.1 (Relative) bounded cohomology

Bounded cohomology was introduced by Johnson and Trauber in the context of Banach algebra and developed into a comprehensive and rich theory by Gromov in his seminal paper "Volume and bounded cohomology" [28]. Since then, it has become a fundamental tool in several fields, most notably in the study of the geometry of manifolds. Many properties, in particular geometric ones, of a group can be characterized by its bounded cohomology. In particular, the notion of bounded cohomology allows us to retrieve certain negatively curved features of the group.

The theory of quasimorphisms has been extensively exploited to study the second bounded cohomology of a group. To be precise, a *quasimorphism* on a group G is a map $\phi : G \to \mathbb{R}$ such that

$$\sup_{g,h \in G} |\phi(gh) - \phi(g) - \phi(h)| < \infty.$$

That is to say, a quasimorphism is locally close to a genuine homomorphism from the group to $\mathbb{R}$. We denote by $\mathrm{QM}(G)$ the $\mathbb{R}$-vector space of quasimorphisms. It follows from the definitions that the coboundary of a quasimorphism is a bounded 2-cocycle and therefore there is a linear map

$$\mathrm{QM}(G) \to H_b^2(G; \mathbb{R}).$$

Moreover, the image of this map is the kernel of the comparison map $H_b^2(G; \mathbb{R}) \to H^2(G; \mathbb{R})$ induced by the inclusion of bounded cochains into ordinary cochains [12, Theorem 2.50].

The first example of this approach is using Brooks' counting quasimorphism [9] to show that the rank-2 free group F_2 has infinite-dimensional second bounded cohomology.

More generally, D. Epstein and K. Fujiwara [18] showed that a nonelementary hyperbolic group has infinite-dimensional second bounded cohomology. This was proved by using a modified version of Brooks

MSC2020: primary 20F65; secondary 57M07.

counting quasimorphism. Later, Fujiwara used this generalization to obtain the same conclusion regarding the dimension of the second bounded cohomology of a group acting properly on a hyperbolic space [24].

When dealing with manifolds with boundary we naturally consider the relative homology and cohomology. Similarly, in group theory, we often investigate the subgroups in order to explore properties of the ambient group. M. Gromov provided a definition of relative bounded cohomology between pairs of topological spaces and also pairs of groups in [28]. This leads to applications in geometry, topology and other fields.

In retrospect, absolute bounded cohomology of many classes of groups are already known to be infinite-dimensional, as demonstrated in

(1) nonelementary Gromov hyperbolic groups [18];

(2) groups with infinitely many ends [25];

(3) groups admitting a nonelementary proper discontinuous action on a Gromov hyperbolic space [24];

(4) groups admitting a nonelementary weakly proper discontinuous action on a Gromov hyperbolic space [5] (see also [30]);

(5) groups admitting a nonelementary weakly proper discontinuous action on a CAT(0) space which contains a rank-one isometry [6];

(6) acylindrically hyperbolic groups [33].

Gromov developed the theory of bounded cohomology in order to compute simplicial volume of manifolds. In particular, the understanding of simplicial volume of manifolds with nonempty boundary could be increased by studying bounded cohomology of pairs of space and pairs of groups, which is the so-called *relative bounded cohomology*. However, very few results on relative bounded cohomology are known. The first computation of the relative bounded cohomology of a specific class of groups was given by C. Pagliantini and P. Rolli [42]. Besides, it was proved independently by Bucher–Burger–Frigerio–Iozzi–Pagliantini–Pozzetti [11] and Kim–Kuessner [35] that the bounded cohomology of a CW-complex X relative to an amenable subcomplex Y is isometrically isomorphic to the absolute bounded cohomology of X; in particular, in this situation, $H_b^*(X, Y; \mathbb{R})$ is infinite-dimensional whenever X fits into the list given in the previous paragraph. Another relevant result was given by Franceschini [21], who proved that if (G, H) is a relatively hyperbolic pair, then the comparison map $H_b^k(G, H; V) \to H^k(G, H; V)$ is surjective for every $k \geq 2$ and any bounded G-module V.

1.2 Contracting element

The purpose of this paper is to generalize the existing results on the second bounded cohomology mentioned above to the relative version. Note that the examples of groups described above all satisfy the contracting properties outlined below.

Let X be a geodesic metric space and A be a closed subset of X. For a constant $C \geq 0$, we denote by $N_C(A)$ the open C-neighborhood of A in X. Let

$$\pi_A : X \to 2^A, \quad x \mapsto \pi_A(x) = \{y \in A : d(x, y) = d(x, A)\},$$

be the map given by the closest point projection. We say that A is *C-contracting* for $C \geq 0$ if $\operatorname{diam}(\pi_A(\gamma)) \leq C$ for any geodesic (segment) γ with $\gamma \cap N_C(A) = \varnothing$. In fact, this notion of contracting is equivalent to the usual one: $\operatorname{diam}(\pi_A(B)) \leq C'$ for any metric ball B disjoint from A; see [6, Corollary 3.4] for a proof. For an isometric group action of a group G on X, an element $g \in G$ is called *contracting* if some (or equivalently, any) orbit of $\langle g \rangle$ is a contracting quasigeodesic.

The prototype of a contracting element is a loxodromic isometry on a Gromov hyperbolic space, but many more examples are known to be contracting:

- rank-1 elements in CAT(0) groups acting on a CAT(0) space, see [2; 6];
- hyperbolic elements in groups with nontrivial Floyd boundary (e.g., relatively hyperbolic groups) acting on their Cayley graph with respect to a generating set, see [17; 40; 47];
- certain infinite-order elements in graphical small cancellation groups acting on their Cayley graph with respect to a generating set, see [1];
- pseudo-Anosov elements in mapping class groups acting on the Teichmüller space equipped with Teichmüller metric, or on the curve complex, see [20; 19; 44].

In this paper, a group is called *nonelementary* if it is not virtually cyclic. Given an isometric group action on a metric space, a subgroup is called *Morse* if some (or equivalently, any) orbit of this subgroup is weakly quasiconvex (see Definition 3.1). Now, we state our main result.

Theorem 1.1 *Let G be a nonelementary countable group acting properly on a geodesic metric space with contracting elements. Consider a finite collection of Morse subgroups $H_1, \dots, H_n$ with infinite index. Then there is an injective $\mathbb{R}$-linear map $\omega : \ell^1 \to H_b^2(G; \mathbb{R})$ such that each coclass in the image $\omega(\ell^1)$ has a representative vanishing on H_i for each i $(1 \leq i \leq n)$.*

Moreover, the dimension of $H_b^2(G, \{H_i\}_{i=1}^n; \mathbb{R})$ as a vector space over $\mathbb{R}$ has the cardinality of the continuum.

Here we denote by ℓ^1 the Banach space of summable sequences of real numbers with the norm $\|(x_i)\| = \sum_{i=1}^{\infty} |x_i|$.

Remark 1.2 (1) If one of the H_i's is of finite index in G, then Corollary 2.24 below implies that $H_b^2(G, \{H_i\}_{i=1}^n; \mathbb{R}) = 0$. Based on this fact, we only consider the second relative bounded cohomology of G with respect to subgroups with infinite index. Since any finitely generated subgroup of a free group is Morse, Theorem 1.1 generalizes the result of Pagliantini–Rolli [42] which states that for a free group F_n $(n \geq 2)$ and a finitely generated subgroup $H \leq F_n$, the subgroup H has infinite index in F_n if and only if the dimension of the second relative bounded cohomology $H_b^2(F_n, H; \mathbb{R})$ as a vector space over $\mathbb{R}$ is infinite.

(2) In [23; 33], the authors extend a nontrivial quasimorphism on a subgroup to the ambient group. Here we take the opposite approach, by constructing a quasimorphism on the ambient group whose restriction to the subgroup is trivial.

To prove Theorem 1.1, we need the following result.

Proposition 1.3 (Proposition 3.3) *Let G be a nonelementary countable group acting properly on a geodesic metric space with contracting elements. Consider a finite collection of Morse subgroups $H_1, \dots, H_n$ with infinite index. Then there is a quasitree on which G acts and each H_i acts elliptically on it for $1 \le i \le n$.*

An interesting corollary of Theorem 1.1 is as follows. This generalizes a result of Kotschick [36, Corollary 11]. See Definition 5.5 for the definition of bounded generation.

Corollary 1.4 (Corollary 5.6) *Under the assumption of Theorem 1.1, G is not boundedly generated by $\{H_i : 1 \le i \le n\}$.*

When a subgroup is a Morse subgroup of infinite index, its limit set is always a proper subset of the limit set of the ambient group, which provides evidence for the existence of enough relative quasimorphisms. (See [31; 49] for more details about convergence boundary.) Thus, we can ask the following question.

Question 1.5 *Let G be a nonelementary countable group acting properly on a geodesic metric space X with convergence boundary. Let $\{H_1, \dots, H_n\}$ be a finite collection of subgroups with proper limit sets. Then is the dimension of $H_b^2(G, \{H_i\}_{i=1}^n; \mathbb{R})$ as a vector space over $\mathbb{R}$ infinite?*

Another natural question is that when the subgroup is taken to be a normal subgroup of infinite index, does the conclusion of Theorem 1.1 still hold? If the normal subgroup has an amenable quotient, then Proposition 2.25 gives a negative answer. If the normal subgroup is normally generated by a higher power of a contracting element, then we have the following result.

Proposition 1.6 (Proposition 6.1) *Let G be a nonelementary countable group acting properly on a geodesic metric space with contracting elements. Then for any contracting element $g \in G$, there exists $k = k(g) > 0$ such that the dimension of $H_b^2(G, \langle\langle g^k \rangle\rangle; \mathbb{R})$ as a vector space over $\mathbb{R}$ has the cardinality of the continuum.*

Remark 1.7 Proposition 1.6 is already known for nonelementary hyperbolic groups. In [15], Delzant showed that for any hyperbolic element g in a nonelementary hyperbolic group G, there exists $k \in \mathbb{N}$ such that $G/\langle\langle g^k \rangle\rangle$ is still hyperbolic. Together with [18, Theorem 1.1] and Proposition 2.22 below, one gets the result.

In general, we raise the following question.

Question 1.8 *Let G be a nonelementary countable group acting properly on a geodesic metric space X with contracting elements. Let H be a normal subgroup of G with nonamenable quotient. Is the dimension of $H_b^2(G, H; \mathbb{R})$ as a vector space over $\mathbb{R}$ infinite?*

1.3 Sketch of proof

There are two main ingredients in the proof of Theorem 1.1. The first is Proposition 1.3, namely the construction of an appropriate projection complex on which each H_i acts elliptically. This notion was introduced by Bestvina–Bromberg–Fujiwara [3]. In Section 3, we are going to make use of the Morse property of $\{H_i\}$ (see Lemma 3.9) to show that there exists a contracting element g such that every $h \in H_i$

has a uniformly bounded projection to the geodesic segment $[o, go]$. To achieve this goal, we use some techniques developed by Han–Yang–Zou [31]. Then we construct the projection complex $\mathcal{P}_K(\mathcal{F})$ whose vertices are the G-translates of the axis $\mathrm{Ax}(g)$ of g and show that each H_i acts elliptically on $\mathcal{P}_K(\mathcal{F})$ (see Lemma 3.10). The projection complex is shown to be a quasitree on which G acts acylindrically (see Section 5).

The second ingredient is Proposition 4.1, which states that if G acts WPD (a weaker notion than acylindrical action) on a δ-hyperbolic space X and each H_i acts elliptically on X, then the dimension of $H_b^2(G, \{H_i\}_{i=1}^n; \mathbb{R})$ as a vector space over $\mathbb{R}$ has the cardinality of the continuum. To prove this proposition, we need a result of Bestvina–Fujiwara [5, Proposition 2] which constructs an infinite collection of words $\{f_i\}$ in a rank-two free subgroup of G. As in [24], we can produce a corresponding collection of quasimorphisms $\{h_i\}$ on G, which satisfy some special properties stated in Proposition 4.13. Then Proposition 4.1 follows from Proposition 4.13. The bridge connecting two ingredients is a result of Bestvina–Bromberg–Fujiwara–Sisto [4, Theorem 5.6]. As a result, we get Theorem 1.1.

As for Proposition 1.6, where H is assumed to be a normal subgroup generated by a higher power of a contracting element, we first construct a quasitree of spaces $\mathcal{C}(\mathcal{F})$ for every contracting element which is a blow-up of the projection complex. The space $\mathcal{C}(\mathcal{F})$ turns out to be a quasitree on which G acts acylindrically (see Lemma 6.4). Later, we construct a hyperbolic cone-off over a scaled $\mathcal{C}(\mathcal{F})$ along $\mathcal{F}$ and obtain a very rotating family (see Lemma 6.5). With the help of the theory of rotating families developed by Dahmani–Guirardel–Osin [14], we can reduce the proof of Proposition 1.6 to the proof of Proposition 4.14, which is completed at the end of Section 4.

Structure of the paper The paper is organized as follows. Section 2 is devoted to recalling some preliminary material about Gromov-hyperbolic spaces, contracting subsets, projection complexes, (relative) bounded cohomology, and quasimorphisms. In Section 3, we prove Proposition 1.3. Specifically, we make use of the recent work of Han–Yang–Zou [31] to construct a projection complex and show that each H_i acts elliptically on it. In Section 4, we review the previous work on Epstein–Fujiwara quasimorphisms in [24] and prove Proposition 4.1. Then we complete the proof of Theorem 1.1 in Section 5. In Section 6, we first recall some facts about hyperbolic cone-offs and rotation families, and then we prove Proposition 1.6.

2 Preliminaries

We first introduce some fundamental notation and definitions that will be used throughout this paper.

2.1 Gromov-hyperbolic spaces

We only introduce some necessary knowledge about Gromov-hyperbolic spaces here. For a more detailed introduction, we refer the readers to [8, Part III. H; 16, Chapter 11]. Let (X, d) be a geodesic metric space. For $S \subset X$ and $r > 0$, we denote by $N_r(S)$ the open r-neighborhood of S. For two subsets $S, T \subset X$, we denoted by $d_H(S, T)$ the *Hausdorff distance* between A and B, which is defined by $d_H(S, T) := \inf\{r > 0 : S \subset N_r(T), T \subset N_r(S)\}$.

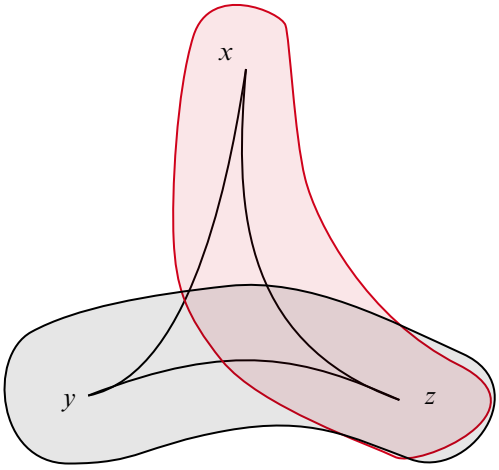


Figure 1: $[x, y] \subset N_\delta([x, z] \cup [y, z])$.

For any two points $x, y \in X$, denote by $[x, y]$ a choice of a geodesic segment between x and y. A *geodesic triangle* in X consists of three points $x, y, z \in X$ and three geodesic segments $[x, y]$, $[y, z]$, and $[z, x]$.

A geodesic metric space (X, d) is called (*Gromov*) *δ-hyperbolic* for a constant $\delta \geq 0$ if every geodesic triangle in X is *δ-thin*: each of its sides is contained in the δ-neighborhood of the union of the other two sides. See Figure 1 for an illustration.

A finitely generated group is called *Gromov-hyperbolic* if its Cayley graph with respect to some finite generating set is a δ-hyperbolic metric space for some $\delta \geq 0$.

Let (X_1, d_1) and (X_2, d_2) be two metric spaces. A (not necessarily continuous) map $f : X_1 \to X_2$ is called a *(λ, ϵ)-quasi-isometric embedding* if there exist constants $\lambda \geq 1$ and $\epsilon \geq 0$ such that for all $x, y \in X_1$ we have

$$\frac{1}{\lambda} d_1(x, y) - \epsilon \leq d_2(f(x), f(y)) \leq \lambda d_1(x, y) + \epsilon.$$

If, in addition, every point of X_2 lies in the ϵ-neighborhood of the image of f, then f is called a *(λ, ϵ)-quasi-isometry*. When such a map exists, the two spaces X_1 and X_2 are said to be *quasi-isometric*.

A *(λ, ϵ)-quasigeodesic* in a metric space X is the image of a (λ, ϵ)-quasi-isometric embedding $c : I \to X$, where I is an interval (possibly bounded or unbounded). For simplicity, a (λ, λ)-quasigeodesic will be referred to as a *λ-quasigeodesic*.

The following result is the well-known Morse lemma in geometric group theory. We refer the readers to [8, Chapter III.H, Theorem 1.7; 16, Theorem 11.72] for a proof.

Lemma 2.1 (Morse lemma) *For all $\delta \geq 0$, $\lambda \geq 1$, $\epsilon \geq 0$, there exists a constant $L = L(\lambda, \epsilon, \delta)$ such that any two (λ, ϵ)-quasigeodesics in a δ-hyperbolic space with the same endpoints are contained in an L-neighborhood of each other.*

Isometries on Gromov-hyperbolic spaces Let (X, d) be a Gromov-hyperbolic space. Two geodesic rays in X are said to be *asymptotic* if the Hausdorff distance between them is finite. Being asymptotic is an equivalence relation on geodesic rays. The *Gromov boundary* of X, denoted by ∂X, is defined to be the set of equivalence classes of geodesic rays in X. We refer the readers to [8; 16] for a detailed discussion about Gromov boundary. By Gromov [29], the isometries of a hyperbolic space X can be subdivided into three classes. A nontrivial element $g \in \mathrm{Isom}(X)$ is called *elliptic* if some $\langle g\rangle$-orbit is bounded. Otherwise, it is called *loxodromic* (resp. *parabolic*) if it has exactly two fixed points (resp. one fixed point) in the Gromov boundary of X. If g is a loxodromic element, any $\langle g\rangle$-invariant quasigeodesic between the two fixed points will be referred to as a *quasiaxis* for g, denoted by L_g.

For an isometry g on a hyperbolic space (X, d), we define the *stable translation length* $\|g\|$ of g as

$$\|g\| := \lim_{n\to\infty} \frac{d(x, g^n x)}{n}$$

for some (or any) $x \in X$. A well-known fact is that an isometry g is loxodromic if and only if $\|g\| > 0$ [13, Chapter 10, Proposition 6.3].

Lemma 2.2 *Let g be an isometry on a metric space (X, d). Then for any $n > 0$ and $x \in X$, one has $d(x, g^n x) \geq n\|g\|$.*

Proof For any $m, n > 0$, it follows from $d(x, g^{mn}x) \leq m \cdot d(x, g^n x)$ that

$$\frac{d(x, g^{mn}x)}{mn} \leq \frac{d(x, g^n x)}{n}.$$

By letting $m \to \infty$, one gets that

$$\|g\| \leq \frac{d(x, g^n x)}{n}. \qquad \square$$

2.2 Contracting subsets

Let (X, d) be a geodesic metric space. For any two $A, B \subset X$, define $d(A, B) := \inf_{x\in A, y\in B} d(x, y)$. For a given closed subset $A \subset X$, define the *closest point projection* $\pi_A : X \to 2^A$ as follows:

- For any $x \in X$, $\pi_A(x) := \{y \in A : d(x, y) = d(x, A)\}$.
- For any subset $B \subset X$, $\pi_A(B) := \bigcup_{x\in B} \pi_A(x)$.

We use the notation $\mathrm{diam}(A)$ to denote the diameter of a subset A in X.

Definition 2.3 (contracting subset) Let (X, d) be a geodesic metric space. A subset $Y \subseteq X$ is called C-*contracting* for $C \geq 0$ if for any geodesic (segment) α in X with $d(\alpha, Y) \geq C$, we have $\mathrm{diam}(\pi_Y(\alpha)) \leq C$. The subset Y is called a *contracting subset* if there exists $C \geq 0$ such that Y is C-contracting, and C is called a *contraction constant* of Y. See Figure 2 for an illustration.

Let G be a finitely generated group. A subgroup $H \leq G$ is called *contracting* if it is a contracting subset in the Cayley graph $\mathcal{G}(G, S)$ of G with respect to some finite generating set S.

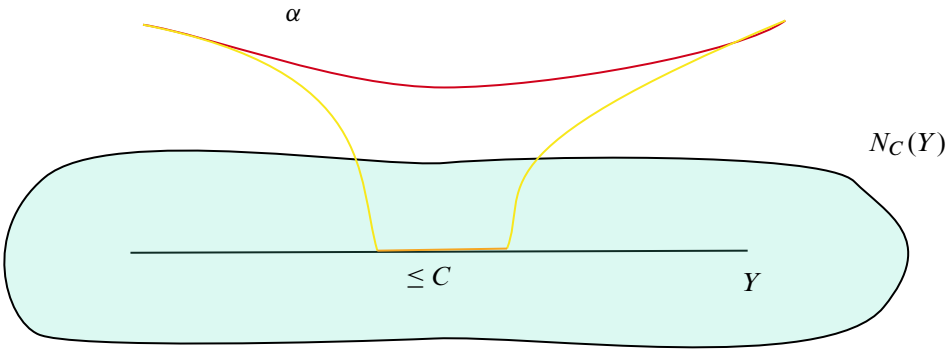


Figure 2: Y is C-contracting.

Example 2.4 The following are well-known examples of contracting subsets and contracting subgroups:

(1) bounded sets in a metric space;

(2) quasigeodesics and quasiconvex subsets in Gromov-hyperbolic spaces [27];

(3) fully quasiconvex subgroups, and maximal parabolic subgroups in particular, in relatively hyperbolic groups [26, Proposition 8.2.4];

(4) the subgroup generated by a hyperbolic element in groups whose Floyd boundary is nontrivial [47, Section 7];

(5) contracting segments and axes of rank-1 elements in CAT(0)-spaces in the sense of Bestvina and Fujiwara [6, Corollary 3.4];

(6) the axis of any pseudo-Anosov element in the Teichmüller space equipped with Teichmüller distance by a theorem of Minsky [37].

It has been proven in [6, Corollary 3.4] that the definition of a contracting subset is equivalent to the following one considered by Minsky [37]. A subset $Y \subseteq X$ is contracting if and only if there exists $C' \geq 0$ such that any open metric ball B with $B \cap Y = \varnothing$ satisfies $\operatorname{diam}(\pi_Y(B)) \leq C'$.

Despite this equivalence, we will always rely on Definition 2.3 for a contracting subset.

Definition 2.5 (contracting system) Let (X, d) be a geodesic metric space. A set $\mathbb{X} = \{X_i : X_i \subset X\}_{i\in\mathbb{N}}$ is called a *contracting system* with a contraction constant C if each X_i is a C-contracting subset in X.

Definition 2.6 (bounded intersection) Two subsets $Y, Z \subseteq X$ have *$\mathcal{R}$-bounded intersection* for a function $\mathcal{R} : [0, +\infty) \to [0, +\infty)$ if $\operatorname{diam}(N_r(Y) \cap N_r(Z)) \leq \mathcal{R}(r)$, for all $r \geq 0$. A contracting system $\mathbb{X}$ has *$\mathcal{R}$-bounded intersection* if any two elements in $\mathbb{X}$ have *$\mathcal{R}$-bounded intersection*.

Admissible paths For a finite rectifiable path p in a metric space X, we denote by $|p|$ the length of p and by p_-, p_+ the initial and terminal points of p respectively.

Definition 2.7 (admissible path) Let (X, d) be a geodesic metric space and $\mathbb{X}$ be a contracting system in X. Let $D, \tau \geq 0$ and $\mathcal{R} : [0, +\infty) \to [0, +\infty)$ be a function, which is called the *bounded intersection gauge*.

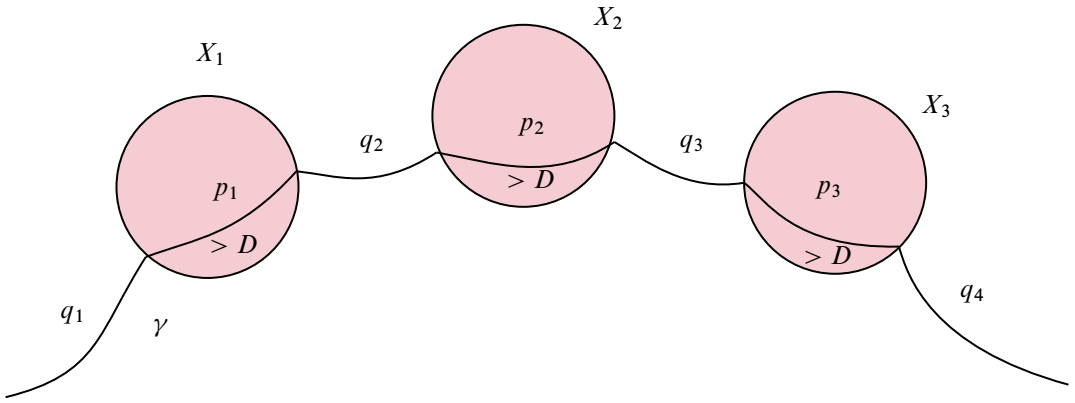


Figure 3: $\gamma = q_1 p_1 q_2 p_2 q_3 p_3 q_4$ is a (D, τ)-admissible path.

A path γ is called a (D, τ)-*admissible path* with respect to $\mathbb{X}$ if the path γ consists of a (finite, infinite, or bi-infinite) concatenation of consecutive geodesic segments $\gamma = \cdots q_i p_i q_{i+1} p_{i+1} \cdots$, satisfying the following "long local" and "bounded projection" properties:

(LL1) For each p_i there exists $X_i \in \mathbb{X}$ such that the two endpoints of p_i lie in X_i, and $|p_i| > D$ unless p_i is the first or last geodesic segment in γ.

(BP) For each X_i we have $\operatorname{diam}(\pi_{X_i}\{(p_i)_+, (p_{i+1})_-\}) \leq \tau$, and $\operatorname{diam}(\pi_{X_i}\{(p_{i-1})_+, (p_i)_-\}) \leq \tau$. Here $(p_{i+1})_- = \gamma_+$ if p_{i+1} does not exist, and $(p_{i-1})_+ = \gamma_-$ if p_{i-1} does not exist.

(LL2) Either $X_i \neq X_{i+1}$ and X_i and X_{i+1} have $\mathcal{R}$-bounded intersection, or $|q_i| > D$.

Remark 2.8 See Figure 3 for an illustration of an admissible path. The collection of $X_i \in \mathbb{X}$ indexed as in (LL1), denoted by $\mathbb{X}(\gamma)$, will be referred to as *contracting subsets* for γ. The union of all $X_i \in \mathbb{X}(\gamma)$ is called the *saturation* of γ.

In the following definitions, a sequence of points x_i in a path α is called *linearly ordered* if x_{i+1} lies in the subpath of α from x_i to α_+ for each i.

Definition 2.9 (fellow travel) Let $\gamma = p_0 q_1 p_1 \cdots q_n p_n$ be a (D, τ)-admissible path, and α be a path such that $\alpha_- = \gamma_-$, $\alpha_+ = \gamma_+$. Given $\epsilon > 0$, the path α *ϵ-fellow travels* γ if there exists a sequence of linearly ordered points $z_0, w_0, z_1, w_1, \ldots, z_n, w_n$ on α such that $d(z_i, (p_i)_-) \leq \epsilon$, $d(w_i, (p_i)_+) \leq \epsilon$.

See Figure 4 for an illustration of ϵ-fellow travel property.

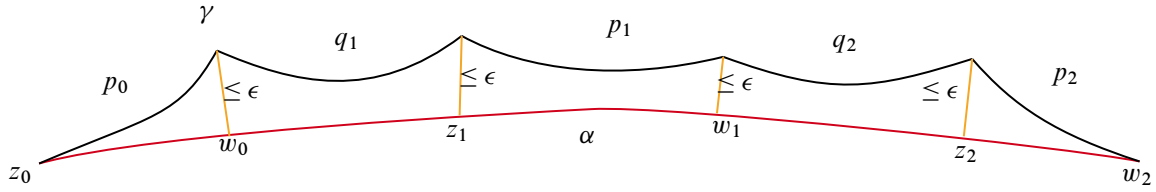


Figure 4: $\gamma = p_0 q_1 p_1 q_2 p_2$ is a (D, τ)-admissible path and α ϵ-fellow travels γ.

Proposition 2.10 [48, Proposition 2.7] *For any $\tau > 0$, and any function $\mathcal{R} : [0, +\infty) \to [0, +\infty)$, there exist $D, \lambda, \epsilon > 0$ depending on τ, $\mathcal{R}$ such that any (D, τ)-admissible path is a λ-quasigeodesic which is ϵ-fellow traveled by any geodesic with the same endpoints.*

Group actions with contracting elements Let G be a group acting isometrically on a geodesic metric space (X, d) with a base point $o \in X$.

Definition 2.11 (contracting element) An element $h \in G$ is called a *contracting element* if $\langle h \rangle \cdot o$ is a contracting subset in X and the map $\mathbb{Z} \to X$, $n \mapsto h^n o$, is a quasi-isometric embedding.

A group action $G \curvearrowright X$ is called *(metrically) proper* if for any $D > 0$, the set $\{g \in G : d(o, go) \leq D\}$ is finite. From now on, we assume that G acts properly on (X, d) with a contracting element.

Definition 2.12 (weakly independent) Suppose $g, h \in G$ are two contracting elements. They are called *weakly independent* if $\langle g \rangle \cdot o$ and $\langle h \rangle \cdot o$ have $\mathcal{R}$-bounded intersection for some $\mathcal{R} : [0, +\infty) \to [0, +\infty)$.

By [48, Lemma 2.11], each contracting element g is contained in a maximal elementary subgroup $E(g)$ defined as

$$E(g) = \{h \in G : \exists r > 0, h\langle g \rangle o \subset N_r(\langle g \rangle o) \text{ and } \langle g \rangle o \subset N_r(h\langle g \rangle o)\},$$

and the index $[E(g) : \langle g \rangle]$ is finite. Roughly speaking, the subgroup $E(g)$ is the set of elements that do not move the orbit of g too much. Hence, if G is not virtually cyclic, then there are at least two weakly independent contracting elements in G. For example, let $g \in G$ be a contracting element by assumption. Since G is not virtually cyclic, $E(g)$ is a proper subgroup of G. By selecting an element $f \in G \setminus E(g)$, one has that g and fgf^{-1} are weakly independent [46, Lemma 7.13]. Actually, by [46, Lemma 2.30], there exist infinitely many pairwise weakly independent elements in G.

For a contracting element g, the subset defined by $\mathrm{Ax}(g) := E(g)o$ is called an *extended-defined quasigeodesic*, which is also called the *axis* of g (depending on o). Compared with the quasiaxis L_g of a loxodromic element g which can be an arbitrary $\langle g \rangle$-invariant quasigeodesic, we use the term "extended-defined" here to emphasize their difference.

The following result, proved in [48, Proposition 2.9 and Lemma 2.14], gives a procedure to construct infinitely many contracting elements.

Lemma 2.13 (extension lemma) *Suppose that a nonelementary group G acts properly on (X, d) with a contracting element. Then there exist a set $F \subset G$ of three contracting elements and $D, \tau, \lambda > 0$ with the following property. For any $g \in G$, there exists $f \in F$ such that the bi-infinite concatenated path $\gamma = \bigcup_{n \in \mathbb{Z}} (gf)^n [o, gfo]$ is a (D, τ)-admissible path with contracting subsets $\{(gf)^n g \,\mathrm{Ax}(f) : n \in \mathbb{Z}\}$. In particular, γ is a C-contracting λ-quasigeodesic, where C depends on $d(o, go)$. Hence gf is a contracting element.*

2.3 Projection complexes

In the assumption of our main result, i.e., Theorem 1.1, the geodesic metric space X on which the group G acts with contracting elements is not hyperbolic. To apply the method proposed by Bestvina–Fujiwara [5],

we need to construct a suitable hyperbolic space on which G acts. This construction makes use of the projection complex techniques developed by Bestvina–Bromberg–Fujiwara [3], which will be introduced in this subsection.

Definition 2.14 (projection axioms) Let $\mathcal{F}$ be a collection of metric spaces equipped with (set-valued) projection maps

$$\{\pi_U : \mathcal{F} \setminus \{U\} \to 2^U\}_{U \in \mathcal{F}}.$$

Define $d_U(V, W) := \operatorname{diam}(\pi_U(V) \cup \pi_U(W))$ for $V \neq U \neq W \in \mathcal{F}$. The pair $(\mathcal{F}, \{U\}_{U \in \mathcal{F}})$ satisfies projection axioms for a constant $\kappa > 0$ if

(1) $\operatorname{diam}(\pi_U(V)) \leq \kappa$ when $U \neq V$;

(2) if U, V, W are distinct and $d_V(U, W) > \kappa$ then $d_U(V, W) \leq \kappa$;

(3) the set $\{U \in \mathcal{F} : d_U(V, W) > \kappa\}$ is finite for $V \neq W$.

We caution the readers that the projection maps in the above definition are just abstract maps, which may not be the closest point projections defined previously. By definition, the triangle inequality

$$d_Y(V, W) \leq d_Y(V, U) + d_Y(U, W) \tag{2-1}$$

holds for any $U, V, Y, W \in \mathcal{F}$. It is well known that the projection axioms hold for a contracting system with bounded intersection (see [47, Appendix]). From now on, we assume that G acts properly on X with contracting elements. Fix a base point $o \in X$.

Lemma 2.15 [49, Lemma 2.18] *Let $g \in G$ be a contracting element. Then the set $\mathcal{F} = \{f \operatorname{Ax}(g) : f \in G\}$ with closest point projections $\pi_U(V)$ satisfies the projection axioms with a constant $\kappa = \kappa(\mathcal{F}) > 0$.*

In [3, Definition 3.1], a modified version of d_U is introduced such that it is symmetric and agrees with the original d_U up to an additive amount 2κ. Thus, the axioms (1)–(3) still hold for 3κ, and the triangle inequality in (2-1) holds up to a uniform error. In what follows, we actually need to work with this modified d_U to define the projection complex, but for the sake of simplicity, we stick to the above definition of d_U.

We consider the interval-like set, for $K > 0$ and $V, W \in \mathcal{F}$,

$$\mathcal{F}_K(V, W) := \{U \in \mathcal{F} : d_U(V, W) > K\}.$$

Define $\mathcal{F}_K[V, W] := \mathcal{F}_K(V, W) \cup \{V, W\}$. It possesses a total order described in the next lemma. Let $\mathcal{F}$ and κ be given by Lemma 2.15.

Lemma 2.16 [3, Theorem 3.3.G] *There exist constants $D = D(\kappa), K = K(\kappa) > 0$ such that the set $\mathcal{F}_K[V, W]$ admits a total order "$<$" with least element V and greatest element W, such that given $U_0, U_1, U_2 \in \mathcal{F}_K[V, W]$, if $U_0 < U_1 < U_2$, then*

$$d_{U_1}(V, W) - D \leq d_{U_1}(U_0, U_2) \leq d_{U_1}(V, W) \quad \text{and} \quad d_{U_0}(U_1, U_2) \leq D \text{ and } d_{U_2}(U_0, U_1) \leq D.$$

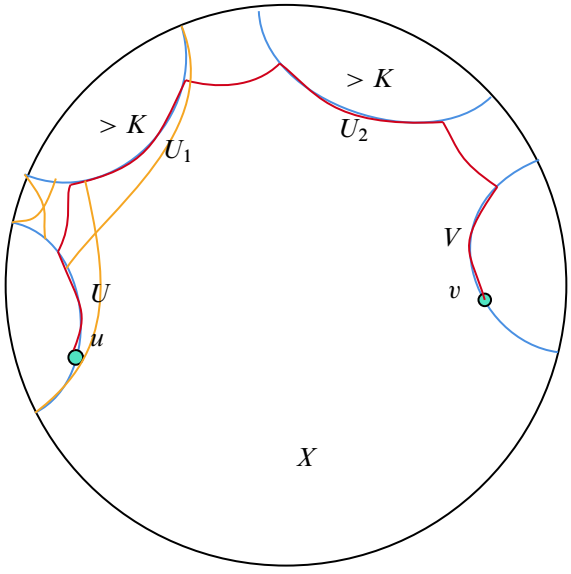


Figure 5: Each blue line is a translate of $\mathrm{Ax}(g)$; $\mathcal{F}_K[U,V]\cup\{U,V\}=\{U<U_1<U_2<V\}$ is a standard path in $\mathcal{P}_K(\mathcal{F})$; the yellow line represents the closest point projection between U and U_1; the red line is a lifted standard path from $u\in U$ to $v\in V$ in X.

We now give the definition of a projection complex.

Definition 2.17 The projection complex $\mathcal{P}_K(\mathcal{F})$ for K satisfying Lemma 2.16 is a graph with the vertex set consisting of the elements in $\mathcal{F}$. Two vertices U and V are connected if $\mathcal{F}_K(U,V)=\varnothing$. We equip $\mathcal{P}_K(\mathcal{F})$ with a length metric $d_{\mathcal{P}}$ induced by assigning unit length to each edge.

The projection complex $\mathcal{P}_K(\mathcal{F})$ is connected, since by [3, Proposition 3.7], the interval set $\mathcal{F}_K[U,V]$ gives a connected path between U and V in $\mathcal{P}_K(\mathcal{F})$: the consecutive elements directed by the total order are adjacent in $\mathcal{P}_K(\mathcal{F})$. The path $\mathcal{F}_K[U,V]\cup\{U,V\}=\{U<U_1<\cdots<U_k<V\}$ is called the *standard path* from U to V. See Figure 5 for an illustration of a standard path. By [4, Corollary 3.7], standard paths are (2, 1)-quasigeodesics in $\mathcal{P}_K(\mathcal{F})$. A *quasitree* is a geodesic metric space quasi-isometric to a tree. The structural result about the projection complex is the following.

Lemma 2.18 [3, Corollary 3.25] *For $K\gg 0$ as in Lemma 2.16, the projection complex $\mathcal{P}_K(\mathcal{F})$ is a quasitree, on which G acts nonelementarily and coboundedly.*

For any two points $u\in U$ and $v\in V$, we often need to *lift* a standard path $\mathcal{F}_K[U,V]\cup\{U,V\}=\{U<U_1<\cdots<U_k<V\}$ in $\mathcal{P}_K(\mathcal{F})$ to a path from u to v in X. The *lifted standard path* in X is a piecewise geodesic path (admissible path) from u to v as concatenation of the normal paths between two consecutive vertices which are G-translates of $\mathrm{Ax}(g)$ in X and geodesics contained in vertices. See Figure 5 for an illustration of a lifted standard path. This is explained by the following lemma proved in [32, Lemma 4.5].

Lemma 2.19 *For any $K > 0$, there exist a constant $D = D(K, \kappa) \geq 0$ with $D \to \infty$ as $K \to \infty$ and a uniform constant $B = B(\kappa) > 0$ with the following property. For any two points $u \in U$, $v \in V$ there exists an (D, B)-admissible path γ in X from u to v with saturation $\mathcal{F}_K[U, V]$.*

In practice, we always assume that K is sufficiently large such that by taking $\tau = B$ and $\mathcal{R}$ as the bounded intersection gauge of $\mathcal{F}$, the constant D in the above lemma satisfies Proposition 2.10, and then the path γ shall be a quasigeodesic.

2.4 (Relative) bounded cohomology

This subsection is devoted to the basic definitions and results on the bounded cohomology of a group and the relative bounded cohomology of a pair of groups. We refer the reader to [10] for full details on ordinary group cohomology theory, to [22; 28; 34] on bounded cohomology of groups, and to [28; 43] for the relative case.

Let G be a group. For a coefficient ring R, the *bar complex* $C_*(G; R)$ is the complex generated in dimension n by n-tuples $(g_1, \dots, g_n)$ with $g_i \in G$ and with boundary map ∂ defined by the formula

$$\partial(g_1, \dots, g_n) = (g_2, \dots, g_n) + \sum_{i=1}^{n-1} (-1)^i (g_1, \dots, g_i g_{i+1}, \dots, g_n) + (-1)^n (g_1, \dots, g_{n-1}).$$

We let $C^*(G; R)$ denote the dual cochain complex $\mathrm{Hom}(C_*(G), R)$, and let d denote the adjoint of ∂. The homology groups of $C^*(G; R)$ are called the *group cohomology of G with coefficients in R*, and are denoted by $H^*(G; R)$.

In this paper, we take $R = \mathbb{R}$. A cochain $\alpha \in C^n(G; \mathbb{R})$ is *bounded* if

$$\sup |\alpha(g_1, \dots, g_n)| < \infty,$$

where the supremum is taken over all n-tuples $(g_1, \dots, g_n)$ with $g_i \in G$. The set of all bounded cochains forms a subcomplex $C_b^*(G; \mathbb{R})$ of $C^*(G; \mathbb{R})$, and its homology is the so-called *bounded cohomology* $H_b^*(G; \mathbb{R})$.

Amenable groups Recall that a *mean* on G is a linear functional on $C_b^1(G; \mathbb{R})$ which maps the constant function $\phi(g) = 1$ to 1, and maps nonnegative functions to nonnegative numbers.

Definition 2.20 A group G is *amenable* if there is a G-invariant mean $\pi : C_b^1(G; \mathbb{R}) \to \mathbb{R}$ where G acts on $C_b^1(G; \mathbb{R})$ by

$$g \cdot \phi(h) = \phi(g^{-1} h)$$

for all $g, h \in G$ and $\phi \in C_b^1(G; \mathbb{R})$.

Examples of amenable groups are finite groups, solvable groups, and Grigorchuk's groups of intermediate growth.

We list three important facts here for future use. Some of them are mentioned in the introduction. See [7; 12, §2.4.2] for details.

Lemma 2.21 (1) $H_b^1(G;\mathbb{R}) = 0$ *for any group* G.

(2) $H_b^*(G;\mathbb{R})$ *vanishes identically when* G *is an amenable group.*

(3) *Let* $H \to G \to K \to 1$ *be a* (*right*) *exact sequence of groups. Then the induced sequence on second bounded cohomology*

$$0 \to H_b^2(K;\mathbb{R}) \to H_b^2(G;\mathbb{R}) \to H_b^2(H;\mathbb{R})$$

is (*left*) *exact.*

Unless otherwise stated, *a pair of groups* (G, H) in this paper means that G is a group and H is a subgroup of G. Now we come to the definition of relative bounded cohomology of the pair (G, H). The kernel of the obvious restriction map $C_b^*(G,\mathbb{R}) \to C_b^*(H,\mathbb{R})$ is denoted by $C_b^*(G, H;\mathbb{R})$, and we have the short exact sequence of complexes

$$0 \to C_b^*(G, H;\mathbb{R}) \to C_b^*(G;\mathbb{R}) \to C_b^*(H;\mathbb{R}) \to 0,$$

which induces a long exact sequence in cohomology

$$\cdots \to H_b^{n-1}(H;\mathbb{R}) \to H^n(C_b^*(G, H;\mathbb{R})) \to H_b^n(G;\mathbb{R}) \to H_b^n(H;\mathbb{R}) \to \cdots. \tag{2-2}$$

The module $H^n(C_b^*(G, H;\mathbb{R}))$ is the n-th bounded cohomology of the pair (G, H). We denote it as $H_b^n(G, H;\mathbb{R})$. Let $\{H_i\}_{i=1}^m$ be a finite collection of subgroups in G. We denote by $C_b^*(G, \{H_i\}_{i=1}^m;\mathbb{R})$ the kernel of the multiple restriction map $C_b^*(G,\mathbb{R}) \to \bigoplus_{i=1}^m C_b^*(H_i,\mathbb{R})$. In the same way as above, we define $H_b^n(G, \{H_i\}_{i=1}^m;\mathbb{R})$ as the n-th bounded cohomology $H^n(C_b^*(G, \{H_i\}_{i=1}^m;\mathbb{R}))$. There are no substantial differences from the relative cohomology in ordinary cohomology group theory.

Proposition 2.22 *Let* (G, H) *be a pair of groups with* $H \lhd G$. *Then* $H_b^2(G/H;\mathbb{R}) \cong H_b^2(G, H;\mathbb{R})$.

Proof Recall that Lemma 2.21(3) gives us a left-exact sequence

$$0 \to H_b^2(G/H;\mathbb{R}) \to H_b^2(G;\mathbb{R}) \to H_b^2(H;\mathbb{R}).$$

From the definition of relative bounded cohomology (2-2), we also have the exact sequence

$$H_b^1(H;\mathbb{R}) \to H_b^2(G, H;\mathbb{R}) \to H_b^2(G;\mathbb{R}) \to H_b^2(H;\mathbb{R}).$$

As any group has a trivial first bounded cohomology, we can conclude that

$$H_b^2(G/H;\mathbb{R}) \cong \ker(H_b^2(G;\mathbb{R}) \to H_b^2(H;\mathbb{R})) \cong H_b^2(G, H;\mathbb{R}). \tag{2-3}$$ □

We remark that the second isomorphism in (2-3) always holds as long as (G, H) is a pair of groups.

Finite-index subgroups Let (G, H) be a pair of groups. The inclusion map from H to G induces the *restriction map* which is denoted by $\mathrm{res} : H^n(G,\mathbb{R}) \to H^n(H,\mathbb{R})$. A subgroup H of a topological group G is called *cocompact* (or uniform) if the quotient space $G/\overline{H}$ is compact, where $\overline{H}$ denotes the closure of H in G. For cocompact subgroups such that the quotient admits a finite invariant measure, the restriction map from the cohomology group of the ambient group to the cohomology group of the subgroup is injective. This notably encompasses the case of uniform lattices (of Lie groups) and finite

index subgroups (of any group). The standard argument uses the existence of a *transfer map* which is the left inverse to the restriction map (or up to a constant multiple depending on different definitions of the transfer map). The transfer map is obtained by integration (or finite sum for finite index subgroups), so it is crucial for the subgroup to be cocompact. See [38, §8.6] for a detailed discussion of the transfer map.

Lemma 2.23 [38, Propositon 8.6.2] *Let (G, H) be a pair of groups. If H has finite index in G, then the natural restriction map $H_b^2(G;\mathbb{R}) \to H_b^2(H;\mathbb{R})$ is isometrically injective.*

Combining with the remark following Proposition 2.22, we have immediately that:

Corollary 2.24 *Let (G, H) be a pair of groups. If H has finite index in G, then $H_b^2(G, H;\mathbb{R}) = 0$.*

The corollary above gives us a better understanding of the relations between the index of a subgroup and the relative bounded cohomology group. In fact, Corollary 2.24 is the main motivation for the authors of [42] and us to consider the relative bounded cohomology of infinite index subgroups, aiming to find some nontrivial results. By combining Lemma 2.21(2) and Proposition 2.22, one gets another simple example as follows.

Proposition 2.25 *Let (G, H) be a pair of groups. If H is a normal subgroup with amenable quotient, then $H_b^2(G, H;\mathbb{R}) = 0$.*

2.5 Quasimorphisms

In this subsection, we recall the notion of quasimorphism on a group G. Typically, one proves that $H_b^2(G;\mathbb{R})$ is infinite-dimensional by demonstrating the existence of infinitely many linearly independent quasimorphisms on G.

Definition 2.26 A map $\phi : G \to \mathbb{R}$ is called a *quasimorphism* if there exists a constant $C > 0$ such that

$$|\phi(gh) - \phi(g) - \phi(h)| < C \quad \text{for all } g, h \in G.$$

The *defect* of a quasimorphism ϕ is defined to be

$$\Delta(\phi) := \sup_{g,h\in G} |\phi(gh) - \phi(g) - \phi(h)|.$$

The defect of a quasimorphism measures how far it is from a genuine homomorphism from the group to $\mathbb{R}$. A quasimorphism ϕ is called *trivial* if there exists a bounded map $b : G \to \mathbb{R}$ and a group homomorphism $\rho : G \to \mathbb{R}$ such that $\phi = b + \rho$.

We denote by $\mathrm{QM}(G)$ the $\mathbb{R}$-vector space of quasimorphisms on G and by $\mathrm{QM}_0(G)$ the subspace of $\mathrm{QM}(G)$ consisting of all trivial quasimorphisms on G, which is exactly $C_b^1(G;\mathbb{R}) \oplus \mathrm{Hom}(G, \mathbb{R})$.

For a quasimorphism ϕ on G, define the 1-coboundary of ϕ by $d^1\phi(g, h) = \phi(g) + \phi(h) - \phi(gh)$. Thus, $d^1\phi$ is a bounded 2-cocycle. We denote by $\omega_\phi := [d^1\phi]_b$ the corresponding bounded cohomology class. Then we have a linear map

$$\mathrm{QM}(G) \to H_b^2(G;\mathbb{R}), \quad \phi \mapsto \omega_\phi,$$

such that the sequence

$$\mathrm{QM}(G) \to H_b^2(G;\mathbb{R}) \to H^2(G;\mathbb{R})$$

is exact [12, Theorem 2.50].

There is another important special kind of quasimorphism called a *homogeneous quasimorphism*.

Definition 2.27 A quasimorphism $\phi : G \to \mathbb{R}$ is homogeneous if $\phi(g^n) = n \cdot \phi(g)$ for every $g \in G$ and every $n \in \mathbb{Z}$.

A *class function* on G is a function that takes the same value on each conjugacy class.

Lemma 2.28 *A homogeneous quasimorphism is a class function.*

Proof Let ϕ be a homogeneous quasimorphism on G. For any two group elements $f, g \in G$, it follows from the definition of homogeneous quasimorphisms that

$$\begin{aligned}|\phi(g) - \phi(fgf^{-1})| = |\phi(g) + \phi(fg^{-1}f^{-1})| &= \frac{|\phi(g^n) + \phi(fg^{-n}f^{-1})|}{n} \\ &\le \frac{|\phi(g^n) + \phi(g^{-n})| + |\phi(f)| + |\phi(f^{-1})| + 2\Delta(\phi)}{n} = \frac{2|\phi(f)| + 2\Delta(\phi)}{n}.\end{aligned}$$

By letting $n \to \infty$, one has that $\phi(g) = \phi(fgf^{-1})$. Since f, g are arbitrary, we complete the proof. □

Remark 2.29 [12, Lemma 2.21, Corollary 2.59] Let ϕ be a quasimorphism on G. Then there exists a unique homogeneous quasimorphism $\bar{\phi}$ which stays at finite distance from ϕ. In fact, the corresponding homogeneous quasimorphism to ϕ is given by

$$\bar{\phi}(g) := \lim_{n\to\infty} \frac{\phi(g^n)}{n} \quad \text{for all } g \in G.$$

This limit exists because the coarse subadditive inequality $\phi(g^{n+m}) \le \phi(g^n) + \phi(g^m) + \Delta(\phi)$ holds. Moreover, the defect $\Delta(\bar{\phi})$ is related to $\Delta(\phi)$ by $\Delta(\bar{\phi}) \le 2\Delta(\phi)$.

We denote the subspace of $\mathrm{QM}(G)$ consisting of all homogeneous quasimorphisms on G as $\mathrm{QM}_h(G)$.

3 Morse subgroups of infinite index

In this section, we assume that G is a nonelementary group acting properly on a geodesic metric space (X, d) with contracting elements. Fix a base point $o \in X$. We first define what is a Morse subgroup.

Definition 3.1 (Morse property) A subset $A \subset X$ is *η-Morse* for a function $\eta : \mathbb{R} \to \mathbb{R}$ if every λ-quasigeodesic with endpoints in A is contained in the $\eta(\lambda)$-neighborhood of A. The function η is called a *Morse gauge* of A. A subgroup $H \le G$ is *Morse* if the subset $H \cdot o$ is η-Morse for some function $\eta : \mathbb{R} \to \mathbb{R}$.

Remark 3.2 A C-contracting set is η_0-Morse for some function $\eta_0 : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ depending only on C [45, Lemma 2.8(1)].

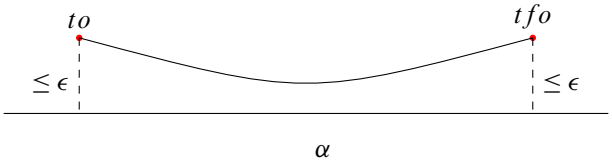


Figure 6: α contains an (ϵ, f)-barrier.

From now on, we assume that $\{H_i\}_{1\le i\le n}$ is a finite collection of Morse subgroups with infinite index in G. The purpose of this section is to prove Proposition 1.3, which is restated as:

Proposition 3.3 *There exists a quasitree on which G acts coboundedly and each H_i acts elliptically for $1 \le i \le n$.*

Proof ideas of Proposition 3.3 For a contracting element $g \in G$, Section 2.3 shows that there exists a projection complex $\mathcal{P}_K(\mathcal{F})$ where $\mathcal{F} = \{f \operatorname{Ax}(g) : f \in G\}$ such that $\mathcal{P}_K(\mathcal{F})$ is a quasitree on which G acts coboundedly. Hence, we only need to find an appropriate contracting element $g \in G$ such that each H_i acts elliptically on $\mathcal{P}_K(\mathcal{F})$ for $1 \le i \le n$. By analyzing the algebraic and geometric properties of $\{H_i : 1 \le i \le n\}$, we can obtain a contracting element g such that the projection of $[o, ho]$ onto $\operatorname{Ax}(g)$ is uniformly bounded for any $h \in \bigcup_{1\le i\le n} H_i$. Finally, we will show that such a contracting element g meets the requirements.

In order to characterize the magnitude of the (closest point) projection to an axis of a contracting element, we introduce a definition of barriers.

Definition 3.4 [48, Definition 4.1] Let $\epsilon \ge 0$ and $f \in G$. We say a geodesic $\alpha \subset X$ contains an (ϵ, f)-*barrier* if there exists an element $t \in G$ such that

$$d(to, \alpha) \le \epsilon, \quad d(tfo, \alpha) \le \epsilon.$$

Otherwise, α is called (ϵ, f)-*barrier-free*. An element $g \in G$ is called (ϵ, f)-*barrier-free* if some choice of geodesic from o to go is (ϵ, f)-barrier-free. See Figure 6 for an illustration.

Recall that the extension lemma, i.e., Lemma 2.13, gives a set $F \subset G$ consisting of three contracting elements. The following result of Han–Yang–Zou shows that for any contracting element $g = g_0 f$ obtained by the extension lemma, any geodesic segment with a large projection to $\operatorname{Ax}(g)$ contains an (ϵ, g_0)-barrier where ϵ depends only on F.

Lemma 3.5 [31, Lemma 2.9] *For any $g_0 \in G$, let $g = g_0 f$ be the contracting element given by Lemma 2.13. Then there exist $\epsilon = \epsilon(F)$ and $\tau = \tau(F, g)$ such that a geodesic segment α with $\operatorname{diam}(\pi_{\operatorname{Ax}(g)}(\alpha)) > \tau$ contains an (ϵ, g_0)-barrier.*

Remark 3.6 Actually, one can strengthen the conclusion of Lemma 3.5 as follows: if a geodesic segment α satisfies $\operatorname{diam}(\pi_{b\operatorname{Ax}(g)}(\alpha)) > \tau$ for some $b \in G$, then α contains an (ϵ, g_0)-barrier. The reason is that from the definition of barriers (see Definition 3.4), it is equivalent to say that α or $b^{-1}\alpha$ contains an (ϵ, g_0)-barrier. Thus, by substituting α with $b^{-1}\alpha$, one can apply Lemma 3.5 to obtain the stronger conclusion.

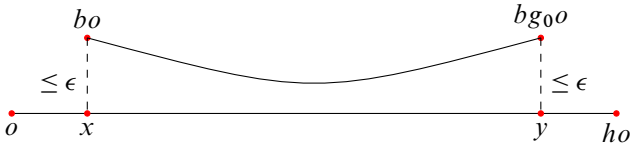


Figure 7: $[o, ho]$ contains an (ϵ, g_0)-barrier.

To get a contracting element such that the projection of each $[o, ho]$ to $\mathrm{Ax}(g)$ is uniformly bounded, we need to find an element g_0 such that h is (ϵ, g_0)-barrier-free for $h \in \bigcup_{1\le i\le n} H_i$.

Lemma 3.7 [39, Lemma 4.1] *Let the group G be the union of finitely many cosets of subgroups $C_1, \dots, C_n$. Then the index of (at least) one of these subgroups in G does not exceed n.*

Lemma 3.8 *For any $\epsilon \ge 0$, there exists an element $g_0 \in G$ such that any $h \in H_i$ with $1 \le i \le n$ is (ϵ, g_0)-barrier-free.*

Proof At first, we claim that the set $G \setminus S\big(\bigcup_{1\le i\le n} H_i\big)S$ is infinite for any finite subset $S \subset G$. Suppose not, by enlarging S by a finite set, we can assume that

$$G = S\Big(\bigcup_{1\le i\le n} H_i\Big)S = \bigcup_{1\le i\le n} (SH_iS) = \bigcup_{s\in S}\bigcup_{1\le i\le n} (sH_is^{-1})(s\cdot S),$$

where each $s\cdot S$ is a finite set. As each conjugate of H_i is still an infinite-index subgroup of G, the above decomposition implies that G can be written as a finite union of cosets of infinite-index subgroups. This contradicts Lemma 3.7. The claim thus follows.

Since each H_i is Morse for $1 \le i \le n$, it follows from Definition 3.1 that there exists a function $\eta : \mathbb{R} \to \mathbb{R}$ such that all orbits $H_i \cdot o (1 \le i \le n)$ are η-Morse. Let $M = \eta(1)$. Define $S := \{g \in G : d(o, go) \le \epsilon + M\}$. As G acts properly on X, the set S is finite.

According to the above claim, the set $G \setminus S\big(\bigcup_{1\le i\le n} H_i\big)S$ is infinite. We are going to show that every element $g_0 \in G \setminus S\big(\bigcup_{1\le i\le n} H_i\big)S$ satisfies the requirement. Suppose to the contrary that there exist some $i \in \{1, \dots, n\}$ and some element $h \in H_i$ such that h is not (ϵ, g_0)-barrier-free. By definition of barriers, for any geodesic segment $[o, ho]$, there exist an element $b \in G$ and two points $x, y \in [o, ho]$ such that $d(bo, x), d(bg_0o, y) \le \epsilon$. See Figure 7 for an illustration.

Since the orbit $H_i \cdot o$ is η-Morse in X, by definition of Morse property, the geodesic segment $[o, ho]$ is contained in $N_M(H_io)$. Since $x, y \in [o, ho]$, there exist $h_1, h_2 \in H_i$ such that $d(x, h_1o), d(y, h_2o) \le M$. Combining two inequalities together, one gets that $d(bo, h_1o), d(bg_0o, h_2o) \le \epsilon + M$. By construction of S, we have $b^{-1}h_1, h_2^{-1}bg_0 \in S$. Therefore,

$$g_0 \in b^{-1}h_2S = b^{-1}h_1 \cdot h_1^{-1}h_2S \subset SH_iS,$$

which contradicts with the choice of g_0. □

By combining Remark 3.6 and Lemma 3.8, we obtain that:

Lemma 3.9 *There exist a contracting element $g \in G$ and $\tau > 0$ such that for any $b \in G$ and $h \in H_i$ with $1 \le i \le n$,*

$$\operatorname{diam}(\pi_{b\,\mathrm{Ax}(g)}([o, ho])) \le \tau.$$

Proof We caution the readers that the constant $\epsilon := \epsilon(F)$ given by Lemma 3.5 only depends on a prefixed set F. Then we choose g_0 by Lemma 3.8 such that each $h \in H_i$ with $1 \le i \le n$ is (ϵ, g_0)-barrier-free. Set $g = g_0 f$ given by Lemma 2.13 for some $f \in F$. Then $\tau = \tau(F, g)$ in Lemma 3.5 is the desired uniform projection constant. Otherwise, one gets a contradiction to Remark 3.6. □

For the contracting element given by Lemma 3.9, we next follow the process in Section 2.3 to construct a projection complex. Finally, we show that the projection complex is the quasitree which satisfies all requirements of Proposition 3.3.

Group actions on projection complex Let g be a fixed contracting element given by Lemma 3.9. Lemma 2.15 shows that $\mathcal{F} = \{f\,\mathrm{Ax}(g) : f \in G\}$ with shortest projection maps satisfies the projection axioms with constants $\kappa = \kappa(\mathcal{F}) > 0$. Hence, one can construct a projection complex $\mathcal{P}_K(\mathcal{F})$ (see Definition 2.17) for $K \gg 0$. As a result of Lemma 2.18, $\mathcal{P}_K(\mathcal{F})$ is a quasitree, on which G acts nonelementarily and coboundedly. Set $K \ge \tau + 2\kappa + 2\epsilon$ where τ is given by Lemma 3.9, κ is given by Lemma 2.15, and ϵ is the fellow travel constant (see Proposition 2.10) with respect to a (D, B)-admissible path given by Lemma 2.19.

Lemma 3.10 *For each $i \in \{1, \dots, n\}$, H_i acts elliptically on $\mathcal{P}_K(\mathcal{F})$.*

Proof Recall that the vertex set of $\mathcal{P}_K(\mathcal{F})$ is $\mathcal{F} = \{f\,\mathrm{Ax}(g) : f \in G\}$, and two vertices U, V are connected by an edge if the set $\mathcal{F}_K[U, V] = \{Z \in \mathcal{F} : d_Z(U, V) > K\}$ is empty. Let $U \in \mathcal{P}_K(\mathcal{F})$ be the point representing $\mathrm{Ax}(g)$ which by definition is the orbit $E(g) \cdot o$. It suffices for us to show that $d_{\mathcal{P}}(U, hU) \le 1$ for any $h \in H_i$ with $1 \le i \le n$.

Suppose not; then there exists at least one element in $\mathcal{F}_K[U, hU]$. Let $V \in \mathcal{F}_K[U, hU]$. It follows from the definition of $\mathcal{F}_K[U, hU]$ that $\operatorname{diam}(\pi_V(U) \cup \pi_V(hU)) > K$. Note that $o \in U$ and thus $ho \in hU$. Lemma 2.19 shows that there exists a (D, B)-admissible path in X from o to ho with saturation $\mathcal{F}_K[U, hU]$. It follows from the ϵ-fellow travel property of an admissible path that

$$\operatorname{diam}(\pi_V([o, ho])) \ge \operatorname{diam}(\pi_V(U) \cup \pi_V(hU)) - 2\kappa - 2\epsilon > K - 2\kappa - 2\epsilon \ge \tau.$$

As V represents a G-translate of $\mathrm{Ax}(g)$, one gets a contradiction to Lemma 3.9. □

As a corollary of Lemma 3.10, we get Proposition 3.3.

Remark 3.11 With Lemma 3.10, we know that the quasitree in Proposition 3.3 is actually a projection complex $\mathcal{P}_K(\mathcal{F})$. Since each vertex in $\mathcal{P}_K(\mathcal{F})$ represents a translate of $\mathrm{Ax}(g)$, any conjugate of $E(g)$ fixes a point in $\mathcal{P}_K(\mathcal{F})$. Hence, the action $G \curvearrowright \mathcal{P}_K(\mathcal{F})$ is not proper. Furthermore, it is generally difficult to obtain subgroups acting elliptically on $\mathcal{P}_K(\mathcal{F})$ other than the vertex stabilizer. However, Proposition 3.3 provides such a family of subgroups, i.e., Morse subgroups of infinite index.

4 Constructing quasimorphisms

In this section, we assume that a nonelementary countable group G acts WPD (see Definition 4.2) on a δ-hyperbolic space X and a finite collection of subgroups $\{H_i : 1 \le i \le n\}$ of G acts elliptically on X. Our goal is the following result.

Proposition 4.1 *There is an injective $\mathbb{R}$-linear map $\omega : \ell^1 \to H_b^2(G;\mathbb{R})$ such that each coclass in the image $\omega(\ell^1)$ has a representative vanishing on H_i for each $1 \le i \le n$.*

Moreover, the dimension of $H_b^2(G, \{H_i\}_{i=1}^n;\mathbb{R})$ as a vector space over $\mathbb{R}$ has the cardinality of the continuum.

Definition 4.2 [5] We say that the action of G on a hyperbolic space X satisfies *WPD* if

- G is not virtually cyclic,
- G contains at least one element that acts on X as a loxodromic isometry, and
- for every loxodromic element $g \in G$, every $x \in X$, and every $C > 0$, there exists $N > 0$ such that

$$\{h \in G : d(x, hx) \le C, d(g^N x, hg^N x) \le C\}$$

  is finite.

In [5], Bestvina–Fujiwara showed that the dimension of $\mathrm{QM}(G)/(H^1(G;\mathbb{R}) \oplus C_b^1(G;\mathbb{R}))$ as a $\mathbb{R}$-vector space has the cardinality of the continuum under the assumption of WPD actions. This implies the absolute version of Proposition 4.1. However, the relative version of second bounded cohomology of a group acting WPD on a hyperbolic space has never been studied before, which is where the value of our Proposition 4.1 lies. Regarding the proof of Proposition 4.1, we generally follow the proof idea in [5], but we will utilize some new techniques (e.g., barriers) developed in this paper.

4.1 Epstein–Fujiwara quasimorphisms on groups acting on hyperbolic spaces

At first, let us recall some basic material about Epstein–Fujiwara quasimorphisms introduced in [24].

Let α be a finite path in X. We denote the length of α by $|\alpha|$. We use the action of $g \in G$ on X to define a path $g \cdot \alpha$ which is the g-translation of the path α. We say that $g \cdot \alpha$ is a *copy* of α. Let w be a finite oriented path, and let w^{-1} be the inverse path. We assume that $|w| \ge 2$ and define

$$|\alpha|_w := \{\text{the maximal number of copies of } w \text{ in } \alpha \text{ without overlapping (except at the vertices)}\}.$$

Suppose that $x, y \in X$ and that W is a number with $0 < W < |w|$. Recall that $[x, y]$ denotes some choice of a geodesic from x to y. We define

$$c_{w,W}([x, y]) = d(x, y) - \inf_{\alpha}\{|\alpha| - W|\alpha|_w\}, \tag{4-1}$$

where α ranges over all the paths from x to y. It follows from the definition that $c_{w,W}([x, y])$ does not depend on the choice of a geodesic $[x, y]$.

Remark 4.3 By choosing α to be a choice of geodesic $[x, y]$, one gets that

$$c_{w,W}([x, y]) \geq d(x, y) - (|[x, y]| - W|[x, y]|_w) = W|[x, y]|_w \geq 0.$$

Moreover, if $c_{w,W}([x, y]) = 0$, then the above inequality implies that $|[x, y]|_w = 0$ and the geodesic $[x, y]$ realizes the infimum in (4-1). However, if $c_{w,W}([x, y]) > 0$, then the realizing path (i.e., a path realizing the infimum in (4-1)) may not exist.

Lemma 4.4 [24, Lemma 3.3] *Suppose that a path α realizes the infimum above. Then the path α is a $\big(\frac{|w|}{|w|-W}, \frac{2W|w|}{|w|-W}\big)$-quasigeodesic.*

Since a realizing path does not always exist, we need another notion which is close to realizing paths in practice. A path β between x and y is called an *almost realizing path* of $c_{w,W}([x, y])$ if it satisfies that

$$|\beta| - W|\beta|_w \leq \min\big\{\inf_\alpha\{|\alpha| - W|\alpha|_w\} + W,\ \big(d(x, y) + \inf_\alpha\{|\alpha| - W|\alpha|_w\}\big)/2\big\}, \tag{4-2}$$

where α ranges over all the paths from x to y. In other words, an almost realizing path β of $c_{w,W}([x, y])$ satisfies that

$$d(x, y) - (|\beta| - W|\beta|_w) \geq \max\big\{c_{w,W}([x, y]) - W,\ c_{w,W}([x, y])/2\big\}.$$

We remark that the requirement $d(x, y) - (|\beta| - W|\beta|_w) \geq c_{w,W}([x, y]) - W$ guarantees β is a uniform quasigeodesic (see Lemma 4.5 below) and the other requirement $d(x, y) - (|\beta| - W|\beta|_w) \geq c_{w,W}([x, y])/2$ is used to obtain $|\beta|_w > 0$ when $c_{w,W}([x, y]) > 0$. By definition and Remark 4.3, an almost realizing path of $c_{w,W}([x, y])$ always exists. Analogous to Lemma 4.4, we have that:

Lemma 4.5 *Let β be an almost realizing path of $c_{w,W}([x, y])$. Then the path β is a $\big(\frac{|w|}{|w|-W}, \frac{3W|w|}{|w|-W}\big)$-quasigeodesic.*

Proof Let $\beta : [0, |\beta|] \to X$ be an arc-length parametrization of β. Let $0 \leq t < s \leq |\beta|$ and set $\beta' = \beta|_{[t,s]}$. Note that $|\beta'| = s - t$. Let γ be a geodesic from $\beta(t)$ to $\beta(s)$.

Claim $$|\beta'| - W(|\beta'|_w + 3) \leq |\gamma| - W|\gamma|_w.$$

Proof of claim Suppose to the contrary that $|\beta'| - W(|\beta'|_w + 3) > |\gamma| - W|\gamma|_w$. Since β is an almost realizing path, $|\beta| - W|\beta|_w \leq \inf_\alpha\{|\alpha| - W|\alpha|_w\} + W$. By setting $\gamma' = \beta|_{[0,t]} \cup \gamma \cup \beta|_{[s,|\beta|]}$, one has that

$$\begin{aligned}
|\gamma'| - W|\gamma'|_w &\leq (t + |\gamma| + |\beta| - s) - W\big(\big|\beta|_{[0,t]}\big|_w + |\gamma|_w + \big|\beta|_{[s,|\beta|]}\big|_w\big) \\
&< |\beta| - W\big(\big|\beta|_{[0,t]}\big|_w + |\beta'|_w + \big|\beta|_{[s,|\beta|]}\big|_w + 3\big) \\
&\leq |\beta| - W(|\beta|_w + 1) \leq \inf_\alpha\{|\alpha| - W|\alpha|_w\}.
\end{aligned}$$

This is impossible since γ' is also a path from x to y. □

Clearly $|\beta'|_w \leq |\beta'|/|w|$. Therefore,

$$\begin{aligned}
d(\beta(t), \beta(s)) = |\gamma| &\geq |\gamma| - W|\gamma|_w \geq |\beta'| - W|\beta'|_w - 3W \\
&\geq |\beta'| - \frac{W}{|w|}|\beta'| - 3W = \frac{|w| - W}{|w|}|\beta'| - 3W.
\end{aligned}$$

□

It is clear from the definition that $c_{w,W}([x,y]) = c_{w^{-1},W}([y,x])$. We then define

$$h_{w,W}([x,y]) = c_{w,W}([x,y]) - c_{w^{-1},W}([x,y]).$$

Take $o \in X$ as a base point. We define functions $c_{w,W}$ and $h_{w,W} : G \to \mathbb{R}$ by

$$c_{w,W}(g) := c_{w,W}([o, g\cdot o]) \quad \text{and} \quad h_{w,W}(g) := h_{w,W}([o, g\cdot o]).$$

Lemma 4.6 [24, Proposition 3.10] *The map $h_{w,W} : G \to \mathbb{R}$ is a quasimorphism. Moreover, the defect $\Delta(h_{w,W}) \le 12L_0 + 6W + 48\delta$ is uniformly bounded where $L_0 = L\big(\frac{|w|}{|w|-W}, \frac{2W|w|}{|w|-W}, \delta\big)$ is given by Lemma 2.1.*

4.2 Constructing infinitely many words

From now on, we borrow some definitions and terminologies from [5]. For a base point $o \in X$ and a loxodromic element g, we denote by $L_g = \bigcup_{i\in\mathbb{Z}} g^i[o, go]$ a quasiaxis of g. Define $\mathrm{Ax}(g) := \bigcup_{k\ge 1} L_{g^k}$. By the Morse lemma, $\mathrm{Ax}(g)$ is still a quasiaxis of g. The quasiaxis $\mathrm{Ax}(g)$ of g is oriented by the requirement that g acts as a positive translation. We call this orientation the *g-orientation* of the quasiaxis. Of course, the g^{-1}-orientation is the opposite of the g-orientation. Let $\mathrm{Ax}(g)$ be a (λ,ϵ)-quasiaxis. By the Morse lemma, any two (λ,ϵ)-quasiaxes of g are within $L(\lambda,\epsilon,\delta)$ of each other. More generally, any sufficiently long path J inside the $L(\lambda,\epsilon,\delta)$-neighborhood of $\mathrm{Ax}(g)$ of g has a natural orientation given by g: a point of $\mathrm{Ax}(g)$ within $L(\lambda,\epsilon,\delta)$ of the terminal endpoint of J is ahead (with respect to the g-orientation of $\mathrm{Ax}(g)$) of a point of $\mathrm{Ax}(g)$ within $L(\lambda,\epsilon,\delta)$ of the initial endpoint of J. We call this orientation of J the *g-orientation*.

Definition 4.7 Let g_1 and g_2 be two loxodromic elements of G. We will write

$$g_1 \sim g_2$$

if there exists a constant $L' > 0$ such that an arbitrarily long segment J in $\mathrm{Ax}(g_1)$ is contained in an L'-neighborhood of $t\,\mathrm{Ax}(g_2)$ for some $t \in G$ and the map $t : J \to t(J)$ is orientation-preserving with respect to the g_1-orientation on J and the g_2-orientation on $t(J)$.

Note that $\sim$ is an equivalence relation. The following lemma gives a relation between the above definition and the definition (see Definition 3.4) of barriers which has nothing to do with the orientation.

Lemma 4.8 *If $g_1 \nsim g_2^{\pm 1}$, then for any $\epsilon' > 0$, there exists $r > 0$ such that g_2^m is (ϵ', g_1^s)-barrier-free for any $m \in \mathbb{Z}$ and $s \ge r$.*

Proof Suppose to the contrary that there exists $\epsilon' > 0$ such that for every $r \in \mathbb{N}$, there exist $m \in \mathbb{Z}$ and $s \ge r$ such that g_2^m is not (ϵ', g_1^s)-barrier-free. According to the definition of barriers, there exists $t \in G$ such that $d(to, [o, g_2^m o]) \le \epsilon'$ and $d(tg_1^s o, [o, g_2^m o]) \le \epsilon'$. See Figure 8 for an illustration.

Let $x, y \in [o, g_2^m o]$ such that $d(to, x) = d(to, [o, g_2^m o])$ and $d(tg_1^s o, y) = d(tg_1^s o, [o, g_2^m o])$. Hence, the path $\gamma = [x, to] \cup [to, tg_1^s o] \cup [tg_1^s o, y]$ is a $(1, 4\epsilon')$-quasigeodesic. By Lemma 2.1, $t[o, g_1^s o] \subset N_{L'}([o, g_2^m o])$

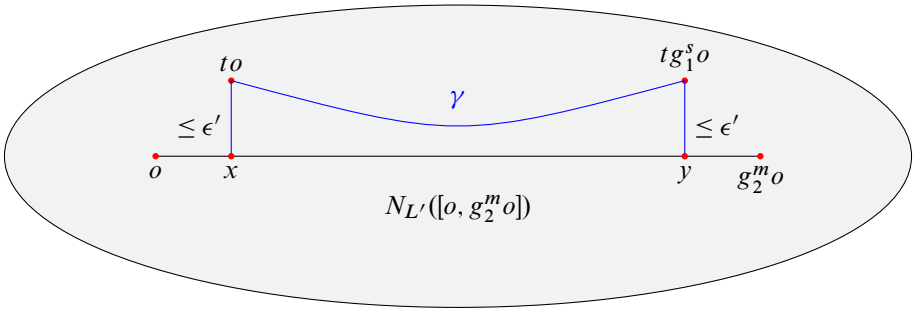


Figure 8: $[o, g_2^m o]$ contains an (ϵ', g_1^s)-barrier. The blue path represents γ. The gray area represents $N_{L'}([o, g_2^m o])$.

where $L' = L(1, 4\epsilon', \delta)$. Up to exchanging g_2 to g_2^{-1}, we can assume that the map t is orientation-preserving. Note that $s \geq r$. By Lemma 2.2, one has $d(o, g_1^s o) \geq s\|g_1\| \geq r\|g_1\|$, which implies that the length of $[o, g_1^s o] \subseteq \mathrm{Ax}(g_1)$ goes to infinity as $r \to \infty$. According to Definition 4.7, this shows that $g_1 \sim g_2$ or $g_1 \sim g_2^{-1}$ which is impossible. □

Proposition 6 in [5] shows that for every WPD action $G \curvearrowright X$, there exist two loxodromic elements g_1 and g_2 such that $g_1 \nsim g_2$. In [5, Proposition 2, Claim 1], Bestvina–Fujiwara proved that two elements f_1, f_2 are nonequivalent if they satisfy

$$f_1 = g_1^{n_1} g_2^{m_1} g_1^{k_1} g_2^{-l_1}, \quad f_2 = g_1^{n_2} g_2^{m_2} g_1^{k_2} g_2^{-l_2},$$

where $0 \ll n_1 \ll m_1 \ll k_1 \ll l_1 \ll n_2 \ll m_2 \ll k_2 \ll l_2$. But in fact, their proof only requires $0 \ll n_1, m_1, |k_1|, l_1 \ll n_2, m_2, |k_2|, l_2$, and does not require k_1, k_2 to be positive. So we can take k_1, l_1, k_2, l_2 to be $-n_1$, m_1, $-n_2$, m_2 respectively. In this case, we have $f_1 = g_1^{n_1} g_2^{m_1} g_1^{-n_1} g_2^{-m_1}$, $f_2 = g_1^{n_2} g_2^{m_2} g_1^{-n_2} g_2^{-m_2} \in [G, G]$ and the remaining proof of [5, Proposition 2, Claim 1] shows that:

Lemma 4.9 *There exist two loxodromic elements g_1 and g_2 in $[G, G]$ on X such that $g_1 \nsim g_2$.*

Since g_1 and g_2 are independent, we may replace g_1, g_2 by high positive powers of conjugates to ensure that the subgroup F of G generated by g_1, g_2 is free with basis $S = \{g_1, g_2\}$, each nontrivial element of F is loxodromic, and F is quasiconvex with respect to the action on X (see [24, Proposition 4.3]). We will call such free subgroups *Schottky groups*. Let $\mathcal{G}(F, S)$ be the Cayley graph of F with respect to the generating set $S = \{g_1, g_2\}$. Then $\mathcal{G}(F, S)$ is a tree and each oriented edge has a label $g_i^{\pm 1}$. Choose a base point $o \in X$ and construct an F-equivariant map $\Phi : \mathcal{G}(F, S) \to X$ that sends 1 to o and sends each edge to a geodesic arc. Quasiconvexity implies that Φ is a (λ_0, ϵ_0)-quasi-isometric embedding for some $\lambda_0 \geq 1, \epsilon_0 \geq 0$ and in particular for every $1 \neq f \in F$ the Φ-image of the axis of f in $\mathcal{G}(F, S)$ is a (λ_0, ϵ_0)-quasiaxis of f in X.

Choose positive constants

$$0 \ll n_1 \ll m_1 \ll k_1 \ll l_1 \ll n_2 \ll m_2 \ll \cdots$$

and define

$$f_i = g_1^{n_i} g_2^{m_i} g_1^{k_i} g_2^{-l_i}$$

for $i = 1, 2, 3, \ldots$.

Proposition 4.10 [5, Proposition 2] *$\{f_i : i \geq 1\}$ is an infinite sequence of loxodromic elements in G such that*

(1) *$f_i \nsim f_i^{-1}$ for $i \geq 1$, and*

(2) *$f_i \nsim f_j^{\pm 1}$ for $j < i$.*

If $f \in F$ is cyclically reduced as a word in g_1, g_2 (equivalently, if its axis passes through $1 \in \mathcal{G}(F, S)$) then by the quasiconvexity of F in G we have

$$d(o, f^m(o)) \geq m\big(d(o, f(o)) - 2L_1\big), \tag{4-3}$$

where $L_1 = L(\lambda_0, \epsilon_0, \delta) > 0$ is given by Lemma 2.1 and is a constant independent of f and m.

4.3 Constructing infinitely many quasimorphisms

From now on, we fix an integer $W \geq 3L_1$ and will only consider a path w with $|w| > W$. Thus, an almost realizing path α as in Lemma 4.5 will be a quasigeodesic with constants independent of w and the endpoints. Moreover, α is contained in a uniform neighborhood, say, L_2-neighborhood, of any geodesic joining the endpoints of α. We will also omit W from the notation and write c_w and h_w for simplicity.

The next lemma is crucial for our discussion, not only in absolute bounded cohomology but also in the relative case. Recall the definition of barriers from Definition 3.4.

Lemma 4.11 *Let $w = [o, fo]$ and $g \in G$ be an (L_2, f)-barrier-free element. Then we have $c_w(g) = 0$ and $c_{w^{-1}}(g) = 0$.*

Proof Assume that $c_w(g) > 0$ and that α is an almost realizing path of $c_w(g)$. From Lemma 4.5 we know that α is a $\big(\frac{|w|}{|w|-W}, \frac{3W|w|}{|w|-W}\big)$-quasigeodesic. Thus, $\alpha \subseteq N_{L_2}([o, go])$. Additionally, we have $d(o, go) - (|\alpha| - W|\alpha|_w) \geq c_w(g)/2 > 0$. Therefore, $|\alpha|_w > (|\alpha| - d(o, go))/W > 0$. From the definition of $|\alpha|_w$, there exists some element $t \in G$ such that $t \cdot w \subseteq \alpha \subseteq N_{L_2}([o, go])$. This leads to a contradiction, as g is (L_2, f)-barrier-free. Therefore, $c_w(g) = 0$. We note that as long as there is no $t \in G$ such that $t \cdot w \subseteq N_{L_2}([o, go])$, there is also no $t \in G$ such that $t \cdot w^{-1} \subseteq N_{L_2}([o, go])$. The conclusion that $c_{w^{-1}}(g) = 0$ then follows. □

For simplicity, for any $f \in G$, we set $c_f := c_{[o, fo]}$ and $h_f := h_{[o, fo]}$. Let $\{f_i : i \geq 1\}$ be the sequence from Proposition 4.10. As the relation $\sim$ is invariant under conjugation, we assume in addition that each f_i is cyclically reduced.

Lemma 4.12 *For all $i \geq 1$, there exists $r_i > 0$ such that for all $j < i$, we have*

(1) *$h_{f_i^{r_i}}(f_i^{r_i m}) \geq L_1 m$ for any $m \geq 0$,*

(2) *$h_{f_i^{r_i}}$ is 0 on $\langle f_j \rangle$.*

Proof We first prove item (2). Fix $i \geq 1$. As each element in the finite set $A := \{f_1^{\pm 1}, \dots, f_{i-1}^{\pm 1}\}$ is not equivalent to f_i, for any sufficiently large $\epsilon' > 0$, Lemma 4.8 gives a constant $r_i > 0$ such that each $a \in A$ satisfies that a^m is $(\epsilon', f_i^{r_i})$-barrier-free for all $m \in \mathbb{Z}$.

Let $\epsilon' \geq L_2$. As a result of Lemma 4.11, we obtain

$$c_{f_i^{r_i}}(a^m) = c_{f_i^{-r_i}}(a^m) = 0$$

for any $a \in A$ and $n \in \mathbb{Z}$.

Hence, $h_{f_i^{r_i}}(f_j^m) = c_{f_i^{r_i}}(f_j^m) - c_{f_i^{-r_i}}(f_j^m) = 0$ for all $m \in \mathbb{Z}$.

Now, we turn to proving item (1).

Claim *For each $i \geq 1$, there exists $r_i' > 0$ such that $c_{f_i^r}(f_i^{-rm}) = 0$ for any $m \geq 0$ and $r \geq r_i'$.*

Proof of claim Suppose not. Then there exists $i \geq 1$ such that for every sufficiently large $r_i' > 0$, there exists $m \geq 0$ and $r \geq r_i'$ such that $c_{f_i^r}(f_i^{-rm}) > 0$. Define $w = [o, f_i^r o]$. Let α be an almost realizing path of $c_w(f_i^{-rm})$ in (4-2) with f_i^{-1}-orientation. Then we have $d(o, f_i^{-rm}o) - (|\alpha| - W|\alpha|_w) \geq c_w(f_i^{-rm})/2 > 0$. Thus, $|\alpha|_w > (|\alpha| - d(o, f_i^{-rm}o))/W > 0$. From the definition of $|\alpha|_w$ there exists some element $t \in G$ such that $t \cdot w \subseteq \alpha \subseteq N_{L_2}([o, f_i^{-rm}o])$ and the map t respects the orientation. As r_i' can be arbitrarily large, this implies that $f_i \sim f_i^{-1}$, which is a contradiction. □

Now we return to the proof of item (1). For each $i \geq 1$, Lemma 4.8 allows us to require $r_i \geq r_i'$ where r_i is the constant appearing in the proof of item (2). Define $w = [o, f_i^{r_i} o]$. For $n \geq 1$, let γ be the concatenated path $\bigcup_{0 \leq k \leq m-1} f_i^{r_i k} w$. According to (4-3), we have

$$|\gamma| = m d(o, f_i^{r_i} o) \leq d(o, f_i^{r_i m} o) + 2L_1 m.$$

Obviously, $|\gamma|_w = m$. Recall that $W \geq 3L_1$. Then (4-1) gives that

$$c_{f_i^{r_i}}(f_i^{r_i m}) \geq d(o, f_i^{r_i m} o) - (|\gamma| - W|\gamma|_w) \geq d(o, f_i^{r_i m} o) - (|\gamma| - 3L_1 m) \geq L_1 m.$$

Therefore,

$$h_{f_i^{r_i}}(f_i^{r_i m}) = c_{f_i^{r_i}}(f_i^{r_i m}) - c_{f_i^{-r_i}}(f_i^{r_i m}) = c_{f_i^{r_i}}(f_i^{r_i m}) - c_{f_i^{r_i}}(f_i^{-r_i m}) \geq L_1 m. \qquad \square$$

Define $h_i : G \to \mathbb{R}$ as $h_i = h_{f_i^{r_i}}$, where $r_i > 0$ is chosen as in Lemma 4.12. Then we obtain the following.

Proposition 4.13 *$\{h_i : i \geq 1\}$ is an infinite sequence of quasimorphisms on G such that*

(1) *$h_i(f_j^m) = 0$ for all $i \neq j$ and for all $m \geq 0$;*

(2) *$h_i(f_i^{r_i m}) \geq L_1 m$ for all $i \geq 1$ and for all $m \geq 0$;*

(3) *$\psi(f_i) = 0$ for all homomorphisms $\psi : G \to \mathbb{R}$;*

(4) *the distance $d(o, f_i o)$ tends to infinity as i tends to infinity;*

(5) *$h_i(h) = 0$ for all $h \in H_j$ with $1 \leq j \leq n$.*

Proof The first two items follow directly from Lemma 4.12. Recall from Lemma 4.9 that $g_1, g_2 \in [G, G]$. Hence, each $f_i \in F = \langle g_1, g_2 \rangle$ is a product of finitely many commutators, which implies item (3). Recall from the paragraph following Lemma 4.9 that $F = \langle g_1, g_2 \rangle$ is a Schottky group which means that the orbital map $\Phi : \mathcal{G}(F, \{g_1, g_2\}) \to X$ is a (λ_0, ϵ_0)-quasi-isometric embedding. Since $f_i = g_1^{n_i} g_2^{m_i} g_1^{k_i} g_2^{-l_i}$ for $0 \ll n_1, m_1, k_1, l_1, \ll n_2, m_2, k_2, l_2 \ll \cdots$, one gets that $d(o, f_i o) \geq \lambda_0^{-1}(n_i + m_i + k_i + l_i) - \epsilon_0$ which implies item (4). It suffices for us to verify item (5).

As each $H_j (1 \leq j \leq n)$ acts elliptically on X, there is a $D > 0$ such that $d(o, ho) \leq D$ for all $h \in \bigcup_{1 \leq j \leq n} H_j$. Since $0 \ll n_1, m_1, k_1, l_1 \ll n_i, m_i, k_i, l_i$ for each $i \geq 2$, we can require $n_1 \gg 0$ such that $d(o, f_i^{r_i} o) \geq \lambda_0^{-1} r_i (n_i + m_i + k_i + l_i) - \epsilon_0 > 2L_2 + D$. Then it follows from the definition of barriers (see Definition 3.4) that each h is $(L_2, f_i^{r_i})$-barrier-free. Hence, as a result of Lemma 4.11, $c_{f_i^{r_i}}(h) = c_{f_i^{-r_i}}(h) = 0$. This shows that $h_i(h) = 0$ for all $h \in \bigcup_{1 \leq j \leq n} H_j$. □

At the end of this section, we prove Proposition 4.1.

Proof of Proposition 4.1 At first, we claim that for each $g \in G$ and $i \gg 0$, one has $h_i(g) = 0$. Indeed, as Proposition 4.13(4) shows, $d(o, f_i o)$ tends to infinity as $i \to \infty$. Hence, for $i \gg 0$, g is $(L_2, f_i^{r_i})$-barrier-free since the diameter of $N_{L_2}([o, go])$ is finite. Thus one gets that $h_i(g) = 0$ by Lemma 4.11. Therefore, if $(a_i)_{i=1}^{\infty} \in \ell^1$, then $\sum_{i=1}^{\infty} a_i h_i$ is well defined as an element of $C^1(G; \mathbb{R})$ since $\sum_{i=1}^{\infty} a_i h_i(g)$ is in fact a finite sum for each $g \in G$. For the same reason, $\sum_{i=1}^{\infty} a_i d^1 h_i$ is a well-defined 2-cocycle. By Lemma 4.6, all the 2-cocycles $d^1 h_i$ have a common bound, which means that there exists a constant $M > 0$ such that $\sup_{g, g' \in G} |d^1 h_i(g, g')| \leq \Delta(h_i) \leq M$ for $i \geq 1$. It follows that if a sequence $(a_i)_{i=1}^{\infty} \in \ell^1$ then $\sum_{i=1}^{\infty} a_i d^1 h_i$ is a bounded 2-cocycle. Therefore,

$$\sum_{i=1}^{\infty} a_i d^1 h_i = d^1 \left(\sum_{i=1}^{\infty} a_i h_i \right).$$

We get a real linear map $\omega : \ell^1 \to H_b^2(G; \mathbb{R})$ which sends the sequence $(a_i)_{i=1}^{\infty}$ to the cohomology class represented by $\sum_{i=1}^{\infty} a_i d^1 h_i$. From Proposition 4.13(5), we know each $d^1 h_i$ lies in $H_b^2(G, \{H_i\}_{i=1}^n; \mathbb{R})$. So the real linear map is actually $\omega : \ell^1 \to H_b^2(G, \{H_i\}_{i=1}^n; \mathbb{R})$. In order to see that ω is injective, suppose $\omega((a_i)) = 0$. Then

$$d^1 \left(\sum_{i=1}^{\infty} a_i h_i \right) = d^1 b$$

for some bounded real-valued map $b \in C_b^1(G; \mathbb{R})$. This means the function

$$\phi := \sum_{i=1}^{\infty} a_i h_i - b$$

is a homomorphism from G to $\mathbb{R}$. Applying this equality of 1-cochains to ${f_i}^{r_i m} \in G$, we find

$$a_i h_i({f_i}^{r_i m}) - b({f_i}^{r_i m}) = \phi({f_i}^{r_i m}) = 0 \quad \text{for all } m \geq 0.$$

Since $h_i(f_i^{r_i}m) \geq L_1 m$ and b is a bounded map, this forces a_i to be 0. As i is arbitrary, (a_i) must be the zero vector. This shows the injectivity of $\omega : \ell^1 \to H_b^2(G, \{H_i\}_{i=1}^n; \mathbb{R})$.

Finally, as ℓ^1 has the dimension equal to the cardinality of the continuum and the space of bounded cochains has cardinality $|\mathbb{R}^{\mathbb{N}}| = |\mathbb{R}|$, we complete the proof. □

As a consequence of Proposition 4.1, we have:

Proposition 4.14 *Let G be a countable group and H a normal subgroup of G. Suppose that G/H acts WPD on a hyperbolic space. Then the dimension of $H_b^2(G, H; \mathbb{R})$ as a vector space over $\mathbb{R}$ has the cardinality of the continuum.*

Proof Proposition 4.1 shows that the dimension of $H_b^2(G/H; \mathbb{R})$ as a vector space over $\mathbb{R}$ has the cardinality of the continuum. Since Proposition 2.22 shows that $H_b^2(G/H; \mathbb{R}) \cong H_b^2(G, H; \mathbb{R})$, the conclusion follows. □

5 The proof of Theorem 1.1

Let us recall the conditions of Theorem 1.1: G acts properly on X with contracting elements and $\{H_i\}$ $(1 \leq i \leq n)$ is a finite collection of Morse subgroups with infinite index in G.

Definition 5.1 The action of a group G on a metric space X is *acylindrical* if for all $L > 0$ there exist $D > 0$ and $B > 0$ such that if $x, y \in X$ and $d(x, y) > D$, then there are at most B elements $g \in G$ with $d(x, gx) \leq L$ and $d(y, gy) \leq L$.

An acylindrical action $G \curvearrowright X$ is called *nonelementary* if the action is unbounded and G is not virtually cyclic.

Lemma 5.2 *A nonelementary acylindrical action on a hyperbolic space must be WPD.*

Proof Let G be a group which admits a nonelementary acylindrical action on a hyperbolic space X. Since $G \curvearrowright X$ is acylindrical, for all $L > 0$, there exists $D > 0$ such that if $x, y \in X$ and $d(x, y) > D$, then the set $\{g \in G : d(x, gx) \leq L, d(y, gy) \leq L\}$ is finite.

Now we verify that $G \curvearrowright X$ is also a WPD action according to Definition 4.2. By [41, Theorem 1.1], G contains infinitely many independent loxodromic elements. Thus, it remains to verify the third item in Definition 4.2. For every loxodromic element $g \in G$, every $x \in X$, and every $L > 0$, let $D > 0$ be the constant given by the above acylindrical action and $N > 0$ be an integer depending only on g such that $d(x, g^N x) \geq N\|g\| > D$. Then the above acylindrical action implies that the set $\{h \in G : d(x, hx) \leq L, d(g^N x, hg^N x) \leq L\}$ is finite. □

Fix a contracting element g given by Lemma 3.9 and $K \gg 0$. Section 3 produces a projection complex $\mathcal{P}_K(\mathcal{F})$.

As shown in [45, Theorem 1.5], $E(g)$ is a hyperbolically embedded subgroup of G. Therefore, as a result of [4, Theorem 5.6], G acts acylindrically on $\mathcal{P}_K(\mathcal{F})$. In particular, as a result of Lemma 5.2:

Lemma 5.3 *The action $G \curvearrowright \mathcal{P}_K(\mathcal{F})$ satisfies WPD.*

Proof of Theorem 1.1 Recall that Lemma 3.10 shows that each H_i $(1 \le i \le n)$ acts elliptically on $\mathcal{P}_K(\mathcal{F})$. Hence, the action $G \curvearrowright \mathcal{P}_K(\mathcal{F})$ satisfies the setup of Section 4. Therefore, Theorem 1.1 follows from Proposition 4.1. □

A natural research direction to generalize our Theorem 1.1 is to assume instead that each subgroup H_i is a subgroup with proper limit sets on a convergence boundary of X. See [31; 49] for more details about convergence boundary. A Morse subgroup with infinite index satisfies this property. Hence, one may wonder whether the following proposition is always true:

Question 5.4 *Let G be a nonelementary countable group acting properly on a geodesic metric space X with convergence boundary. Let H be a subgroup of G with proper limit sets. Is the dimension of $H_b^2(G, H; \mathbb{R})$ as a vector space over $\mathbb{R}$ infinite?*

At the end of this section, we provide an application of Theorem 1.1.

Definition 5.5 A group G is *boundedly generated* by a finite collection of subgroups $H_1, \dots, H_k$ if for every $g \in G$ there is a number N such that all powers g^n can be written in the form

$$g^n = \prod_{i=1}^{N} h_i(n),$$

where each $h_i(n)$ is conjugate to some element in $\bigcup_{1 \le j \le k} H_j$.

This definition is a variation of [36, Definition 4]. There Kotschick required each subgroup to be cyclic.

Corollary 5.6 *Under the assumption of Theorem 1.1, G is not boundedly generated by $\{H_i : 1 \le i \le n\}$.*

Proof Suppose to the contrary that G is boundedly generated by $\{H_i : 1 \le i \le n\}$. Then for every $g \in G$, there exists an $N \in \mathbb{N}$ such that every power g^m can be written as a product of N elements in conjugations of $\bigcup_{1 \le i \le n} H_i$.

As Proposition 4.13 shows, there is at least one unbounded quasimorphism ϕ on G such that $\phi(h) = 0$ for each $h \in H_i$ with $1 \le i \le n$. Let $\bar{\phi}$ be the homogenization of ϕ given by Remark 2.29. Proposition 4.13 also gives an element $g \in G$ such that $\bar{\phi}(g) > 0$. Then there exists $N \in \mathbb{N}$ such that $g^m = h_1' \cdots h_N'$ for any $m > 0$ and each $h_i' = g_i h_i g_i^{-1}$ with $g_i \in G, h_i \in \bigcup_{1 \le j \le n} H_j$. Note that homogeneous quasimorphisms take constant values on conjugacy classes. Hence, $\bar{\phi}(h_i') = \bar{\phi}(h_i) = 0$ for each i. Then one has that

$$m|\bar{\phi}(g)| = |\bar{\phi}(g^m)| = |\bar{\phi}(h_1' \cdots h_N')| \le |\bar{\phi}(h_1' \cdots h_{N-1}')| + \Delta(\bar{\phi}) \le \cdots \le N\Delta(\bar{\phi}),$$

where $\Delta(\bar{\phi})$ is the defect of $\bar{\phi}$. By letting $m \to \infty$, we reach a contradiction. □

6 Rotation family and relative bounded cohomology

In a group G, the normal closure of an element g is denoted as $\langle\langle g \rangle\rangle$. The goal of this section is as follows:

Proposition 6.1 *Let G be a nonelementary countable group acting properly on a geodesic metric space X with contracting elements. Then for any contracting element $g \in G$, there exists $k = k(g) > 0$ such that the dimension of $H_b^2(G, \langle\langle g^k \rangle\rangle; \mathbb{R})$ as a vector space over $\mathbb{R}$ has the cardinality of the continuum.*

Proof ideas of Proposition 6.1 Since $\langle\langle g^k\rangle\rangle$ is a normal subgroup of G, by Proposition 4.14, we only need to find a hyperbolic space such that the quotient group $G/\langle\langle g^k\rangle\rangle$ acts acylindrically on it. According to the theory of rotating family developed by Dahmani–Guirardel–Osin [14], we need to find a hyperbolic space such that G acts acylindrically on it and $(\mathcal{F} = \{f\,\mathrm{Ax}(g) : f \in G\}, \{f\langle g^k\rangle f^{-1} : f \in G\})$ forms a rotating family. To obtain such a hyperbolic space, we need a construction of quasitrees of spaces $\mathcal{C}(\mathcal{F})$, which can be seen as a blow-up of the projection complex $\mathcal{P}_K(\mathcal{F})$. Theorem 6.9 of [4] shows that $\mathcal{C}(\mathcal{F})$ is a quasitree on which G acts acylindrically and g is a loxodromic element. In order to get a suitable rotating family, we will consider a cone-off space $\dot{Z}_r(\mathcal{F})$ with apexes $\mathcal{F}$ over a scaled metric space $(\mathcal{C}(\mathcal{F}), l\cdot d_{\mathcal{C}})$. For $r \gg 0$, $\dot{Z}_r(\mathcal{F})$ is also hyperbolic [14, Corollary 5.39]. Moreover, [32, Lemma 5.3] shows that $(\mathcal{F}, \{f\langle g^k\rangle f^{-1} : f \in G\})$ is a suitable rotating family on $\dot{Z}_r(\mathcal{F})$. As a result of [14, Proposition 5.28], we get that $\dot{Z}_r(\mathcal{F})/\langle\langle g^k\rangle\rangle$ is still hyperbolic. Finally, we verify that both the extended action $G \curvearrowright \dot{Z}_r(\mathcal{F})$ and the quotient action $G/\langle\langle g^k\rangle\rangle \curvearrowright \dot{Z}_r(\mathcal{F})/\langle\langle g^k\rangle\rangle$ are acylindrical.

6.1 Quasitrees of spaces

Fix a contracting element $g \in G$. We define $\mathcal{F} = \{f\,\mathrm{Ax}(g) : f \in G\}$. Section 2.3 gives a projection complex $\mathcal{P}_K(\mathcal{F})$ whose vertex set is $\mathcal{F}$ and two vertices $U, V \in \mathcal{F}$ are connected by an edge if and only if $\mathcal{F}_K(U,V) := \{W \in \mathcal{F} : d_W(U,V) > K\} = \varnothing$. Fix a positive number L such that $1/2K \le L \le 2K$. We now define a blowup version, $\mathcal{C}(\mathcal{F})$, of the projection complex $\mathcal{P}_K(\mathcal{F})$ by preserving the geometry of each $U \in \mathcal{F}$. Namely, we replace each $U \in \mathcal{F}$, a vertex in $\mathcal{P}_K(\mathcal{F})$, with the corresponding subspace $U \subset X$, while maintaining the adjacency relation in $\mathcal{P}_K(\mathcal{F})$: if U and V are adjacent in $\mathcal{P}_K(\mathcal{F})$ (i.e., $d_{\mathcal{P}}(U,V) = 1$), then we attach an edge of length L from every point $u \in \pi_U(V)$ to $v \in \pi_V(U)$. This choice of L, as stated in [3, Lemma 4.2], ensures that $U \subset X$ is geodesically embedded in $\mathcal{C}(\mathcal{F})$ (so the index L is omitted here).

For any contracting element $g \in G$, the infinite cyclic subgroup $\langle g\rangle$ is of finite index in $E(g)$ by [48, Lemma 2.11], so $\mathrm{Ax}(g) = E(g)o$ is quasi-isometric to a line $\mathbb{R}$. Thus, the set $\mathcal{F}$ (derived from Lemma 2.15) consists of uniform quasilines. By [3, Theorem B], we have the following:

Theorem 6.2 [3] *The quasitree of spaces $\mathcal{C}(\mathcal{F})$ is a quasitree of infinite diameter, with each $U \in \mathcal{F}$ totally geodesically embedded into $\mathcal{C}(\mathcal{F})$. Moreover, the shortest projection from U to V in $\mathcal{C}(\mathcal{F})$ agrees with the projection $\pi_U(V)$ up to a uniform finite Hausdorff distance.*

Any two points $u \in U$ and $v \in V$ in $\mathcal{C}(\mathcal{F})$ are connected via a standard path obtained from the standard path α between U and V in $\mathcal{P}_K(\mathcal{F})$, which passes through each vertex space U on α via a geodesic in U (see [3, Definition 4.3] for more details). Hence, standard paths in $\mathcal{C}(\mathcal{F})$ are also uniform quasigeodesics.

6.2 Hyperbolic cone-off and rotation family

We first introduce a construction of a hyperbolic metric space by conning off a collection of Morse subsets from a given hyperbolic space.

Let Z be a hyperbolic space with a collection $\mathcal{F}$ of uniformly Morse subsets, which means that these Morse subsets have a uniform Morse gauge (see Definition 3.1). Assume that $\mathcal{F}$ has bounded intersection (see Definition 2.6). For $r \geq 0$, we first define the hyperbolic cone-off $\dot{Z}_r(\mathcal{F})$ of Z along $\mathcal{F}$.

For each $U \in \mathcal{F}$, the *hyperbolic cone* $C_r(U)$ is the quotient space of the product $U \times [0, r]$ by collapsing $U \times 0$. The collapsed point denoted by $a(U)$ is called the *apex* of the cone, and $U \times 1$ the *base* of the cone. The cone is equipped with a geodesic metric such that it is the metric completion of the universal covering of a closed hyperbolic disk punctured at the origin.

The *hyperbolic cone-off* $\dot{Z}_r(\mathcal{F})$ is the quotient space of the disjoint union

$$Z \sqcup \coprod_{U \in \mathcal{F}} C_r(U)$$

by gluing U with the base of the cone $C_r(U)$, equipped with the length metric. Since $\mathcal{F}$ has bounded intersection, for $r \gg 0$, $\dot{Z}_r(\mathcal{F})$ is also hyperbolic [14, Corollary 5.39].

Assume that G acts isometrically on Z and leaves $\mathcal{F}$ invariant. The action naturally extends by isometry to the hyperbolic cone $C_r(U)$ by the rule $g(x, t) = (gx, t)$ for any $g \in G$, $x \in U$, $0 \leq t \leq r$. This is a prototype of the notion of a rotating family introduced in [14].

Definition 6.3 Assume G acts isometrically on a metric space $\dot{Z}$. Let A be a G-invariant set in $\dot{Z}$ and a collection of subgroups $\{G_a : a \in A\}$ of G such that $G_a(a) = a$, $gG_ag^{-1} = G_{ga}$ for any $a \in A$, $g \in G$. We call such a pair $(A, \{G_a : a \in A\})$ a *rotating family*.

Returning to the above cone-off construction, the apexes $A(\mathcal{F}) = \{a(U) : U \in \mathcal{F}\}$ and the stabilizers G_a for $a \in A(\mathcal{F})$ together consist of a rotating family. Moreover, we say that A is *ρ-separated* if any two distinct apexes are at distance at least ρ.

Roughly speaking, a rotating family $(A, \{G_a : a \in A\})$ is called *very rotating* if every nontrivial element in G_a rotates around a with a very large angle. This big angle is usually achieved by taking a sufficiently deep subgroup (which is generated by a higher power of some element) of G_a.

6.3 Proof of Proposition 6.1

From now on, we suppose that G is a nonelementary countable group acting properly on a geodesic metric space X with contracting elements. Fix a base point $o \in X$ and a contracting element $g \in G$. Define $\mathcal{F} = \{f \operatorname{Ax}(g) : f \in G\}$. Theorem 6.2 produces a quasitree of space $\mathcal{C}(\mathcal{F})$ in which each G-translate of $\operatorname{Ax}(g)$ is totally geodesically embedded.

Lemma 6.4 [4, Theorem 6.9] *For $K \gg 0$, $\mathcal{C}(\mathcal{F})$ is a quasitree on which G acts acylindrically and g is a loxodromic element on $\mathcal{C}(\mathcal{F})$.*

Denote by $d_{\mathcal{C}}$ the metric on $\mathcal{C}(\mathcal{F})$. The following result gives a way to produce a very rotating family on some cone-off of a "scaled" quasitree of spaces. Here, a scaled metric means a constant multiple of the original metric. By scaling the metric of a hyperbolic space, one can require the hyperbolicity constant to be uniform.

Lemma 6.5 [32, Lemma 5.3] *There exist universal constants $\delta_U > 0$, $r > 20\delta_U$ and $k = k(g), l = l(g) > 0$ with the following property. Consider the cone-off space $\dot{Z}_r(\mathcal{F})$ with apexes $A(\mathcal{F})$ over the scaled metric space $Z_l = (\mathcal{C}(\mathcal{F}), l \cdot d_{\mathcal{C}})$. For every $n \geq 1$, set*

$$E_n = \{f\langle g^{nk}\rangle f^{-1} : f \in G\}.$$

Then $(A(\mathcal{F}), E_n)$ is a $2r$-separated very rotating family on the δ_U-hyperbolic space $\dot{Z}_r(\mathcal{F})$.

As G acts acylindrically on $(\mathcal{C}(\mathcal{F}), d_{\mathcal{C}})$, it is straightforward to verify that G also acts acylindrically on $(\mathcal{C}(\mathcal{F}), l \cdot d_{\mathcal{C}})$. Moreover, [14, Proposition 5.40] shows that acylindricity is preserved by taking (suitable) cone-off, thus one has:

Lemma 6.6 *The extended action $G \curvearrowright \dot{Z}_r(\mathcal{F})$ is acylindrical.*

Fix $r > 10^{10}\delta_U$. As [14, Proposition 5.33] also shows that acylindricity is preserved by taking quotient of a normal subgroup generated by a $2r$-separated very rotating family, one has:

Lemma 6.7 *The quotient action $G/\langle\langle g^k\rangle\rangle \curvearrowright \dot{Z}_r(\mathcal{F})/\langle\langle g^k\rangle\rangle$ is acylindrical.*

Proof In order to apply [14, Proposition 5.33] to get the conclusion, we need to verify that there exists $K > 0$ such that for all $a \in A(\mathcal{F})$ and for all x with $|x - a| = 50\delta_U$,

$$\#\{h \in G : h(a) = a, |x - h(x)| \leq 10\delta_U\} \leq K.$$

From the construction of $A(\mathcal{F})$ above, the stabilizer of each apex in $A(\mathcal{F})$ is exactly a conjugate of $E(g)$. As $[E(g) : \langle g\rangle]$ is finite, there exists $K = K(g)$ with the desired property. □

Moreover, the quotient space $\dot{Z}_r(\mathcal{F})/\langle\langle g^k\rangle\rangle$ is $60000\delta_U$-hyperbolic by [14, Proposition 5.28].

Proof of Proposition 6.1 As Lemma 6.7 implies that the quotient group acts WPD on a hyperbolic space, the conclusion follows from Proposition 4.14. □

We conclude this section by posing the following question:

Question 6.8 *Let G be a nonelementary countable group acting properly on a geodesic metric space X with contracting elements. Let H be a normal subgroup of G with a nonamenable quotient. Is the dimension of $H_b^2(G, H; \mathbb{R})$ as a vector space over $\mathbb{R}$ infinite?*

Acknowledgements

We are grateful to Prof. Wenyuan Yang and Prof. Shi Wang for many helpful suggestions on the first draft. Many thanks to Bingxue Tao for pointing out Proposition 1.6 for us. We also thank the anonymous referees for numerous and useful comments. Wan is supported by NSFC 12471065 & 12326601 and in part by Science and Technology Commission of Shanghai Municipality (22DZ2229014). Both authors are supported by National Key R&D Program of China (SQ2020YFA070059) and NSFC (12131009).

References

[1] **G N Arzhantseva**, **C H Cashen**, **D Gruber**, **D Hume**, *Negative curvature in graphical small cancellation groups*, Groups Geom. Dyn. 13:2 (2019) 579–632 MR

[2] **W Ballmann**, *Lectures on spaces of nonpositive curvature*, DMV Seminar 25, Birkhäuser, Basel (1995) MR

[3] **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

[4] **M Bestvina**, **K Bromberg**, **K Fujiwara**, **A Sisto**, *Acylindrical actions on projection complexes*, Enseign. Math. 65:1-2 (2019) 1–32 MR

[5] **M Bestvina**, **K Fujiwara**, *Bounded cohomology of subgroups of mapping class groups*, Geom. Topol. 6 (2002) 69–89 MR

[6] **M Bestvina**, **K Fujiwara**, *A characterization of higher rank symmetric spaces via bounded cohomology*, Geom. Funct. Anal. 19:1 (2009) 11–40 MR

[7] **A Bouarich**, *Suites exactes en cohomologie bornée réelle des groupes discrets*, C. R. Acad. Sci. Paris Sér. I Math. 320:11 (1995) 1355–1359 MR

[8] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR

[9] **R Brooks**, *Some remarks on bounded cohomology*, from "Riemann surfaces and related topics" (Stony Brook, NY, 1978) (I Kra, B Maskit, editors), Ann. of Math. Stud. 97, Princeton Univ. Press (1981) 53–63 MR

[10] **K S Brown**, *Cohomology of groups*, Graduate Texts in Mathematics 87, Springer (1994) MR

[11] **M Bucher**, **M Burger**, **R Frigerio**, **A Iozzi**, **C Pagliantini**, **M B Pozzetti**, *Isometric embeddings in bounded cohomology*, J. Topol. Anal. 6:1 (2014) 1–25 MR

[12] **D Calegari**, *scl*, MSJ Memoirs 20, Mathematical Society of Japan, Tokyo (2009) MR

[13] **M Coornaert**, **T Delzant**, **A Papadopoulos**, *Géométrie et théorie des groupes*: *les groupes hyperboliques de Gromov*, Lecture Notes in Math. 1441, Springer (1990) MR

[14] **F Dahmani**, **V Guirardel**, **D Osin**, *Hyperbolically embedded subgroups and rotating families in groups acting on hyperbolic spaces*, Mem. Amer. Math. Soc. 1156, Amer. Math. Soc. (2017) MR

[15] **T Delzant**, *Sous-groupes distingués et quotients des groupes hyperboliques*, Duke Math. J. 83:3 (1996) 661–682 MR

[16] **C Druţu**, **M Kapovich**, *Geometric group theory*, American Mathematical Society Colloquium Publications 63, Amer. Math. Soc., Providence, RI (2018) MR

[17] **C Druţu**, **M Sapir**, *Tree-graded spaces and asymptotic cones of groups*, Topology 44:5 (2005) 959–1058 MR

[18] **D B A Epstein**, **K Fujiwara**, *The second bounded cohomology of word-hyperbolic groups*, Topology 36:6 (1997) 1275–1289 MR

[19] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Mathematical Series 49, Princeton Univ. Press (2012) MR

[20] **A Fathi**, **F Laudenbach**, **V Poenaru**, *Travaux de Thurston sur les surfaces*: *Séminaire Orsay*, Astérisque 66-67, Soc. Math. France, Paris (1979) MR

[21] **F Franceschini**, *A characterization of relatively hyperbolic groups via bounded cohomology*, Groups Geom. Dyn. 12:3 (2018) 919–960 MR

[22] **R Frigerio**, *Bounded cohomology of discrete groups*, Mathematical Surveys and Monographs 227, Amer. Math. Soc., Providence, RI (2017) MR

[23] **R Frigerio**, **M B Pozzetti**, **A Sisto**, *Extending higher-dimensional quasi-cocycles*, J. Topol. 8:4 (2015) 1123–1155 MR

[24] **K Fujiwara**, *The second bounded cohomology of a group acting on a Gromov-hyperbolic space*, Proc. London Math. Soc. (3) 76:1 (1998) 70–94 MR

[25] **K Fujiwara**, *The second bounded cohomology of an amalgamated free product of groups*, Trans. Amer. Math. Soc. 352:3 (2000) 1113–1129 MR

[26] **V Gerasimov**, **L Potyagailo**, *Quasiconvexity in relatively hyperbolic groups*, J. Reine Angew. Math. 710 (2016) 95–135 MR

[27] **É Ghys**, **P de la Harpe** (editors), *Sur les groupes hyperboliques d'après Mikhael Gromov*, Progr. Math. 83, Birkhäuser, Boston, MA (1990) MR

[28] **M Gromov**, *Volume and bounded cohomology*, Inst. Hautes Études Sci. Publ. Math. 56 (1982) 5–99 MR

[29] **M Gromov**, *Hyperbolic groups*, from "Essays in group theory" (S M Gersten, editor), Math. Sci. Res. Inst. Publ. 8, Springer (1987) 75–263 MR

[30] **U Hamenstädt**, *Bounded cohomology and isometry groups of hyperbolic spaces*, J. Eur. Math. Soc. 10:2 (2008) 315–349 MR

[31] **S Han**, **W Yang**, **Y Zou**, *Counting double cosets with application to generic* 3*-manifolds*, preprint (2023) arXiv 2307.06169

[32] **Z He**, **J Liu**, **W Yang**, *Large quotients of group actions with a contracting element*, from "Proceedings of the International Consortium of Chinese Mathematicians 2017" (S Y Cheng, S-T Yau, L Ji, X-P Zhu, editors), Int. Press, Boston, MA (2020) 319–338 MR

[33] **M Hull**, **D Osin**, *Induced quasicocycles on groups with hyperbolically embedded subgroups*, Algebr. Geom. Topol. 13:5 (2013) 2635–2665 MR

[34] **N V Ivanov**, *Foundations of the theory of bounded cohomology*, Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) 143 (1985) 69–109 MR In Russian; translated in J. Soviet Math. 37 (1987) 1090–1115

[35] **S Kim**, **T Kuessner**, *Simplicial volume of compact manifolds with amenable boundary*, J. Topol. Anal. 7:1 (2015) 23–46 MR

[36] **D Kotschick**, *Quasi-homomorphisms and stable lengths in mapping class groups*, Proc. Amer. Math. Soc. 132:11 (2004) 3167–3175 MR

[37] **Y N Minsky**, *Quasi-projections in Teichmüller space*, J. Reine Angew. Math. 473 (1996) 121–136 MR

[38] **N Monod**, *Continuous bounded cohomology of locally compact groups*, Lecture Notes in Mathematics 1758, Springer (2001) MR

[39] **B H Neumann**, *Groups covered by permutable subsets*, J. London Math. Soc. 29 (1954) 236–248 MR

[40] **D V Osin**, *Relatively hyperbolic groups: intrinsic geometry, algebraic properties, and algorithmic problems*, Mem. Amer. Math. Soc. 843, Amer. Math. Soc. (2006) MR

[41] **D Osin**, *Acylindrically hyperbolic groups*, Trans. Amer. Math. Soc. 368:2 (2016) 851–888 MR

[42] **C Pagliantini**, **P Rolli**, *Relative second bounded cohomology of free groups*, Geom. Dedicata 175 (2015) 267–280 MR

[43] **H Park**, *Relative bounded cohomology*, Topology Appl. 131:3 (2003) 203–234 MR

[44] **P Przytycki**, **A Sisto**, *A note on acylindrical hyperbolicity of mapping class groups*, from "Hyperbolic geometry and geometric group theory" (K Fujiwara, S Kojima, K Ohshika, editors), Adv. Stud. Pure Math. 73, Math. Soc. Japan, Tokyo (2017) 255–264 MR

[45] **A Sisto**, *Contracting elements and random walks*, J. Reine Angew. Math. 742 (2018) 79–114 MR

[46] **R Wan**, **X Xu**, **W Yang**, *Marked length spectrum rigidity in groups with contracting elements*, preprint (2024) arXiv 2402.10165

[47] **W-y Yang**, *Growth tightness for groups with contracting elements*, Math. Proc. Cambridge Philos. Soc. 157:2 (2014) 297–319 MR

[48] **W-y Yang**, *Statistically convex-cocompact actions of groups with contracting elements*, Int. Math. Res. Not. 2019:23 (2019) 7259–7323 MR

[49] **W Yang**, *Conformal dynamics at infinity for groups with contracting elements*, preprint (2022) arXiv 2208.04861

ZHENGUO HUANGFU huangfuzhg@shanghaitech.edu.cn
Institute of Mathematical Sciences, ShanghaiTech University, Shanghai, China

RENXING WAN rxwan@math.ecnu.edu.cn
School of Mathematical Sciences, East China Normal University, Shanghai, China

Received: October 13, 2024 Revised: March 6, 2025

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. 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 `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 26 Issue 3 (pages 825–1227) 2026