Asymptotic Geometry, Bounded Generation and Subgroups of Mapping Class Groups

Chun-yi Sun

Abstract

In this paper we will discuss a geometric proof of the non-bounded generation of subgroups of the mapping class group with exponential growth. We also provide a proof that any list of irredundant non-elliptic mapping classes with large enough powers generate a right-angled Artin group which is isometrically embedded in the mapping class group.

Let $S = S_{g,p}$ be a surface of finite type with complexity $\xi(S) = 3g + p - 3 > 1$. The mapping class group $\mathcal{MCG}(S)$ consists of isotopy classes of orientation preserving homeomorphisms of S.

A group G is said to be **boundedly generated** if there is a tuple $(g_i \in G)_1^k$ such that for any $g \in G$, there exist $(n_i) \in \mathbb{Z}^k$ such that $g = g_1^{n_1} g_2^{n_2} \cdots g_k^{n_k}$ (in the given order). In other words, $G = \langle g_1 \rangle \langle g_2 \rangle \cdots \langle g_k \rangle$ where products of groups $\Gamma_1 \Gamma_2 \cdots \Gamma_k = \{\gamma_1 \gamma_2 \cdots \gamma_k | \gamma_i \in \gamma_i\}$. We say G is bounded generated by subgroups $\Gamma_1, \ldots, \Gamma_k$ if $G = \Gamma_1 \cdots \Gamma_k$. It is easy to see that if a group G is boundedly generated, it is finitely generated, and that all abelian groups of finite rank are boundedly generated.

Farb, Lubotzky and Minsky proved $\mathcal{MCG}(S)$ is not boundedly generated through pro-p completion of some finite-index subgroup [FLM01]. In this paper we will give a geometric approach to this problem and show that all subgroups of exponential growth of $\mathcal{MCG}(S)$ are not boundedly generated with a growth argument.

Theorem 4.1. *If $G < \mathcal{MCG}(S)$ has exponential growth, G is not boundedly generated.*

By the Tits alternative for subgroups of mapping class groups, a subgroup of mapping class groups either is virtually abelian or has exponential growth [Iva84], [McC85]. Combine this and Theorem 4.1, we deduce:

Theorem 0.1. *A subgroup G of a mapping class group is boundedly generated if and only if G is virtually abelian.*

$\mathcal{MCG}(S)$ is finitely generated [Deh87] and we will fix a finite generating set and define word length $\|\cdot\|$ with this generating set. We focus our study of $\mathcal{MCG}(S)$ on its action on the curve complexes $\mathcal{C}(S)$ and $\mathcal{C}(Y)$ where Y is a subsurface of S. The complex of curves $\mathcal{C}(S)$, or the curve complex [Har81], is a finite dimensional simplicial complex associated to S. The vertices of $\mathcal{C}(S)$ are isotopy classes of essential non-peripheral simple closed curves on S. The mapping class group $\mathcal{MCG}(S)$ acts on $\mathcal{C}(S)$. We follow the work by Masur and Minsky [MM00] and study $\mathcal{MCG}(S)$ through the geometry and combinatorics of $\mathcal{C}(S)$. They showed in [MM99] that $\mathcal{C}(S)$ has infinite diameter and is δ-hyperbolic. Gromov and Canon extract a property shared by negatively curved spaces such as trees and hyperbolic spaces $\mathbb{H}^n$ and define δ-hyperbolicity. (See Section 1.2 for definition.) Actions of groups on δ-hyperbolic spaces have been studied extensively ever since.

Nielsen and Thurston classified mapping classes into three mutually exclusive types: elliptic, reducible, and pseudo-Anosov. Some powers of elliptic mapping classes are isotopic to the identity map. A reducible mapping class g fixes a collection A of homotopy classes of simple closed curves on S. We call $\mathrm{supp}(g) = S - \bigcup_{\alpha \in A} \alpha$ the support of g. Pseudo-Anosov mapping classes have no fixed points in $\mathcal{C}(S)$ and $\mathrm{supp}(g) = S$. For any torsion-free subgroup $G < \mathcal{MCG}(S)$, we define its support $\mathrm{supp}(G)$ to be the union of the supports of all non-elliptic elements $g \in G$. By adding Dehn twists along the boundary curve for each $\mathrm{supp}(g)$ into the tuple, we can make each g_i in any given tuple $(g_i \in G)_1^k$ fix pointwise the boundary of its support.

The stable length of g acting on a metric space X is

$$l_X(g) = \lim_{n \to \infty} \frac{1}{n} d_X(x, g^n x).$$

The stable length of elliptic and reducible mapping classes acting on $\mathcal{C}(S)$ is 0. Masur and Minsky showed that the stable lengths of pseudo-Anosov mapping classes are positive [MM99].

We may also treat the curve complex of a subsurface as a subcomplex of $\mathcal{C}(S)$. Masur and Minsky defined subsurface projection $\pi_Y : \mathcal{C}(S) \to \mathcal{C}(Y)$ for $Y \subseteq S$ [MM00]. They establish a coarse equality between the word length of any mapping class g and some coarse sum of projection distance in $\mathcal{C}(Y)$ of a simplex μ of $\mathcal{C}(S)$ and $g\mu$ for every $Y \subseteq S$. This sum is coarsely equal to the distance between μ and $g\mu$ in the marking complex $\mathcal{M}$, where μ is maximal dimensional simplex in $\mathcal{C}(S)$. Definitions of markings and marking complexes can be found in Section 1.4. A value A is *coarsely greater than* B if there is $C \geq 1$ and $K \geq 0$ such that

$$CA + K \geq B.$$

Denote $A \succcurlyeq B$. If B is also coarsely greater than A, then we say A and B are coarsely equal.

To show that subgroups of $\mathcal{MCG}(S)$ with exponential growth are not boundedly generated by any given tuple $(g_i)_1^k$, we will estimate the cyclicly reduced word length of words $g_1^{n_1} \cdots g_k^{n_k}$ with a certain property by $\sum |n_i|$. The cyclicly reduced word length $\|w\|_c$ of a word w is the minimum word length among its conjugates $w' = gwg^{-1}$. Here we require that there is no obvious cancellation in the ordered list $(g_i)_1^k$: if supports g_i and g_j are disjoint with support of every g_l where $i < l < j$, then g_i and g_j are not commensurable. This is the **non-cancellation condition** (see Section 3.3.1).

Theorem 0.2 (Weaker version of Theorem 3.4 Good word estimate)**.** *Given non-elliptic $(g_i \in \mathcal{MCG}(S))_1^k$ satisfying the non-cancellation condition, then*

$$\|g_1^{n_1} \cdots g_k^{n_k}\|_c \succcurlyeq \sum_{i=1}^{k} |n_i|$$

provided each $|n_i|$ is sufficiently large. The constants in the coarse inequality depend on $(g_i)_1^k$.

We call such $g_1^{n_1} \cdots g_k^{n_k}$ a *good word*.

Fix a collection of mapping classes $(h_i)_1^j$ which may or may not contain elliptic elements. To complete the proof of Theorem 4.1, we shall show that we may rewrite some conjugates of $h_1^{n_1} \cdots h_j^{n_j}$ as $g_1^{m_1} \cdots g_k^{m_k} \gamma$ for some γ of bounded word length in finite steps so that $(g_i)_1^k$ are non-elliptic and satisfy the non-cancellation condition while each m_i is sufficiently large and $k \leq j$. This rewriting process is of finite time and the bound of $\|\gamma\|$ is determined by (h_i). Thus, we can show that the cardinality of $\{w = h_1^{n_1} \cdots h_j^{n_j} \| \|w\| \leq R\}$ is at most a polynomial of R of degree j. So any subgroup $G < \mathcal{MCG}(S)$ of exponential growth cannot be boundedly generated.

Masur and Minsky showed in [MM00] that pseudo-Anosovs have quasi-axes. A *quasi-axis* of pseudo-Anosov g is a geodesic β in $\mathcal{C}(S)$ such that the Hausdorff distance between β and $g\beta$ is less than 2δ. Bowditch proved that there exists $m = m(S) \in \mathbb{N}$ such that g^m fixes at least one quasi-axis for every pseudo-Anosov g [Bow08]. For every pseudo-Anosov g, without loss of generality, we arbitrarily pick one quasi-axis and named it axis(g). For a reducible mapping class g, we arbitrarily pick a quasi-axis $\mathrm{axis}(g)$ in $\mathcal{C}(Y)$ where Y is the support $\mathrm{supp}(g)$.

To show Theorem 3.4, we need to dig deeper into the geometry and combinatorics of $\mathcal{C}(S)$. Behrstock inequality 3.1 [Beh06] captures the contraction property of subsurface projections. At the same time, nearest point projections onto geodesics in a δ-hyperbolic space also have similar inequality. We then define projections from $\mathcal{C}(S)$ onto geodesics in $\mathcal{C}(Y)$ for any $Y \subset S$ by first applying the subsurface projection then the nearest point projection. We can relate these two inequalities by the following Lemma of triples.

Lemma 3.2 (Lemma of triples)**.** *Let X, Y, and Z be subdomains of S or S itself. Let α, β and γ be geodesics in $\mathcal{C}(X)$, $\mathcal{C}(Y)$, $\mathcal{C}(Z)$ respectively if X, Y, and Z are not annular. If either X, Y or Z is an annulus, we associate the core curve to that annular domain. Then there exists $L > 0$ depending only on $\xi(S)$ such that at most one of*

$$d_\alpha(\beta,\gamma), d_\beta(\gamma,\alpha), d_\gamma(\alpha,\beta)$$

is greater than L.

This symmetric formulation is inspired by the consistency condition C1 of [BKMM08] while the name of the lemma is from [BBF10]. With some careful induction, for a sequence of "geodesics" supported in each's own curve complex and satisfying some local conditions, we derive Lemma 3.3(Generalized local to global). Suppose $\mathcal{A} = (\alpha \subset \mathcal{C}(Y_\alpha))$ is a discrete partially ordered list of geodesics in complexes of subsurfaces (Y_α). Lemma 3.3 states that if the distances of local projections onto each α are sufficiently large, we can estimate the marking distance between the maxima and minima of $\mathcal{A}$ by summing up all the local projections, hence the name *local to global*. The local projection distance onto α is the distance between $\max\{\beta \prec \alpha \mid \beta \in \mathcal{A}\}$ and $\min\{\alpha \prec \beta \mid \beta \in \mathcal{A}\}$. We put on some restrictions on $(\mathcal{A},\prec)$ so that the local projections onto α is not empty unless α is either maximum or minimum in $\mathcal{A}$. The coefficients in this estimate depend only on S.

We then construct a bi-infinite sequence $\mathcal{A}$ of axes of non-elliptics, which is fixed by a good word $g_1^{n_1}\cdots g_k^{n_k}$. Finally, we apply Lemma 3.3 to segments of $\mathcal{A}$ and obtain Theorem 3.4 (Good word estimate).

A result on quasi-homomorphisms of $\mathcal{MCG}(S)$ by Bestvina and Fujiwara implies Theorem 4.1 [Fuj09]. A quasi-homomorphism h on a discrete group G is a function $h : G \to \mathbb{R}$ such that $\max|h(\gamma_1\gamma_2) - h(\gamma_1) - h(\gamma_2)| < \infty$.

They showed that there is an infinite-dimensional subspace of the vector space $QH(\mathcal{MCG}(S))$ of quasi-homomorphisms with each nontrivial element bounded on every cyclic group and curve stabilizer [BF02]. This actually implies an extension of Theorem 4.1 [BF07]:

Theorem 4.2. *If $G < \mathcal{MCG}(S)$ has exponential growth and has pseudo-Anosov elements, and subgroups $C_1, \dots C_k < G$ are either cyclic or contained in some curve stabilizers respectively, then*

$$G \neq C_1 \cdots C_k.$$

We give an alternative proof of Theorem 4.2 by projecting the bi-infinite sequence $\mathcal{A}$ of a word $g = h_1 \cdots h_k$ for $h_i \in C_i$ to its axis. Bowditch showed that there are uniformly many mapping classes taking each of any two sufficiently far apart vertices to the respective bounded neighborhood. He named this property *acylindricity* . Acylindricity of the action of $\mathcal{MCG}(S)$ on $\mathcal{C}(S)$ guarantees that if we project $\gamma\beta$ onto $\beta = \text{axis}(g)$ for any pseudo-Anosov g, the projection cannot be of arbitrarily large diameter. (See Lemma 1.11.) If $G < \mathcal{MCG}(S)$ contains pseudo-Anosov and is of exponential growth, then G contains an infinite collection of pseudo-Anosovs that are not commensurable even up to conjugacy. However, by comparing the constructed bi-infinite sequence $\mathcal{A}$ of g and $\text{axis}(g)$, we conclude that $C_1 \cdots C_k = \{h_1 \cdots h_k | h_i \in C_i\}$ do not have an infinite collection of pseudo-Anosovs that are not commensurable up to conjugacy. Therefore, $G \neq C_1 \cdots C_k$.

0.1 Application of the Good Word Estimate

A finitely generated, finitely presented group G is a *right-angled Artin group* if it can be generated by $(x_i)_1^k$ with relations solely in the form of commutators $[x_s, x_t]$. We may associate a finite graph $\Gamma(G)$ to G by assigning each x_i with a vertex V_i and each relation $[x_s, x_t] = x_s x_t x_s^{-1} x_t^{-1}$ with an edge between V_s and V_t. Free groups of finite rank and free abelian groups of finite rank are examples of right-angled Artin groups.

An easy application of Theorem 3.4 generalized the result from [CLM10] and [Kob10]. Given an irredunddant set of non-elliptic mapping classes, then sufficiently large powers of elements in this set generate a right-angled Artin group which is quasi-isometrically embedded into $\mathcal{MCG}(S)$. A set of mapping classes is *irredundant* if no pair of its elements are commensurable.

We can also extend the result by Dahmani and Guirardel [DG] that for some $m = m(\xi(S))$, the normal subgroup $\langle\langle g^n m \rangle\rangle$ is free.

Theorem 5.2. *Suppose $(g_i)_1^k$ is a collection of non-elliptic elements in which each pair (g_s, g_t) is not commensurable up to conjugacy if $s \neq t$. There exists an integer $m' = m'(\xi(S))$ such that for sufficiently large n the normal subgroup $\langle\langle g_i^{nm'} \rangle\rangle$ has a representation described as follows. The generators are $\{\gamma g_i^{nm'} \gamma^{-1} | 1 \leq i \leq k, \gamma \in \mathcal{MCG}(S)\} / \sim$ where $\gamma g_i^{nm'} \gamma^{-1} \sim \gamma' g_i^{nm'} \gamma'^{-1}$ if they fix the same end points at boundary of $\mathcal{C}(S)$; that is, if $\gamma \gamma'^{-1}$ commute with g_i^l for some $l \neq 0$. The relations are in the form of commutators of the generators.*

0.2 Relation to Superrigidity

Γ is an **irreducible lattice** in a semi-simple Lie group G if Γ is discrete and of finite co-volume, and ΓN is dense for any non-compact closed normal subgroup $N \rhd G$.

Many irreducible lattices of semi-simple Lie group of higher rank are boundedly generated by results of Tavgen, Keller, Carter, Witte Morris, Rapinchuk and Rapinchuk[Tav90], [CK83]. It is conjectured that all non-uniform irreducible lattices in higher rank semi-simple Lie groups are boundedly generated. Igor Rivin kindly mentioned in an email that with Margulis' Normal Subgroup Theorem, we can show:

Theorem 0.3 (Superrigidity with $\mathcal{MCG}(S)$ target, [FM98])**.** *Let Γ be a boundedly generated irreducible lattice in a connected semi-simple Lie group G of $\mathbb{R}$-rank at least two with finite center. Then any homomorphism $\phi : \Gamma \to \mathcal{MCG}(S)$ has finite image.*

Proof. $\phi(\Gamma) < \mathcal{MCG}(S)$ is also boundedly generated, hence virtually abelian by Theorem 4.1. By Margulis' Normal Subgroup Theorem, $\ker(\phi) \lhd \Gamma$ is either finite or of finite index. $\phi(\Gamma)$ is isomorphic to $\Gamma / \ker(\phi)$.

If $\ker(\phi)$ is finite, then Γ is virtually abelian. Since G is semi-simple, this contradicts with the assumption that Γ is irreducible.

If $\ker(\phi) \lhd \Gamma$ is of finite index, then $\phi(\Gamma) \cong \Gamma / \ker(\phi)$ is finite. $\square$

Outline

In Section 1, we will go over some basic properties of δ-hyperbolicity, mapping class groups and curve complexes. We discuss some implications of acylindricity.

In Section 2, we demonstrate the motivation from Klenina groups and hyperbolic space $\mathbb{H}^3$.

In Section 3, we first demonstrate simple cases of Lemma 3.4. Then we prove Lemma 3.2(Lemma of triples) by combining Behrstock inequality and contraction property of nearest point projection. Lemma 3.3 (Generalized local to global) and Theorem 3.4 (Good word estimate) are stated and proved in this chapter.

In Section 4, we prove Theorem 4.1 by a growth argument. We give an alternative proof of Theorem 4.2 with nearest point projections onto axes of pseudo-Anosovs.

In Section 5, we apply Theorem 3.4 to extend a result in [CLM10] and [Kob10]. We further show that normal subgroup generated by a good collection of mapping classes has nice representation as a right-angled Artin group except that it is not finitely generated.

1 Background

In this section we review quickly δ-hyperbolicity, mapping class groups and curve complexes. We show some facts of δ-hyperbolicity and restate acylindricity in terms of nearest point projections.

1.1 Geometric Group Theory and Large Scale Geometry

1.1.1 Quasi-isometry

Suppose $f : X \to Y$ is a map between metric spaces. We say f is a quasi-isometry if there are $K \geq 0$ and $C \geq 1$ such that

1. $$\frac{1}{C}d(x,x') - K \leq d(f(x), f(x')) \leq Cd(x,x') + K$$
   for any $x, x' \in X$.

2. Hausdorff distance $d_{\text{Hausdorff}}(f(X), Y) \leq K$.

In this thesis, coarse geometry, asymptotic geometry, or large-scale geometry mean the studies of properties which are invariant up to quasi-isometry, such as δ-hyperbolicity (see Section 1.2). When distance between two sets is mentioned, we mean the minimum (infimum) distance unless we specifically use Hausdorff distance, which is denoted by $d_{\text{Hausdorff}}$.

1.1.2 Geometric Group Theory

Geometric group theory utilizes geometric methods to solve group theory problems. For example, one can show that subgroups of free groups are free by studying graphs. Ol'shanskii [Ol′84] gave elegant proofs for many new and old results on free groups by looking at the action of a free group (as Deck transformation) on its Cayley graph.

There have been many beautiful results on Kleinian groups, which are finitely generated discrete subgroups of $\mathrm{PSL}(2,\mathbb{C})$, over the past 30 years. Not only do those results have tremendous applications on low-dimensional geometry, quite often they are sources of inspiration for studies on other groups.

If S and S' are two finite generating sets of a group G, then Cayley graphs $Cayley(G;S)$ and $Cayley(G;S')$ are quasi-isometric.

One can use coarse geometry to study properties of groups that can be passed through finite-indexed subgroups. There are many interesting properties of groups of this kind. For examples, if a group G has exponential growth, then so does any of its finite-indexed subgroups. Bounded generation is also an example.

We say that a group is virtually (P) for some property P if there is a finite index subgroup $H < G$ which has this property.

1.2 δ-hyperbolicity

Gromov and Cannon introduced the idea of δ-hyperbolic space (also known as Gromov hyperbolic space, or word hyperbolic space) which captures a common property of hyperbolic spaces $\mathbb{H}^n$ and trees [Gro87], [Can91]. For much of the discussion in this chapter, see [GHV91], [BH99] for references.

First we will define δ-thin triangles: a triangle is δ-thin if any side is contained in the union of the δ-neighborhoods of the other sides. For a geodesic space (X,d), if there is a constant δ such that all geodesic triangles (that is, triangles with geodesic sides) are δ-thin, then we say X is δ-hyperbolic.

We say that two infinite geodesic rays are equivalent if they have finite Hausdorff distance. Define $\partial_\infty X$ to be equivalent classes of infinite geodesic rays in X.

If bi-infinite geodesics α and β have same endpoints at infinity, $d_{\mathrm{Hausdorff}}(\alpha,\beta) \leq 2\delta$.

A (λ,ϵ)-quasi-geodesic in a metric space X is a map from interval $I \subset \mathbb{R}$ to X which is a (λ,ϵ) quasi-isometric embedding. If X is δ-hyperbolic, then (λ,ϵ)-quasi-geodesic triangles are δ'-thin where δ' depends on δ, λ and ϵ.

All of our statements are in the language of coarse (large-scale) geometry. In order to simplify the arguments, we often obscure the differences between quasi-geodesics and geodesics, one geodesic and a finite set of fellow-traveling geodesics with same endpoints (because they have Hausdorff distance less than 2δ), points and bounded sets, etc., as long as the constants are fixed and independent of choices.

1.3 Nearest Point Projections

For geodesic α, we define the *nearest point projection* $\pi_\alpha : X \to \alpha$ which project a point $x \in X$ to the shortest segment on α that contains $\{y \in \alpha | d(x,y) = d(x,\alpha)\}$. The definition is viable because by δ-hyperbolicity, we can easily show $\operatorname{diam} \pi_\alpha(x) \le 4\delta$. We write the projection distance $d_\alpha(x,z) = d(\pi_\alpha(x), \pi_\alpha(y))$.

Denote $[xy]$ to be one of the geodesics linking x and y.

For any geodesics α and β, and $x, y \in \alpha$, and $x', y' \in \beta$, if $\max\{d(x,x'), d(y,y')\} \le d$ and $\min\{d(x,y), d(x',y')\} \ge 2d + 2\delta$, then there are $z \in \alpha$ and $z' \in \beta$, such that $d(z,z') \le 2\delta$. In fact, the mid-segment of length $d(x,y) - 2d - 2\delta$ on $[xy] \subset \alpha$ and the mid-segment of length $d(x',y') - 2d - 2D$ on $[x'y'] \subset \beta$ have Hausdorff distance at most 2δ.

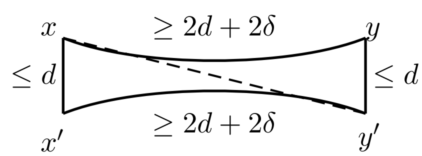

Figure 1: Divide a long 4-gon into 2 triangles.

For any $x \in X$ and $y \in \alpha$, assume $x' \in \pi_\alpha(x)$. Then $[xy]$ passes through the 3δ-neighborhood of x': by hyperbolicity, there is a point z on $[xy]$ which is in the intersection of regular neighborhoods $N_\delta([xx'])$ and $N_\delta([x'y])$. Assume $d(x'', z) \le \delta$ where $x'' \in [xx']$. Then $x' \in \pi_\alpha(x'')$ due to the choice of x'. Then $d(x'', x') \le 2\delta$. Thus $d(z,x') \le d(z,x'') + d(x'',x') \le 3\delta$.

Moreover, $d(x,y) = d(x,z) + d(z,y) \le d(x,x') + d(x',y) - 6\delta$.

Fact 1.1. *If $d_\beta(x,y) > 18\delta$, there are x', y' on $[xy]$ such that 1) $[x'y']$ is in the 2δ-neighborhood of β, and 2) $d(x',y') > d_\beta(x,y) - 18\delta$.*

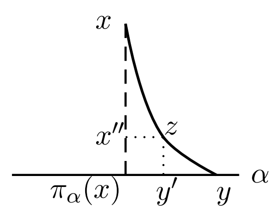

Figure 2: Nearest point projections. $d(\pi_\alpha(x), [xy]) < 3\delta$.

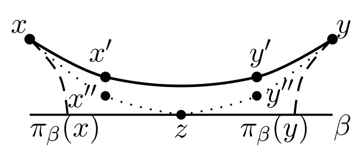

Figure 3: Fact 1.1: If $d_\beta(x, y)$ is large, then a segment of $[xy]$ 2δ fellow travel with a segment of β.

Fact 1.2 (Contraction property)**.** *1) Pick any $x' \in \pi_\alpha(x)$ and $y' \in \pi_\alpha(y)$. If $d(x, x') - d(x, y) \geq 4\delta$, then $d(x', y') \leq 8\delta$. Otherwise, $d(x', y') \leq 3d(x, y) + 8\delta$. 2)Suppose α, β are geodesics (bi-infinite, rays, or segments.) If $d(\alpha, \beta) > 18\delta$, $\mathrm{diam}_\beta(\alpha) < 8\delta$.*

So π_α is coarsely Lipschitz for points close to α, and contract balls away from α to sets of diameter less than 8δ.

Proof. **Fact 1.1** By triangle inequality, $|d(x, x') - d(y, y')| \leq d(x, y)$. From our previous discussion, there is a point z on $[xy']$ which is 3δ-close with x'. Then $d(x, z) \leq d(x, x') + d(x', z) \leq d(x, x') + 3\delta$. If $d(x, y) \leq d(x, z') + \delta$, then z' is in δ-neighborhood of $y'' \in [yy']$. Thus $d(y'', y') \leq d(y'', z) + d(z, x') \leq 4\delta$. Hence, $d(x', y') \leq d(x', z) + d(z, y'') + d(y'', y') \leq 8\delta$.

If $d(x, y) \leq d(x, x') + 4\delta$, then $d(x', y') \leq d(x, x') + d(x, y) + d(y, y') \leq 3d(x, y) + 8\delta$.

Fact 1.2: Pick any $x \in \alpha$ and take $y = \pi_\beta(x) \in \beta$. Suppose $d(\alpha, \beta) > 18\delta$. Then $d(x, y) > 10\delta$. In the argument of Fact 1.1, for any $x' \in \alpha$ and $y' = \pi_\beta(x') \in \beta$, $[x'y']$ intersects the 4δ-neighborhood of y. Thus, because π_β is nearest point projection, $y' = \pi_\beta(x')$ is in the 8δ neighborhood of y. Therefore, $\mathrm{diam}_\beta(\alpha) \leq 8\delta$.

□

If bi-infinite geodesics α and β are of bounded Hausdorff distance, then they 2δ-fellow travel. By definition, it is easy to show that $d(\pi_\alpha(x), \pi_\alpha\pi_\beta(x)) \leq 2\delta$ for any x. Therefore, instead of discussing the nearest point projection onto one geodesic, we will think of π_α as the nearest point projection onto any geodesic which 2δ-fellow travel with α.

1.4 Mapping Class Groups and Curve Complex

Fix a p-punctured genus-g surface $S = S_{g,p}$ where $\xi(S) = 3g + p - 3 \geq 1$.

1.4.1 Mapping Class Groups

Please refer to [FM11] for this subsection.

Fix an essential simple closed curve a on S. Cut along a, twist one side for a full counterclockwise turn, and then glue the two sides back together. This gives us a homeomorphism on S, and we call the isotopy class of such a homeomorphism a **Dehn twist** $T_a \in \mathcal{MCG}(S)$ along a.

Other than Dehn twists, there are also **pseudo-Anosov** elements in $\mathcal{MCG}(S)$. Except at some singularities, locally some powers of the pseudo-Anosov elements behave like Anosov elements (e.g. $\left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$) in $\mathrm{SL}(2, \mathbb{Z})$ acting on $\mathbb{H}^2$: it stretches on one direction and shrinks on the perpendicular direction.

Dehn showed $\mathcal{MCG}(S)$ is finitely generated by Dehn twists.

Nielsen-Thurston classification says all elements in $\mathcal{MCG}(S)$ can only be one of the following:

Elliptic: some representative g of $[g] \in \mathcal{MCG}(S)$ has some power n such that $g^n = \mathrm{Id}$.

Reducible: $[g]$ fixes some isotopy classes of disjoint essential close curves in S. Dehn twists and pseudo-Anosovs supported on proper subsurfaces of S are reducible.

Pseudo-Anosov.

We call the subgroup of $\mathcal{MCG}(S)$ which fixes a simple closed curve a a curve stabilizer. It is generated by all the reducible elements supported on $S - a$ and Dehn twist T_a around a. Here the reducible might permute boundaries a_+ and a_- of $S-a$, which were resulted from cutting along a on S. Indeed, the curve stabilizer of a is a semi-direct product $\mathcal{MCG}(S-a) \ltimes \langle T_a \rangle$.

1.4.2 Free Subgroups and Growth

Tits alternative [Tit72] states that any subgroup of finitely generated linear group either is virtually solvable or contains a free subgroup. [Iva84] and [McC85] exhibited that for pairwise independent pseudo-Anosovs $g_1, \ldots, g_k$, there is an integer $N > 0$ such that $g_1^N, \ldots, g_k^N$ generate a free subgroup. This result gives the "Tits alternative" for mapping class groups: any subgroup of $\mathcal{MCG}(S)$ either is virtually abelian or contains a free subgroup [McC85].

Clay-Leininger-Mangahas and Koberda showed that for non-elliptic elements $\{g_1, \ldots, g_k\}$, for sufficiently large N, $\langle g_1^N, \ldots, g_k^N \rangle$ is a right-angled Artin group [CLM10], [Kob10]. If each pair (g_i, g_j) is not commensurable, then $\{g_i^N\}$ is exactly the generators of $\langle g_1^N, \ldots, g_k^N \rangle$.

A group G is said to have exponential growth with respect to a finite generating set $A = A^{-1}$, if

$$\omega(G, A) = \lim_{n \to \infty} \sqrt[n]{|\{g \in G \| |g||_A \leq n\}|} > 1,$$

where $||\cdot||_A$ is the word length with respect to A. Having exponential growth is independent of finite generating sets.

[AAS07] states that $\mathcal{MCG}(S)$ has uniform exponential growth by surjecting $\mathcal{MCG}(S)$ into linear group and applying results from [EMO05].

1.4.3 Curve Complexes

If $\xi(S) = 3g + p - 3 > 1$, the curve complex $\mathcal{C}(S)$ is a 1-complex in which vertices are homotopy classes of essential simple closed curves in S and there is an edge between any two disjoint curves.

If $S = S_{1,1}$ is once-punctured torus, there is an edge between two essential simple closed curves if their minimal intersection number is 1.

If $S = S_{0,4}$ is a four-holed sphere, there is an edge between two essential simple closed curves if their minimal intersection number is 2.

When $S = S_{0,3}$, which we call either a thrice-punctured sphere or a pair of pants, $\mathcal{C}(S)$ is empty; there is no essential simple closed curve on $S_{0,3}$.

Even though the annulus $S_{0,2}$ is not negatively curved, for later discussions, we will define the curve complex $\mathcal{C}(Y)$ when the core curve a of annulus $Y \subset S$ is an essential simple closed curve on S. We lift S to the annular cover $\tilde{Y}$, to which Y lifts homeomorphically. Take $\hat{Y}$ to be a natural compactification of $\tilde{Y}$ obtained as the usual compactification of the universal cover $\tilde{S} = \mathbb{H}^2$ by the close disk. Define vertices of $\mathcal{C}(Y)$ to be paths linking the two boundary components of $\hat{Y}$ modulo homotopies which fix the endpoints. Put an edge between any two elements of $\mathcal{C}_0(Y)$ which have representatives with disjoint interiors. $\mathcal{C}(Y)$ is quasi-isometric to $\mathbb{Z}$ [MM00]. We also write $\mathcal{C}(a) = \mathcal{C}(Y)$.

Naturally, $\mathcal{C}(S)$ has a metric $d = d_S$ if we assign each edge in $\mathcal{C}(S)$ with length 1.

$\mathcal{C}(S)$ is of infinite diameter if $\xi(S) \neq 0$ [MM99]. It is connected.

Masur and Minsky showed that $\mathcal{C}(S)$ ($\xi(S) \geq 0$) is δ-hyperbolic where δ depends on $\xi(S)$. The boundary at infinity of $\mathcal{C}(S)$ as a δ-hyperbolic space can be identified with laminations [Kla99].

========== Define laminations here! ==========

Bowditch gave an alternative proof and an effective bound on δ [Bow06].

For any two curves $x, y \in \mathcal{C}(S)$, we obtain the fill $F(x,y)$ by taking geodesic representatives of x and y, and gluing disks and annuli which are components of $S - (x \cup y)$ to them. Note that if $x, y \in \mathcal{C}_0(S)$ has $d(x,y) \geq 3$, then $F(x,y) = S$; that is, any essential simple closed curve on S either intersects x or intersects y. In this case, we say that x and y fill S.

For easier discussion, we will fix a hyperbolic metric on S and when we talk about simple closed curves in $\mathcal{C}(S)$, vertices in $\mathcal{C}(Y)$ and core curve of Y when Y is an annulus, and boundary components of other subsurfaces in S, we always take the unique geodesic representatives.

A *subdomain* Y in S is an incompressible, non-peripheral, connected open subsurfaces.

We say a subdomain Y is *nested* in another subdomain Z if $Y \subset Z$ when Y is not an annulus. When Y is an annulus, we say Y is nested in Z if $Y \subset Z$ and also Y is not homotopic to any boundary component of Z. We will keep the notation and denote $Y \subset Z$.

Y and Z *overlap* (essentially) if the intersection of their interiors are not empty and they are not nested.

1.4.4 Subsurface Projections

For a non-annular subdomain $Y \subset S$, we define *subsurface projection* π_Y which sends $x \in \mathcal{C}(S)$ to a simplex in $\mathcal{C}(Y)$. $\pi_Y(x)$ consists of boundary components of $Y - N_\epsilon(x)$ where $N_\epsilon(x)$ is the ϵ-regular neighborhood of $x \cap Y$ for arbitrarily small $\epsilon > 0$. If $x \cap Y = \varnothing$, define $\pi_Y(x) = \varnothing$.

For annular subdomain Y, we lift each vertex in $\mathcal{C}(S)$ to the annular cover $\tilde{Y}$ of S and take the compactification to be its image in $\mathcal{C}(Y)$.

From[MM00], we learn that π_Y is coarsely Lipschitz with constant $K = 3$, $C = 3$.

We define *projection distance* $d_Y : \mathcal{C}(S) \times \mathcal{C}(S) \to \mathbb{N}$ by pre-composing the minimal distance in $\mathcal{C}(Y)$ with $\pi_Y \times \pi_Y$. If Y is an annulus with core curve a, we also write $d_a = d_Y$. Note that numerous papers such as [MM00] on curve complexes we refer to adopt the Hausdorff distance between two subsets. But in the setting of those paper, distance between two subsets is only used when both subsets are simplices, or both subsets are bi-infinite geodesics. In those discussions, the Hausdorff distance and the minimal distance are quasi-isometric. Therefore, we will apply the results from those papers freely assuming proper adjustments of constants.

The following statement can be seen as an analogue of Fact 1.2 for nearest point projections onto geodesics.

Theorem 1.3 ([MM00]Bounded geodesic image)**.** *Let Y be a proper subdomain of Z with $\xi(Y) \neq 0$ and let g be a geodesic segment, ray, or bi-infinite line in $\mathcal{C}(Z)$, such that $\pi_Y(v) \neq \varnothing$ for every vertex v of g. There is a constant M depending only on $\xi(Z)$ so that*

$$\operatorname{diam} \pi_Y(g) \leq M.$$

This theorem says that if $a, b \in \mathcal{C}(Z)$ intersect Y and have large projection distance on Y, then any geodesic $[a, b]$ has to pass through the 1-neighborhood of $\mathcal{C}(Y)$ or ∂Y in $\mathcal{C}(Z)$.

The following is just rephrasing of some of the facts from our discussion of δ-hyperbolicity. Consider any subsurface $Y \subset S$ and a geodesic $\alpha \in \mathcal{C}(Y)$.

Lemma 1.4. *For any geodesic $\alpha \subset \mathcal{C}(Y)$, if $d_\alpha(x, y) > 18\delta$,*

$$d_\alpha(x, y) - 8\delta \leq d_Y(x, y).$$

Therefore, for any simplex $A \in \mathcal{C}(S)$,

$$\operatorname{diam} \pi_\alpha(A) \leq 2 + 8\delta.$$

1.4.5 Hierarchy

If $\xi(S) = 3g+p-3 \geq 1$, for each point in $\mathcal{C}(S)$, there are infinitely many points in the 1-neighborhood. There are infinitely many geodesics connecting any two points that are far apart.

So [MM00] introduced tight geodesics. A *tight geodesic* h is geodesic with consecutive simplices $\{v_0, v_1, \ldots, v_m\}$ where m is the length of h. $\{v_0, v_1, \ldots, v_m\}$ satisfy the following properties:

1. $d(v_i, v_j) = |i - j|$;
2. $v_i \subset F(v_{i-1}, v_{i+1})$ for all $1 \leq i \leq m$. (Here we adopt the definition by Bowditch in [Bow08].)

They showed that between any two simplexes in $\mathcal{C}(S)$, there exists at least one tight geodesic connecting them.

$\mathcal{MCG}(S)$ acts on $\mathcal{C}(S)$ isometrically but not freely. A pseudo-Anosov element g has quasi-axes in $\mathcal{C}(S)$ ([MM00]) and permutes a finite set of fellow traveling tight geodesics sharing the same endpoints on the boundary of $\mathcal{C}(S)$ ([Bow08]). Since all the quasi-geodesics and geodesics with same endpoints are within bounded Hausdorff distance (only depends on δ and the quasi-geodesic constants) with each other, in this paper we will just pick one of the tight geodesics and call it the axis of g.

A **marking** in S is a system of pairwise disjoint simple closed curves $\mu = (\alpha_i)$ (base curves) which equipped with transversals (t_i) where t_i is a diameter-1 subset of $\mathcal{C}(\alpha_i)$. $\beta \in \mathcal{C}(S)$ is a *clean transversal* of α if the subdomain filled by α and β has complexity 1 (i.e. it is a torus with one puncture or a sphere with 4 punctures.) A marking is **complete** if its base has $\xi(s)$ disjoint curves. A marking is **clean** if each curve α_i has a transversal t_i which is realized by clean transversal $\pi_{\alpha_i}(\beta_i)$ and β_i does not intersect α_j for all $i \neq j$. All complete markings (i.e. it consists of exactly $\xi(S)$ pairwise disjoint simple closed curves) have at least one and at most some finite bounds of clean complete markings that share the same base curves ([MM00]).

Clean markings $\{\alpha_i, \pi_{\alpha_i}(\beta_i)\}$ have two elementary moves: twist and flip. We twist $\{\alpha_i, \pi_{\alpha_i}(\beta_i)\}$ by applying Dehn twist once along one α_i on β_i. $\{\alpha_i, \pi_{\alpha_i}(\beta_i)\}$ is flipped by replacing α_i by β_i and replace the new marking with a clean compatible one ([MM00]). We can then define clean complete marking graph $\mathcal{M}$: the vertices of $\mathcal{M}$ are clean complete markings and two clean complete markings have distance $d_{\tilde{\mathcal{M}}}$ one if they are only differ by one elementary move.

A **hierarchy** H between two markings $\mathbf{I}$ and $\mathbf{T}$ has a base tight geodesic g_H in $\mathcal{C}(S)$, and a collection of tight geodesics $\{h \subset \mathcal{C}(Y_h)\}$ where $Y_h = \operatorname{supp}(h) \Subset S$. Each Y_h is a component of $Y_f - v$ for some tight geodesic f in H and a vertex v on f. [MM00] showed that between any two clean complete markings, there exists a unique complete hierarchy H between them.

We will use the following theorem to sum up the long segment lengths in different subsurfaces in Lemma 3.3 (Generalized local to global).

Theorem 1.5 (Move distance and projections [MM00].). *There is a constant $M_6(S)$ such that, given $M \geq M_6$, there are e_0, e_1 for which, if μ and ν are any two complete clean markings then*

$$e_0^{-1} d_{\tilde{\mathcal{M}}}(\mu,\nu) - e_1 \leq \sum_{\substack{Y \subseteq S \\ d_Y(\mu,\nu) \geq M}} d_Y(\mu,\nu) \leq e_0 d_{\tilde{\mathcal{M}}}(\mu,\nu) + e_1.$$

Lemma (Large link[MM00]) guarantees the summation in the inequality in Theorem 1.5 is finite.

Lemma 1.6 (Large link[MM00]). *There exist constants M_1, M_2 depending only on S such that, for any hierarchy H and domain Y in S,*

$$\mathrm{diam}_Y(\mathbf{I}(H), \mathbf{T}(H)) > M_2,$$

then Y is the support of a geodesic h in H.

Conversely if $h \in H$ is any geodesic with $Y = D(h)$,

$$\Big| |h| - d_Y(\mathbf{I}(H), \mathbf{T}(H)) \Big| \leq 2M_1.$$

In [Beh06], Behrstock uses Lemma (Order and projections. [MM00]) to show Lemma 3.1 (Projection estimates; Behrstock inequality [Beh06]).

Lemma 1.7 (Order and projections. [MM00]). *Let H be a hierarchy and $h, \quad k \in H$ with $D(h) = Y$ and $D(k) = Z$ and $Y \pitchfork Z$. Then if $h \prec_t k$ then*

$$d_Y(\partial Z, \mathbf{T}(H)) \leq M_1 + 2$$

and

$$d_Z(\mathbf{I}(H), \partial Y) \leq M_1 + 2.$$

1.4.6 $\mathcal{MCG}(S)$ acting on $\mathcal{C}(S)$

Here we adopt the convention that $\mathcal{MCG}(S)$ acts on $\mathcal{C}(S)$ from the left and $(gh)x = g(hx)$ for any $g, h \in \mathcal{MCG}(S)$ and $x \in \mathcal{C}(S)$.

In Bers' proof of Nielsen-Thurston classification, he considers action of mapping classes on Teich(S) and discusses the translation length $\tau(g) = \inf_{x \in \mathrm{Teich}(S)} d(x, gx)$. Unfortunately, either translation length or stable length of mapping classes acting on $\mathcal{C}(S)$ cannot distinguish the elliptic and the reducible elements. Nonetheless, the actions of different mapping classes on $\mathcal{C}(S)$ have very different behavior. We will discuss them in this section.

For any group G acting on a metric space (X, d), the stable length

$$l_X(h) = \lim_{n\to\infty} \frac{d(x, h^n x)}{n}.$$

is well-defined. Because the sequence $\{a_n = d(x, h^n x)\}$ is non-negative and subadditive, by Fekete's Subadditive Lemma [Fek23], $\lim_{n\to\infty} \frac{a_n}{n}$ exists and is equal to $\inf_n \frac{a_n}{n}$. l_X does not depend on choice of base point x since

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} d(x, h^n x) &= \lim_{n\to\infty} \frac{1}{n}\big(d(y, h^n y) - 2d(x, y)\big)\\ &= \lim_{n\to\infty} \frac{1}{n} d(y, h^n y).\end{aligned}$$

l_X is invariant under conjugacy because

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} d(x, gh^n g^{-1} x) &= \lim \frac{1}{n} d(g(g^{-1}x), gh^n(g^{-1}x))\\ &= \lim_{n\to\infty} \frac{1}{n} d(gx, gh^n x),\\ &= \lim_{n\to\infty} \frac{1}{n} d(x, h^n x),\end{aligned}$$

Pseudo-Anosovs

Then the following proposition by Masur and Minsky states that pseudo-Anosovs have positive stable lengths in $\mathcal{C}(S)$; further, the stable lengths of all pseudo-Anosovs are uniformly bounded away from zero.

Proposition 1.8 (Proposition 4.6, [MM99]). *There exists $c > 0$ such that for any pseudo-Anosov $g \in \mathcal{MCG}(S)$, any $x \in \mathcal{C}(S)$ and any $m \in \mathbb{Z}$.*

$$d(g^m x, x) \geq c|m|.$$

This proposition implies $\mathcal{C}(S)$ has infinite diameter. Moreover, pseudo-Anosov elements act on $\mathcal{C}(S)$ as loxodromic elements act on $\mathbb{H}^3$:

Proposition 1.9 (Proposition 7.6 (Axis), [MM00]). *Let h be a pseudo-Anosov element in $\mathcal{MCG}(S)$. There exists a bi-infinite tight geodesic β such that for each j, $h^j(\beta)$ and β are 2δ-fellow travelers. Moreover there exists a hierarchy H with main geodesic β.*

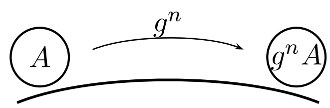

Figure 4: A pseudo-Anosov g moves set A and its projection $\pi_\beta(A)$ along $\beta = \mathrm{axis}(g)$.

Call β a *quasi-axis* of h. Because all the quasi-axes of h share endpoints at infinity, they are all 2δ-fellow travelers. h acts as a permutation among its quasi-axes. Bowditch showed that there exists a universal $m = m(S) > 0$ such that for any pseudo-Anosov $g \in \mathcal{MCG}(S)$, g^m fixes some bi-infinite (tight) geodesic β [Beh06, Theorem 1.4]. Denote $\beta = \mathrm{axis}(g)$.

Dehn Twists and Reducible Elements

A Dehn twist T_a around simple closed curve a fixes a and $\mathcal{C}(S-a)$ point wise. For other

Similarly, any reducible element g supported on $Y = \mathrm{supp}(g)$ fixes ∂Y and $\mathcal{C}(S-Y)$. g also has quasi-axes in $\mathcal{C}(Y)$ and they all 2δ-fellow travel. Take quasi-axis $\beta = \mathrm{axis}(g)$ in $\mathcal{C}(Y)$ and name $\beta = \mathrm{axis}(g)$. Since $\mathcal{MCG}(S)$ act on $\mathcal{C}(S)$ isometrically, $d(g(x), \partial Y) = d(x, \partial Y)$. Therefore, $\lim_{m\to\infty} \frac{1}{m} d(g^m(x), x) = 0$. If simple closed curve $x \in \mathcal{C}(S)$ intersect ∂Y essentially, its orbit $\{g^m\}_{\infty}^{\infty}$ in $\mathcal{C}(S)$ is on an infinite spiral with fixed distance from simple closed curves in $S - Y$ as shown in Figure 1.4.6. The orbit is of constant distance with the quasi-geodesic $\{g^m(\partial Y(x))\}$, which fellow travels with β. We can compare the set of curves $H(x,Y) = \{z \in \mathcal{C}(S) | d(z,y) = d(x,y) \forall y \in \mathcal{C}(Y)\}$ to a horosphere in $\mathbb{H}^3$ and g to a parabolic element in $\mathrm{PSL}(2, \mathcal{C})$.

Elliptic Elements Given an elliptic $g \in \mathcal{MCG}(S)$, take orbifold or manifold $\mathcal{O} = S/\langle g\rangle$. S is a covering of $\mathcal{O}$. For complicated enough $\mathcal{O}$, Teich($\mathcal{O}$) has dimension greater than 1 and contains pseudo-Anosovs on $\mathcal{O}$ that can be lift to pseudo-Anosovs on S. See [Thu80] and [Sco83] for discussions on orbifolds.

Removing open neighborhoods of cone points from $\mathcal{O}$ and define $\mathcal{C}(\mathcal{O})$ to be the curve complex of this surface. Rafi and Schleimer show that the one-to-many map $\Pi : \mathcal{C}(\mathcal{O}) \to \mathcal{C}(S))$ which arises from the covering map is

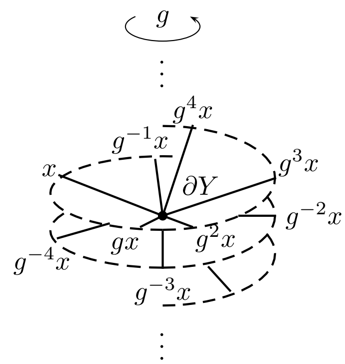

Figure 5: A reducible g moves x along an infinite spiral.

a quasi-isometric embedding with constant depending only on the order of g and $\xi(S)$ [RS09]. The image $\Pi(\mathcal{C}(\mathcal{O}))$ is fixed by $\langle g\rangle$. We can imagine the action of $\langle g\rangle$ on $\mathcal{C}(S)$ as rotation around $\Pi(\mathcal{C}(\mathcal{O}))$.

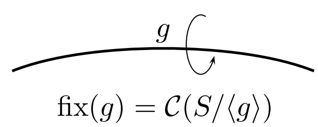

Figure 6: An elliptic g rotate the space around $\mathrm{fix}(g)$.

This completes our comparison between the action of mapping class groups on curve complexes and the action of $\mathrm{PSL}(2,\mathbb{C})$ on $\mathbb{H}^3$.

1.4.7 Acylindricity

$\mathcal{C}(S)$ is not locally finite, and there are infinitely many geodesics between two intersecting curves. However, there are only finitely many tight geodesics between any two curves. Bowditch showed that there are only at most $N(\xi(S))$ of $g \in \mathcal{MCG}(S)$ taking x, and $y \in \mathcal{C}(S)$ to $g(x) \in \mathrm{B}(x,r)$ and

$g(y) \in \mathrm{B}(y, r)$ respectively for any given $r > 0$ if x and y are sufficiently far apart in $\mathcal{C}(S)$ [Bow08]. He named this property **acylindricity**. The weakly properly discontinuous (WPD) property shown by Bestvina and Fujiwara is weaker because it does not require N to be independent of g [BF02].

As a result, Bowditch shows that there is $m = m(\xi(S)) > 0$ such that for every non-elliptic g, g^m fixes bi-infinite geodesic $\operatorname{axis}(g)$ in $\gamma(Y)$ where g is supported on Y. Denote $l(g) = l_S(g)$.

For every γ, consider the orbit of $\operatorname{axis}(g)$, $\{\gamma^n \operatorname{axis}(g)\}$. By acylindricity, if $\{\gamma^n \operatorname{axis}(g)\}$ fellow travel for a long distance, then they actually share endpoints at the infinity, hence of Hausdorff distance 2δ. So γ commutes with some power of g. This property is related to the fact that a non-elementary word hyperbolic group does not contain a Baumslag-Solitar group as its subgroup. The reason is that flats or long and large cylinders cannot be isometrically embedded into a δ-hyperbolic space. With acylindricity, we can translate small angles in $\mathbb{H}^n$ to large projections between axes in $\mathcal{C}(S)$ as shown in the following lemma.

Lemma 1.10. *There exists $N = N(S) > 0$ such that for arbitrary pseudo-Anosovs $g, h \in \mathcal{MCG}(S)$, if $\operatorname{axis}(g)$ and $\operatorname{axis}(h)$ have 2δ-fellow traveling segments longer than $Nm^2 l(g) l(h) + 16\delta$, then g and h are commensurable.*

Proof. Take $r = 12\delta$. By Bowditch's acylindricity, there are $R = R(r) > 0$ and $N_0 = N_0(r) > 0$ such that for any $a, b \in \mathcal{C}(S)$ where $d(a, b) > R$,

$$|\{f \in \mathcal{MCG}(S) | f(a) \in \mathrm{B}(a, r), f(b) \in \mathrm{B}(b, r)\}| \le N_0.$$

We may assume R and N_0 are integers.

Take $N = (N_0 + 2)R$, and denote $g' = g^{Rm^2 l(h)}$ and $h' = h^{Rm^2 l(g)}$. Then $l(g') = l(h')$. Assume $[xx'] \subset \operatorname{axis}(g)$ and $[yy'] \subset \operatorname{axis}(h)$ 2δ-fellow travel, $d(x, x') > Nm^2 l(g) l(h) + 12\delta$ and $d(y, y') > Nm^2 l(g) l(h) + 12\delta$. Then for every $0 \le k \le N_0 + 1$, $[g'^k(x) g'^{k+1}(x)] \subset [xx']$. Pick $z \subset [yy']$ such that $[h'^{-1}(z) z] \subset [yy']$ and $d(g'^{k+1}(x), z) \le 2\delta$. Apply $(h')^{-k}$ on $[h'^{-1}(z) z]$ and $g'^k [x g'(x)]$. Because $l(g') = l(h')$, $d(h'^{-k-1}(z), z) = d(x, g'^{k+1}(x))$. Moreover, because $[xx']$ and $[yy']$ 2δ fellow travel and $d(z, g'^{k+1}(x)) \le 2\delta$, by triangle inequality, $d(\pi_{\operatorname{axis}(h)}(x), h'^{-k-1}(z)) \le 4\delta$ and $d(\pi_{\operatorname{axis}(h)}(g'(x)), h'^{-k}(z)) \le 4\delta$. Therefore,

$$\begin{aligned} d(x, (h')^{-k} g^k(x)) &\le d(x, \pi_\beta(x)) + d(\pi_\beta(x), h'^{-k-1}(z)) + d(h'^{-k-1}(z), h'^{-k-1} g'^k(x)) \\ &\le 2\delta + 4\delta + 6\delta = 12\delta. \end{aligned}$$

Similarly,

$$d(g'(x), (h')^{-k} g^{k+1}(x)) \le 12\delta.$$

By acylindricity,

$$|\{h'^{-k}g'^k | 1 \le k \le N_0 + 1\}| \le N_0.$$

So there are nonzero $l \neq k$ such that $h'^{-l}g'^l = h'^{-k}g'^k$; that is, g and h are commensurable. □

Apply the nearest point projection bound, we can restate the above lemma in terms of projection:

Corollary 1.11. *There exists $N = N(S) > 0$ such that for pseudo-Anosovs g and h with $\beta = \text{axis}(h)$ if $\text{diam}_\beta(\text{axis}(g)) > Nl(g)l(h) + 64\delta$, then g and h are commensurable.*

2 Motivation

We will investigate how to estimate word lengths of $g_1^{n_1} \cdots g_k^{n_k}$ by discussing two special cases when (g_i) satisfy the condition that there is no obvious cancellation and each $|n_i|$ is sufficiently large.

2.1 Case 1. Every g_i is pseudo-Anosov.

Pick an arbitrary marking μ and look at the orbit $\{\mu, g_1^{n_1}\mu, \dots, g_1^{n_1} \cdots g_k^{n_k}\mu\}$. δ-hyperbolicity of $\mathcal{C}(S)$ leads us to the length estimate provided g_i and g_{i+1} are not commensurable and each $|n_i|$ is large. This is a generalization of the following well-known fact in hyperbolic spaces $\mathbb{H}^n$:

Fact 2.1. *Suppose $(\alpha_i)_1^k$ is a piecewise geodesic in X. That is, there are distinct points $x_0, \dots, x_k$ such that α_i is the geodesic linking x_{i-1} and x_i. Denote l_i to be the length of α_i and θ_i to be the (smaller) angle between α_i and α_{i+1} at x_i. If $\theta_i \ge \epsilon$ for some constant $\epsilon > 0$, then there exists $L = L(\epsilon) > 0$ and $K = K(\epsilon) > 0$ such that if each $l_i > L$, $(\alpha_i)_1^k$ is a $(4, K)$-quasi-geodesic. In other words,*

$$d(x_0, x_k) \ge \sum_{i=1}^{k} (l_i - K).$$

Note that choice of L and K does not depend on the number of segments k. The reader can, for instance, find a proof of this fact in [ECH+92].

The condition that the angles are uniformly bounded away from zero in $\mathbb{H}^n$ can be rewritten in terms of nearest point projections. Specifically, diameter of $\pi_{\alpha_i}(\alpha_{i\pm1}$ for each $0 < i < k$ is uniformly bounded. Then one can generalize this fact to any δ-hyperbolic space; in particular, $\mathcal{C}(S)$.

Figure 7: δ-hyperbolicity makes the geodesic $[\mu, g_1^{n_1}g_2^{n_2}g_3^{n_3}\mu]$ 2δ-close to long segments of $[\mu, g_3^{n_3}\mu]$, $[g_3^{n_3}\mu, g_2^{n_2}g_3^{n_3}\mu]$, and $[g_2^{n_2}g_3^{n_3}\mu, g_1^{n_1}g_2^{n_2}g_3^{n_3}\mu]$.

2.2 Case 2. Every g_i is reducible and every pair of supports overlap.

Clay, Leininger and Mangahas deal with the special case that every g_i is reducible and not Dehn twist in [CLM10].

Similar to Case 1, we first observe sequence $\Sigma = \{\mu, g_1^{n_1}\mu, g_1^{n_1}g_2^{n_2}\mu, \dots, g_1^{n_1}\cdots g_k^{n_k}\mu\}$ for marking μ which intersect with all $\mathrm{supp}(g_i)$ essentially. $g_1^{n_1}\cdots g_{i-1}^{n_{i-1}}\mu$ and $g_1^{n_1}\cdots g_i^{n_i}\mu$ will have large distance in $g_1^{n_1}\cdots g_{i-1}^{n_{i-1}}\,\mathrm{supp}(g_i)$ if $|n_i|$ is large. We want to add all these pieces up and give an estimate for $d_{\mathcal{M}}(\mu, g_1^{n_1}\cdots g_k^{n_k}\mu)$ by $\sum|n_i|$. We can achieve this with help of Theorem 1.5 (Move distance and projections [MM00]) if we know the distance of projections of μ and $g_1^{n_1}\cdots g_k^{n_k}\mu$ on $g_1^{n_1}\cdots g_{i-1}^{n_{i-1}}\,\mathrm{supp}(g_i)$ is comparable to $|n_i|$.

Compare μ and $g_1^{n_1}g_2^{n_2}\mu$ in $\mathrm{supp}(g_1)$. Suppose $|n_2|$ is sufficiently large such that in $X_1 = \mathrm{supp}(g_1)$, $\pi_{X_1}(g_2^{n_2}\mu)$ should be close to the projection to X_1 of one of the ending laminations $\lambda_2^\pm$ of g_2 so long as $\mathrm{supp}(g_1) \pitchfork \mathrm{supp}(g_2)$. Thus, if $|n_1|$ is sufficiently large, $d_{X_1}(\mu, g_1^{n_1}\lambda_2^\pm) \geq C_1|n_1|$.

On the other hand, take $X_2 = \mathrm{supp}(g_2)$ and compute

$$d_{g_1^{n_1}X_2}(\mu, g_1^{n_1}g_2^{n_2}\mu) = d_{X_2}(g_1^{-n_1}\mu, g_2^{n_2}\mu).$$

When $|n_1|$ is sufficiently large, in X_2, $g_1^{-n_1}\mu$ is close to the projections of one of the ending laminations $\lambda_1^\pm$. Then we can estimate $d_{g_1^{n_1}X_2}(\mu, g_1^{n_1}g_2^{n_2})$ by $d_(X_2)(\lambda_1^\pm, g_2^{n_2}\mu)$.

We should continue in this fashion until can estimate each $d_{g_1^{n_1}\cdots g_{i-1}^{n_{i-1}}X_i}(\mu, g_1^{n_1}\cdots g_k^{n_k}\mu)$ by $d_{X_i}(\lambda_{i-1}^\pm, g_i^{n_i}\lambda_{i+1}^\pm)$. We start with pairs of markings which are 1 apart in the sequence Σ. Then with the lower bounds in mind, we discuss pairs of markings that are 2 apart in Σ and obtain larger yet finite lower bounds. We go on until we exhaust the sequence. However, we expect this estimate will depend on the choice of μ. To avoid this issue, we may instead consider the boundaries ∂X_i of supports and look at the sequence

$$X_1, g_1^{n_1}X_2, g_1^{n_1}g_2^{n_2}X_3, \ldots, g_1^{n_1}\cdots g_{k-1}^{n_{k-1}}X_k, g_1^{n_1}\cdots g_k^{n_k}X_1.$$

where $X_i = \mathrm{supp}(g_i)$. Then the discussion above still applies to this new sequence.

We will utilize Behrstock inequality with some induction in this case.

Now, if all the g_i's are non-elliptic reducible and $Y_i \pitchfork Y_{i+1}$ for all i. Consider set

$$\Sigma = \{Z_1 = Y_1, Z_2 = g_1^{n_1}Y_2, g_1^{n_1}g_2^{n_2}Y_3, \ldots, Z_k = g_1^{n_1}\cdots g_{k-1}^{n_{k-1}}Y_k, Z_{k+1} = g_1^{n_1}\cdots g_k^{n_k}Y_1\}.$$

We picked these specific subsurfaces because

$$Z_{i+1} = g_1^{n_1}\cdots g_i^{n_i}\,\mathrm{supp}(g_{i+1}) = \mathrm{supp}((g_{i+1}^{n_{i+1}})^{g_1^{n_1}\cdots g_i^{n_i}}).$$

for each i. Then because $Y_i = \mathrm{supp}(g_i)$, for sufficiently large $|n|$,

$$d_{Z_i}(\partial Z_{i-1}, \partial Z_{i+1}) = d_{Y_i}(\partial Y_{i-1}, g_i^n \partial Y_{i+1}) > 4b.$$

By Behrstock inequality, both $d_{Y_{i-1}}(\partial Y_i, g_i^n \partial Y_{i+1})$ and $d_{Y_{i+1}}(\partial Y_i, g_i^{-n}\partial Y_{i-1})$ are less than b. Then we can use induction on the cardinality of $\{Y_i\}$ to show that there is some M which does not depend on k, such that if each $d_{Z_i}(\partial Z_{i-1}, \partial Z_{i+1}) > M$, then $d_{Z_i}(\partial Z_1, \partial Z_{k+1}) \geq d_{Z_i}(\partial Z_{i-1}, \partial Z_{i+1}) - 2b$. We only need to take $M > 4b$ and the rest is just triangle inequality and Behrstock inequality.

The following is due to Brock, Masur and Minsky. Given a set of subsurfaces $\{Y_\alpha\}_{\alpha \in A}$ in S, where A is a partial ordered set. The only requirement for the partial order is that if $Y_\alpha \pitchfork Y_{\alpha'}$, α and α' are ordered. For each $\alpha \in A$, we can define its left immediate marking $\mu^-(\alpha, A)$ to be the set of boundaries of $\{Y_\beta\}$ where $\beta \prec \alpha$ is the maximum among all elements preceding α. Similarly we can define $\mu^+(\alpha, A)$ for each α. We will also define the initial marking $\mu_{\mathbf{I}}$ to be the collection of boundaries of $\{Y_\alpha | \alpha \text{ is minimal in } A\}$ and similarly define the terminal marking $\mu_{\mathbf{T}}$.

Lemma 2.2 (Local to Global [BMM06, Notes from correspondence between Brock, Masur and Minsky])**.** *There exist $m_1 > m_2 > 0$, which depend only on S, such that the following holds: For any indexed family $\{Y_\alpha\}_{\alpha \in A}$ as defined above, if for every internal α,*

$$d_{Y_\alpha}(\mu^-(\alpha, A), \mu^+(\alpha, A)) > m_1$$

then the hierarchy $H(\mu_{\mathbf{I}}, \mu_{\mathbf{T}})$ contains all the domains Y_α for internal α, and moreover,

$$d_{Y_\alpha}(\mu_{\mathbf{I}}, \mu_{\mathbf{T}}) > d_{Y_\alpha}(\mu^-(\alpha, A), \mu^+(\alpha, A)) - m_2.$$

Notice that in the above statement, m_1 and m_2 do not depend on the cardinality of A.

Now we take $A = \Sigma$ and define the partial order on A to be exactly the original order of Σ. Then $\mu^-(i, \Sigma) = \partial Z_{i-1}$ and $\mu^+(i, \Sigma) = \partial Z_{i+1}$ while $\mathbf{I} = \partial Z_1$ and $\mathbf{T} = \partial Z_k = g_1^{n_1} \cdots g_k^{n_k} \partial Z_1$. In order to make $d_{Z_i}(\partial Z_{i-1}, \partial Z_{i+1}) > m_1$, we can find $N_i > 0$ for each $1 < i \leq k$ such that if $|n_i| > N_i$,

$$d_{Z_i}(\partial Z_{i-1}, \partial Z_{i+1}) = d_{Y_i}(\partial Y_{i-1}, g_i^{n_i} \partial Y_{i+1}) > m_1.$$

Because $d_{Y_i}(\partial Y_{i-1}, g_i^{n_i} \partial Y_{i+1}) \sim |n_i|$ with some scalar constant depends on g_i. Then by Lemma 2.2,

$$d_{Z_i}(\partial Z_1, g_1^{n_1} \cdots g_k^{n_k} \partial Z_1) \geq d_{Y_i}(\partial Y_{i-1}, g_i^{n_i} \partial Y_{i+1}) - m_2 \ |n_i| - m_2,$$

for each $1 < i \leq k$. Replace m_1 with a larger constant so that $m_1 - m_2 \geq M_6$ in Theorem 1.5. Then we have

$$d_{\tilde{\mathcal{M}}}(\partial Z_1, g_1^{n_1} \cdots g_k^{n_k} \partial Z_1) \gtrsim e_0^{-1} \sum (|n_i| - m_2) - e_0^{-1} e_1,$$

and each Z_i shows up in the hierarchy connecting ∂Z_1 and $\partial Z_{k+1} = g_1^{n_1} \cdots g_k^{n_k} \partial Z_1$.

2.3 Nested Domains

From the discussion in Case 2, we see that ending laminations play important roles for finding the lower bounds of the powers. Instead of discussing both laminations as separate cases, we should connect them with axes of elements in their supports.

Now we use axes for pseudo-Anosovs in Case 1. and supporting subsurfaces for reducible elements in Case 2. We would like to have a more general statement which combines both . We may still look at the sequence

$$\Sigma = \{Z_1 = Y_1, Z_2 = g_1^{n_1} Y_2, Z_3 = g_1^{n_1} g_2^{n_2} Y_3, \ldots, Z_k = g_1^{n_1} \cdots g_{k-1}^{n_{k-1}} Y_k, Z_{k+1} = g_1^{n_1} \cdots g_k^{n_k} Y_1\}.$$

If $\partial Y_{i\pm1} \pitchfork Y_i$, we can just take $\partial Y_{i\pm1}$ to be the objects and we apply $g_i^{n_i}$ on one and then project both to Y_i. However, if, for example, Y_i is nested in $Y_{i+1} = S$, then the projection of $\partial Y_{i+1} = \partial S$ onto Y_i is empty. We can instead consider ∂Y_{i+2} and $d_{Y_i}(\cdot, g_i^{n_i} g_{i+1}^{n_{i+1}} \partial Y_{i+2})$. If $\partial Y_{i+2} \neq S$, and $|n_{i+1}|$ is sufficiently large, $g_{i+1}^{n_{i+1}} \partial Y_{i+2}$ will be quite close to either the vertical lamination or the horizontal lamination of g_{i+1} in $Y_i \neq S$. In other words, $g_{i+1}^{n_{i+1}} \partial Y_{i+2}$ is quite close to the projection of $\mathrm{axis}(g_{i+1})$ in Y_i. This motivates us to take axes of the elements of the bigger domains and project them into the smaller domains.

Moreover, for non-elliptic g and h with $Y = \mathrm{supp}(g) \notin \mathrm{supp}(h) = Z$, $d_Y(\partial Z, \mathrm{axis}(h))$ is bounded by some constant depending only on S. Along with the fact that π_Y and $\pi_{\mathrm{axis}(g)}$ are coarsely bi-Lipschitz, we should just consider the axis of each g_i and discuss the distance of their projections on each other.

2.4 Revisit Case 1.

For a more general result, we hope to express the first two cases in some similar language.

Note that in the previous section, we mentioned the similarity between Fact 1.2 (Contraction property) and Theorem ([MM00]Bounded geodesic image). This indicates that the nearest point projection and the subsurface projection have some common properties. Also notice that in Σ, each $Z_i = g_1^{n_1} \cdots g_{i-1}^{n_{i-1}} Y_i$ is the support of $g_1^{n_1} \cdots g_{i-1}^{n_{i-1}} g_i^{n_i} (g_1^{n_1} \cdots g_{i-1}^{n_{i-1}})^{-1}$. So in Case 1, instead of instead of orbits of arbitrary markings, we should consider the fix sets of each pseudo-Anosov element. The axes come in play very well. (Or rather a set of quasi-geodesics which the pseudo-Anosov g permutes, which are in the δ-Hausdorff-neighborhood of some tight geodesics connecting two projective measured laminations which g fixes. See [MM00, Section 6.] and [Bow08]).

Suppose $\alpha_i = \mathrm{axis}(g_i)$. Then we can repeat the discussion in the overlapping case on the following set of geodesics supported on $\mathcal{C}(S)$

$$\{\beta_1 = \alpha_1, \beta_2 = g_1^{n_1} \alpha_2, \beta_3 = g_1^{n_1} g_2^{n_2} \alpha_3, \ldots, \beta_k = g_1^{n_1} \cdots g_{k-1}^{n_{k-1}} \alpha_k, \beta_{k+1} = g_1^{n_1} \cdots g_k^{n_k} \alpha_1\}.$$

With Fact 1.2 (Contraction property), we can show

$$d_{\beta_i}(\beta_1, \beta_{k+1}) > d_{\alpha_i}(\alpha_{i-1}, g_i^{n_i} \alpha_{i+1}) - M_2$$

for some constant M_2 independent of the cardinality of $\mathcal{A} = \{\alpha_i\}$. Moreover, we can show

$$d(\beta_1, \beta_{k+1}) \geq \sum (d_{\alpha_i}(\alpha_{i-1}, g_i^{n_i} \alpha_{i+1}) - M_3).$$

The proof of these two inequalities is included in the proof of Lemma 3.3 (Generalized local to global).

3 Key Lemmas

The goal of this section is to prove Theorem 3.4 Good word estimate. We boil the ideas from the previous section down and relate the contraction property of nearest point projections to Lemma 3.1 (Projection estimates; Behrstock inequality [Beh06]). We define a projection from $\mathcal{C}(S)$ to any geodesic α in $\mathcal{C}(Y)$ for any subsurface Y by composing the subsurface projection with the nearest point projection to α in $\mathcal{C}(Y)$. We change the argument of Lemma 3.1 to a symmetric statement as in Lemma 3.2 (Lemma of triples). This is inspired by the consistency condition of [BKMM08]. Then we generalize Lemma 2.2 ([BMM06]), and state and show Lemma 3.3 (Generalized local to global). At the end of this section, we construct a sequence of axes of non-elliptic elements for each good word. This sequence is fixed by the good word, hence fellow travel with the axis of this word. We then apply Lemma 3.2 and obtain Theorem 3.4 Good word estimate.

3.1 Lemma of Triples

We first project curve a in $\mathcal{C}(S) - B_1(\partial X)$ to $\mathcal{C}(X)$ by subsurface projection π_X. For any geodesic $\alpha \subset \mathcal{C}(X)$, we can then project $\pi_X(a)$ to α by nearest point projection. If X is an annulus, $\mathcal{C}(X)$ is quasi-isometric to $\mathbb{Z}$. Therefore we will assume $\mathcal{C}(X) = \alpha$. Abusing the notation, we define $\pi_\alpha = \pi_\alpha \circ \pi_X$ and $d_\alpha = d_X \circ (\pi_\alpha \circ \pi_X)$. Lemma 1.4 says that $d_\alpha(y,z) \leq d_X(y,z) + 8\delta$.

In this section we will prove the lemma of triples which is a generalized form of Behrstock inequality.

Lemma 3.1 (Projection estimates; Behrstock inequality [Beh06])**.** *There is a constant b such that for any two overlapping subsurfaces Y and Z of S with $\xi(Y) \neq 0 \neq \xi(Z)$ and any curve c that intersects with Y and Z essentially,*

$$\min\{d_Y(\partial Z, c), d_Z(\partial Y, c)\} < b.$$

Note: The original proof uses 1.7 (Order and projections. [MM00]), hence does not have a good estimate for the constant b. For Y and Z with $\xi(Y) > 1$ and $\xi(Z) > 1$, see [Man10] for an elementary proof which shows $b = 10$ due to Chris Leininger.

Lemma 3.2 (Lemma of triples)**.** *Let X, Y, and Z be subdomains of S or S itself. Let α, β and γ be geodesics in $\mathcal{C}(X)$, $\mathcal{C}(Y)$, $\mathcal{C}(Z)$ respectively if X, Y, and Z are not annular. If either X, Y or Z is an annulus, we associate the core curve to that annular domain. Then there exists $L > 0$ depending* *only on $\xi(S)$ such that at most one of*

$$d_\alpha(\beta,\gamma), d_\beta(\gamma,\alpha), d_\gamma(\alpha,\beta)$$

is greater than L.

Recall that we use the minimal distance between sets.

Proof of lemma of triples. We will prove this statement case by case modulo symmetry according to the relationships between X, Y and Z.

We repeat a common strategy in each case. We will show that each of the 3 pairs (P,Q) among $\{d_\alpha(\beta,\gamma), d_\beta(\alpha,\gamma), d_\gamma(\alpha,\beta)\}$ satisfy the statement similar to Lemma 3.1 (Projection estimates; Behrstock inequality [Beh06]); that is,

$$\exists b' > 0 \ (P > b' \Rightarrow Q < b'). \tag{$\mathbb{B}$}$$

If (P,Q) satisfies ($\mathbb{B}$), so does (Q,P). So we only need to do this at most three times for each case. So it helps to choose the right one in a pair to set to be large so that the argument is easier. The bound grows as we go through every case. Because the curve complexes of annuli are not subsets of $\mathcal{C}(S)$, we will discuss it separately.

Case 0. At least one of X, Y, z is an annulus.

Subcase 0.1. All of X, Y and Z are annuli. Thus α, β, γ are the core curves. Apply Lemma 3.1 directly on pairs that are not disjoint. That is, if X and Y are not disjoint, set $d_\gamma(\alpha,\beta)$ to be large and Lemma 3.1 says that $d_\alpha(\beta,\gamma)$ and $d_\beta(\alpha,\gamma)$ are small. For pairs that are disjoint, their projection distance to the other annulus is at most 1.

Subcase 0.2. Exactly two of X, Y and Z are annuli. Suppose X and Y are annuli and Z is a non-annular domain. We should show that $(d_\alpha(\beta,\gamma), d_\beta(\alpha,\gamma))$ and $(d_\alpha(\beta,\gamma), d_\gamma(\alpha,\beta))$ satisfy ($\mathbb{B}$). The other pair is symmetric to the latter.

If $d_\alpha(\beta,\gamma) > b$ where b is the constant in Lemma 3.1, then for every $z \in \gamma$, $d_\alpha(\beta,x) > b$. By Lemma 3.1, $d_\beta(\alpha,x) < b$, hence $d_\beta(\alpha,\gamma) < b$.

If $d_\gamma(\alpha,\beta) > b+3(2+8\delta)$, then α and β either are nested in Z or intersect ∂Z essentially. We assert that geodesics connecting α and $\pi_\gamma(\alpha)$ (in $\mathcal{C}(Z)$ if $\alpha \in Z$; in $\mathcal{C}(S)$ if α intersects ∂Z) is at least distance 2 away from β. This is due to the fact that π_Z and π_γ is coaresly Lipschitz. Therefore,

$$d_\beta(\alpha,\gamma) \le d_\beta(\alpha,\pi_\gamma(\alpha)) + d_\beta(\pi_\gamma(\alpha),\gamma) \le 2M,$$

by Theorem 1.3.

Subcase 0.3. Only one of X, Y and Z is an annulus. Suppose X is an annulus while Y and Z are non-annular. Then we can deal with $(d_\beta(\alpha,\gamma), d_\gamma(\alpha,\beta))$ in Cases 1 through 4 because α is just a simple closed curve in S. For $(d_\alpha(\beta,\gamma), d_\beta(\alpha,\gamma))$, suppose $d_\beta(\alpha,\gamma) > b + 3(2+8\delta)$. Similarly to Case 0.2, geodesics connecting $z \in \gamma$ and $\pi_\beta(\gamma)$ is at least distance 2 away from α (in $\mathcal{C}(S)$ if α intersects ∂Y; in $\mathcal{C}(Y)$ if $\alpha \subset Y$). Therefore, $d_\alpha(\beta,\gamma) \leq 2M$.

Case 1. All the domains overlap each other essentially, and the minimum distance between the boundaries in $\mathcal{C}(S)$ is greater than 2: In this case, for any $x \in \alpha$, $y \in \beta$ and $z \in \gamma$, they all intersect each of X, Y, Z essentially.

For any $x \in \alpha$, $d(x,\partial X) = 1$, $d(x,\partial Y) \geq 2$ and $d(x,\partial Z) \geq 2$. By [MM00, Lemma 2.3(Lipschitz projection)], $d_Y(x,\partial X) \leq 2$ and $d_Z(x,\partial X) \leq 2$ for any $x \in \alpha$, hence $d_Y(\alpha,\partial X) \leq 2$ and $d_Z(\alpha,\partial X) \leq 2$. Then $d_\beta(\alpha,\partial X) \leq 2+8\delta$ and $d_\gamma(\alpha,\partial X) \leq 2+8\delta$.

The above inequalities also work if we exchange $x \in \alpha$ with $y \in \beta$ or $z \in \gamma$, and interchange between (α, X), (β, Y), and (γ, Z).

Now, suppose $d_\alpha(\beta,\gamma) > b+4+8\delta$, then $d_X(\partial Y,\partial Z) \geq d_\alpha(\partial Y,\partial Z) - 8\delta \geq d_\alpha(\beta,\gamma) - 4 - 8\delta > b$. By Behrstock inequality, $d_Y(\partial X,\partial Z) < b$ and $d_Z(\partial X,\partial Y) < b$. Hence $d_\beta(\alpha,\gamma) < b+4+8\delta$ and $d_\gamma(\alpha,\beta) < b+4+8\delta$.

Case 2. X and Y overlap, but $0 < d(\partial X,\partial Y) \leq 2$; and ∂Z intersects both X and Y essentially. There might be $x \in \alpha$ which is disjoint from ∂Y but there are at most three such vertices of α and they are adjacent because α is geodesic.

If every $x \in \alpha$ which is disjoint from ∂Y is also in Y, then we can still use [MM00, Lemma 2.3(Lipschitz projection)] and similar argument as above because .

If there is $x \in \alpha$ disjoint from Y, then $\pi_Y(\alpha)$ consists of two sets of diameter less than 4 in $\mathcal{C}(Y)$. Use the above argument on each of these components. Since we take the minimum definition of distance between sets, Behrstock inequality still works.

Rest of the cases involve at least two subdomains that are either disjoint, nested, or the same.

Case 3. Two of the subdomains are disjoint. If X and Y are disjoint, then $\pi_Y(\alpha) = \varnothing = \pi_X(\beta)$, which implies two of the triples are 0. $d_\gamma(\alpha,\beta) \leq 2+8\delta$ because each vertex of α and each vertex of β are disjoint.

Case 4. Two of the subdomains are nested or the same. Assume without loss of generality that $X \subset Y$ or $X = Y$, and Z is not disjoint with either of X or Y.

Now we will discuss the possibilities of $\pi_Y(\gamma)$:

1. If $Y = Z$, $\pi_Y(\gamma) = \gamma$ is a geodesic in $\mathcal{C}(Y)$.
2. If $Z \subset Y$, then $\operatorname{diam} \pi_Y(\gamma) \leq 2$.
3. If $Y \subset Z$, and each vertex of γ overlaps with Y, $\operatorname{diam} \pi_Y(\gamma) \leq M$, by Theorem 1.3. Otherwise, if $Y \Subset Z$, and there is a vertex of γ which is disjoint from Y, then $\pi_Y(\gamma)$ consists of two sets of diameter less than M in $\mathcal{C}(Y)$, by Theorem 1.3.
4. If Y and Z overlap, $\operatorname{diam} \pi_Y(\gamma) \leq 4$ because for any $z \in \gamma$, $d_Y(z, \partial Z) \leq 2$.

$(d_\alpha(\beta,\gamma), d_\gamma(\alpha,\beta))$ satisfies ($\mathbb{B}$). Assume $d_\alpha(\beta,\gamma)$ is large. Then ∂Z intersects both ∂X and ∂Y essentially because $X \subseteq Y$. However, $d(\partial X, \partial Y) \leq 1$. Therefore, $d(\alpha,\beta) \leq d(\alpha, \partial X) + d(\partial X, \partial Y) + d(\partial Y, \beta) \leq 3$. So $d_\gamma(\alpha,\beta)$ is bounded.

Take $d_\beta(\alpha,\gamma)$ to be large. We will assume that $d_\beta(\alpha,\gamma)$ is sufficiently large and show the other two are bounded.

In any of the above cases except when $Y = Z$ and β and γ share at least one end points in the boundary at infinity of $\mathcal{C}(Y)$, $\pi_Y(\gamma)$ is either one bounded set with bound $\max\{4, M\}$, or union of two bounded sets which are bounded by M.

Notice that when $Z = Y$, if β and γ fellow travel the entire time, then $d_\beta(\alpha,\gamma) = d_\gamma(\alpha,\beta) = 0$. On the other hand, if $\pi_\alpha(\beta)$ has infinite diameter, then so does $\pi_\beta(\alpha)$, and α and β fellow travel for infinite length and they have at least one same endpoint at infinity.

1. If $Z = Y$, take any $z \in \gamma$ and $x \in \alpha$, when $d_\beta(\alpha,\gamma)$ is sufficiently large, the geodesic $[xz]$ in $\mathcal{C}(Y)$ has a long segment $[x'z']$ which 2δ-fellow travels with β (Fact 1.1). So we can find w in $[x'z']$ and $y \in \beta$ such that $d_Y(w,y) \leq 2\delta$ and w far away from α and γ. So $d_\alpha(\beta, z) \leq d_\alpha(\beta, w) + d_\alpha(w, z) \leq 3(2\delta + 16\delta) + M$. Similarly, $d_\alpha(x,\gamma) \leq 3(2\delta + 16\delta) + M$.
2. If $Z \subset Y$, and $d_\beta(\alpha,\gamma)$ is larger than $54\delta + M$, then any geodesic connecting ∂X and ∂Z has a long segment $[x'z']$ which 2δ-fellow travel with some segment of β (Fact 1.1). Because subsurface projections are coarsely Lipshitz, by similar argument as above, we have $d_\alpha(\beta,\gamma) < L$ and $d_\gamma(\alpha,\beta) < L$.

3. If $X \subseteq Y \subsetneq Z$, then $d_Z(\alpha, \beta)$ and $d_X(\beta, \gamma)$ are bounded because ∂Y and $partial Z$ are disjoint and subsurface projections are coarsely Lipschitz. Hence $d_\gamma(\alpha, \beta)$ and $d_\alpha(\beta, \gamma)$ are bounded.
4. If Y and Z overlap, $\pi_Y(\gamma)$ has diameter less than 4. To show $d_\alpha(\beta, \gamma)$ is bounded, we use the same strategy as previous cases: we connect any $z \in \gamma$ and $x \in \alpha$ with a geodesic $[xz]$ in $\mathcal{C}(Y)$. Then there is a long segment $[x'z']$ on $[xz]$ which 2δ-fellow travels with γ. So by similar argument, we can show that $d_\alpha(\beta, \gamma)$ is bounded. $d_\gamma(\alpha, \beta)$ is bounded because ∂X and Z overlap, and $d_Z(\partial X, \partial Y) \leq 3$.

In summary, we just need to take $L > 54\delta + 9 + b + M$.

□

Note that Behrstock inequality can be considered as a special case in lemma of triples: suppose $d_X(\mu, \partial Y) > L + 4\max\{M, 4\} + 8\delta$ and $X \pitchfork Y$. Pick one vertex v in μ which overlap X and some component Z of $S - v$ has complexity $\xi(Z) \geq 1$. Now consider a geodesic $\alpha \subset \mathcal{C}(X)$ connecting $\pi_X(\mu)$ and $\pi_X(\partial Y)$, a geodesic $\beta \subset \mathcal{C}(Y)$ connecting $\pi_Y(\mu)$ and $\pi_Y(\partial X)$, and any $\gamma \in \mathcal{C}(Z)$. Then $d_\alpha(\mu, \partial Y) > L + 2\max\{M, 4\} + 8\delta$ and $d_\alpha(\beta, \gamma) > L$. By lemma of triples, $d_\beta(\alpha, \gamma) < L$, thus $d_Y(\mu, \partial X) < L + 4\max\{M, 4\}$. Take $b = L + 4\max\{M, 4\} + 8\delta$. Here we disregard the actual constants of Behrstock inequality. We cannot produce $b = 10$ with lemma of triples.

3.2 Local to Global Lemma

3.2.1 Local to Global

Now we would like to replace Behrstock inequality in the proof of Theorem 2.2 with the lemma of triples and to replace subsurfaces Y_α with "geodesics" $\alpha \in \mathcal{C}(Y_\alpha)$. But we will need to define a partial order Σ and the corresponding $\mu^\pm(\alpha, A)$ so their projection onto α can be used as the word length of g^n if $\alpha = \mathrm{supp}(g)$.

Suppose $\mathcal{A} = \{\alpha\}$ where each α is a geodesic in $\mathcal{C}(Y_\alpha)$ and $Y_\alpha = \mathrm{supp}(\alpha)$.

We want to eventually make the distance of projections of some elements from the left of α and some elements from the right of α onto α to be sufficiently large just like in Lemma 2.2. So we will need to give $\mathcal{A}$ some partial order.

So we assume further that $\mathcal{A}$ has a partial order $\prec$ which satisfies two conditions:

1. If α and β are not ordered, then $Y_\alpha \cap Y_\beta = \varnothing$; and
2. If β is in $W_\alpha^- = W_\alpha^-(\mathcal{A})$, the set of all maximum elements in $\mathcal{A}$ preceding α, or $W_\alpha^+ = W_\alpha^+(\mathcal{A})$, the set of all minimum elements in $\mathcal{A}$ succeeding α, then $Y_\alpha \cap Y_\beta \neq \varnothing$.

In the discussion in case 4 in the proof of lemma of triples, for any $\alpha, \beta \in \mathcal{A}$, $\pi_\alpha(\beta)$ is some bounded set except when $Y_\alpha = Y_\beta$ and α and β fellow travel for infinite distance. If α and W_α^+ (or W_α^-) have the same support and share both endpoints at infinity of $\mathcal{C}(Y_\alpha)$, then potentially we may have a "backtracking" segment, which will affect any estimate gravely. So we will dismiss this case.

Denote $\mathbf{I} = \mathbf{I}(\mathcal{A})$, the initial elements, to be the set of elements α with $W_\alpha^- = \varnothing$ and $\mathbf{T} = \mathbf{T}(\mathcal{A})$, the terminal elements, to be the set of elements α with $W_\alpha^+ = \varnothing$. Note that all the elements in W_α^+ (or W_α^-, $\mathbf{I}$, $\mathbf{T}$) have disjoint domains because they are not ordered and condition 1. So $\mathrm{card}(W_\alpha^\pm) \leq \xi(S)$. Also, if $\beta \in W_\alpha^+$ has support $Y_\beta \supsetneq Y_\alpha$ or $Y_\beta = Y_\alpha$, then $W_\alpha^+ = \{\beta\}$ because if there are other $\gamma \in W_\alpha^+$, then Y_γ is disjoint from Y_β hence from Y_α, which contradicts with condition 2 of the partial order on $\mathcal{A}$. Similarly for W_α^-. We call elements which are neither in $\mathbf{I}$ nor in $\mathbf{T}$ **internal**.

Lemma 3.3 (Generalized local to global)**.** *Given $\mathcal{A}$ as above. Suppose if α and W_α^+ (or α and W_α^-) share the same support, they at most share one endpoint at infinity of Y_α. There exist $M_1 > M_3 > M_2 > 0$ and $C_0, C_1 > 0$ such that if we further assume for every α internal,*

$$d_\alpha(W_\alpha^-, W_\alpha^+) > M_1, \tag{$\star$}$$

then

1.
$$d_\alpha(\mathbf{I}, \mathbf{T}) \geq d_\alpha(W_\alpha^-, W_\alpha^+) - M_2, \tag{$\star\star$}$$
and

2.
$$d_{\mathcal{M}}(\mathbf{I}, \mathbf{T}) \geq \sum_{\alpha \in \mathcal{A}} C_0(d_\alpha(W_\alpha^-, W_\alpha^+) - M_3) - C_1. \tag{$\star\star\star$}$$

Remarks:

1. ($\star$) guaranteed that if α, W_α^+ and W_α^- are geodesics in $\mathcal{C}(Y_\alpha)$, then they cannot all share one same endpoint at infinity. But W_α^+ may have one same endpoint at infinity with α while W_α^- shares the other endpoint of α at infinity.

2. If we only do straight forward induction for each internal i with lemma of triples, we will not achieve $(\star\star)$ with M_2 independent of cardinality of $\mathcal{A}$. It takes some effort to do induction on the cardinality of $\mathcal{A}$ and to check after taking one element out of $\mathcal{A}$, rest of the sequence will be divided into 3 parts and on each part and its corresponding $\mathbf{I}$, $\mathbf{T}$, and $W_\alpha^\pm$ still satisfies $(\star)$.

3. For readers with deeper understanding of hierarchy, time order and partial order $\prec_p$ in [MM00], it might be interesting to know that for every α internal, Y_α supports a geodesic in the every hierarchy linking $\mathbf{I}$ and $\mathbf{T}$ because of Lemma (Large link[MM00]). The given partial order $\prec$ coincides with partial order $\prec_p$ defined in [MM00] for overlapping pairs.

Proof of Generalized local to global. Take $M_2 > 2L+8\delta$, $M_3 > M_2+2L+36\delta+6$ and $M_1 > M_3 + M_6(S) + 2L + 44\delta + 6$ where $L > 0$ is the constant from lemma of triples and $M_6(S) > 0$ is the threshold constant in Theorem 1.5 ([MM00]).

Consider the base case $\mathbf{card}(\mathcal{A}) = 3$: If all three are initial elements, then the statement is automatically true. Otherwise, there can be at most one non-initial element α and $\mathbf{I} = W_\alpha^-$, $\mathbf{T} = W_\alpha^+$, which also render the statement trivially.

Suppose the statement is true for any $\mathcal{A}$ with cardinality less than n.

Now suppose $\mathcal{A}$ has n elements and satisfies the local condition $\star$. So pick any internal α in $\mathcal{A}$ and define $\mathcal{A}^- = \{\beta \in \mathcal{A} | \beta \prec \alpha\} \cup \{\alpha\}$, $\mathcal{A}^+ = \{\beta \in \mathcal{A} | \beta \succ \alpha\} \cup \{\alpha\}$ and $\mathcal{B} = \{\beta \in \mathcal{A} \| \beta \text{ and } \alpha \text{ are not ordered}\}$. We will show $\mathcal{A}^+$, $\mathcal{A}^-$ and $\mathcal{B}$ still satisfy the local condition $(\star)$.

Claim. *For $\alpha \neq \beta \in \mathcal{A}^-$, $W_\beta^-(\mathcal{A}^-) = W_\beta^-(\mathcal{A})$ and $W_\beta^+(\mathcal{A}^-) \subset W_\beta^+(\mathcal{A})$; for $\beta \in \mathcal{A}^+$, $W_\beta^+(\mathcal{A}^+) = W_\beta^+(\mathcal{A})$ and $W_\beta^-(\mathcal{A}^+) \subset W_\beta^-(\mathcal{A})$.*

Proof of claim. If $\beta \in \mathcal{A}^-$ and $\gamma \in W_\beta^-(\mathcal{A})$, then $\gamma \prec \beta \prec \alpha$ and $Y_\gamma \pitchfork Y_\beta$. Therefore, $\gamma \in W_\beta^- \mathcal{A}^-$ and hence $W_\beta^-(\mathcal{A}) \subset W_\beta^-(\mathcal{A}^-)$. Similarly, if $\beta \in \mathcal{A}^+$, $W_\beta^+(\mathcal{A}) \subset W_\beta^+(\mathcal{A}^+)$. The other relations are obvious.

Therefore, $\mathbf{I}(\mathcal{A}^-) = \mathbf{I} \cap \mathcal{A}^-$ and $\mathbf{T}(\mathcal{A}^+) = \mathbf{T} \cap \mathcal{A}^+$. Moreover, $\mathbf{T}(\mathcal{A}^-) = \{\alpha\}$ and $\mathbf{I}(\mathcal{A}^+) = \{\alpha\}$ from definition.

Thus, if β is internal in $\mathcal{A}^-$,

$$d_\beta(W_\beta^-(\mathcal{A}^-), W_\beta^+(\mathcal{A}^-)) \geq d_\beta(W_\beta^-(\mathcal{A}), W_\beta^+(\mathcal{A})) > M_1,$$

since we define distance between sets to be the minimal distance between points in each sets. Symmetrically, if β is internal in $\mathcal{A}^+$,

$$d_\beta(W_\beta^-(\mathcal{A}^+), W_\beta^+(\mathcal{A}^+)) \geq d_\beta(W_\beta^-(\mathcal{A}), W_\beta^+(\mathcal{A})) > M_1.$$

As for $\mathcal{B}$ and $\beta \in \mathcal{B}$, $W^{\pm}_{\beta}(\mathcal{B}) = W^{\pm}_{\beta} \cap \mathcal{B}$, $\mathbf{I}(\mathcal{B}) = \mathbf{I} \cap \mathcal{B}$ and $\mathbf{T}(\mathcal{B}) = \mathbf{T} \cap \mathcal{B}$. That is because any element in $\mathcal{B}$ is not ordered with α and hence it is not ordered with any elements in $\mathcal{A}^-$ or $\mathcal{A}^+$ by the definition of $\mathcal{A}^{\pm}$ and partial order.

$\square$

[End of proof of claim.]

By triangle inequality,

$$d_\alpha(\mathbf{I}, \mathbf{T}) \geq d_\alpha(W^-_\alpha, W^+_\alpha) - d_\alpha(\mathbf{I}, W^-_\alpha) - d_\alpha(W^+_\alpha, \mathbf{T}).$$

Because for any $\beta \in W^-_\alpha$, β is internal in $\mathcal{A}^-$, by the induction assumption,

$$\begin{aligned} d_\beta(\mathbf{I}, \{\alpha\}) \quad & \geq d_\beta(\mathbf{I}(\mathcal{A}^-), \mathbf{T}(\mathcal{A}^-)) \\ & > d_\beta(W^-_\beta(\mathcal{A}^-), W^+_\beta(\mathcal{A}^-)) - M_2 \\ & \geq d_\beta(W^-_\beta(\mathcal{A}), W^+_\beta(\mathcal{A})) - 2 - 8\delta - M_2 \\ & > L. \end{aligned}$$

Thus, by the lemma of triples,

$$d_\alpha(\mathbf{I}, W^-_\alpha) < L.$$

Similarly we can get

$$d_\alpha(W^+_\alpha, \mathbf{T}) < L.$$

Therefore,

$$d_\alpha(\mathbf{I}, \mathbf{T}) \geq d_\alpha(W^-_\alpha, W^+_\alpha) - 2L > d_\alpha(W^-_\alpha, W^+_\alpha) - M_2.$$

We will now show that we can just sum the distance up. From $(\star)$, we know that if $Y_\alpha \Subset Y_\beta$ or $Y_\alpha \pitchfork Y_\beta$ for $\beta \in W^{\pm}_\alpha$, then Y_α will at most show up once. However, for example when $Y_\alpha = S$, we may have to deal with several geodesics in $\mathcal{C}(Y_\alpha)$. Thus we connect $\pi_{Y_\alpha}(\mathbf{I})$ to $\pi_{Y_\alpha}(\mathbf{T})$ with a (tight) geodesic h_{Y_α} in $\mathcal{C}(Y_\alpha)$ and need to find large segment on each of these geodesics on the same domain with disconnected projections on h_{Y_α} so we can estimate the length of h_{Y_α} by summing up $d_\alpha(W^-_\alpha, W^+_\alpha)$. From Fact 1.1, and because

$$d_\alpha(\mathbf{I}, \mathbf{T}) > d_\alpha(W^-_\alpha, W^+_\alpha) - M_2 > M_1 - M_2 > 18\delta,$$

for any $x \in \mathbf{I}$ and $y \in \mathbf{T}$, there is x_α and y_α on $[\pi_{Y_\alpha}(x)\pi_{Y_\alpha}(y)]$ such that $d(x_\alpha, y_\alpha) > d_\alpha(\mathbf{I}, \mathbf{T}) - 18\delta$ and $[x_\alpha, y_\alpha]$ is in 2δ-neighborhood of α.

Now for $\beta \in W_\alpha^+$, $[x_\beta, y_\beta]$ is in 2δ-neighborhood of β. Because $d_\alpha \le d_{Y_\alpha} + 8\delta$,

$$\pi_\alpha \circ \pi_{Y_\alpha}([x_\beta, y_\beta]) \subset N_{10\delta}(\pi_\alpha(\beta)).$$

So if for each $\alpha \in \mathcal{A}$, we take $[x'_\alpha y'_\alpha]$ to be the middle segment of $[x_\alpha y_\alpha]$ of length $d(x'_\alpha, y'_\alpha) = d_\alpha(W_\beta^-, W_\beta^+) - 36\delta - 6$ provided $d_\alpha(W_\beta^-, W_\beta^+)$ sufficiently large, then we have $[x'_\alpha y'_\alpha]$ and $[x'_\beta y'_\beta]$are disjoint for any adjacent α and β.

Therefore, from Thm 1.5 ([MM00]),

$$\begin{aligned}
d_{\tilde{\mathcal{M}}}(\mathbf{I}, \mathbf{T}) &\ge e_0^{-1}\Big(\big(\sum_{\substack{Y \subset S \\ d_Y(\mathbf{I},\mathbf{T}) \ge M_6(S)}} d_Y(\mathbf{I}, \mathbf{T})\big) - e_1\Big) \\
&\ge e_0^{-1}\Big(\sum_{\alpha \in \mathcal{A}} d_{Y_\alpha}(\mathbf{I}, \mathbf{T}) - e_1\Big), \\
&\qquad\qquad\qquad (\text{because } d_{Y_\alpha}(\mathbf{I}, \mathbf{T}) > M_1 - M_2 > M_6(S),) \\
&\ge e_0^{-1}\Big(\sum_{\alpha \in \mathcal{A}} \big(d_\alpha(\mathbf{I}, \mathbf{T}) - 8\delta\big) - e_1\Big), \\
&\qquad\qquad\qquad (\text{because } d_{Y_\alpha} \ge d_\alpha(\mathbf{I}, \mathbf{T}) - 8\delta,) \\
&\ge e_0^{-1}\Big(\sum_{\alpha \in \mathcal{A}} \big(d_\alpha(W_\alpha^-, W_\alpha^+) - 44\delta - 6 - M_2\big) - e_1\Big), \\
&\qquad\qquad\qquad (\text{because } d_\alpha(\mathbf{I}, \mathbf{T}) \ge \delta(W_\alpha^-, W_\alpha^+) - 36\delta - 6.)
\end{aligned}$$

So we will take $C_0 = e_0^{-1}$, $C_1 = e_0^{-1} e_1$.

For the last part, it is just the result of the Theorem of Large Link in [MM00] because M_6 is larger than the constant in that theorem. As to the partial order, we know from Lemma 4.18 (Time Order) in [MM00] and definition of $\prec_p$, and the fact that if $d_\alpha(\beta_s, \beta_t)$ large, $d_{Y_\alpha}(\partial Y_{\beta_s}, \partial Y_{\beta_t})$ is also large and thus $\partial Y_{\beta_s} \pitchfork \partial Y_{\beta_t}$.

□

In fact, if $\mathrm{supp}(\beta) = \mathrm{supp}(\gamma)$, the partial order is the same as $\prec_p$ in the sense that on the tight geodesic h with $D(h) = \mathrm{supp}(\beta) = \mathrm{supp}(\gamma)$, $\pi_h(\beta)$ shows up before $\pi_h(\gamma)$.

3.3 Good Word Estimate

3.3.1 Good Word Estimate

Fix a finite ordered set $(g_i)_1^k \in \mathcal{MCG}(S)^k$ of non-elliptic elements. Name $X_i = \mathrm{supp}(g_i)$ to be the minimal essential subsurface where g_i is supported and α_i to be an element in the finite set of the tight geodesics in $\mathcal{C}(Y_i)$ which g_i permutes. Suppose further each X_i is connected. (Otherwise we can decompose g_i into a set of reducible elements supported on connected components of X_i.)

Given any $(n_i)_1^k \in \mathbb{Z}^k$, denote $\Lambda = (g_i^{n_i})_1^k$. We extend our list infinitely by taking $g_i = g_j$, $n_i = n_j$, and $\alpha_i = \alpha_j$ if $i \equiv j(\mathrm{mod} k)$. Take $\gamma_0 = \mathrm{Id}$ and inductively define $\gamma_{i+1} = \gamma_i g_i^{n_i}$ for all $i \in \mathbb{Z}$. Set the infinite sequence β_Λ of sets in $\mathcal{C}(S)$ to be

$$(\ldots, \gamma_{-1-s}\alpha_{-s}, \gamma_{-s}\alpha_{-s+1}, \ldots, \gamma_{-1}\alpha_0, \gamma_0\alpha_1, \gamma_1\alpha_2, \gamma_2\alpha_3, \ldots, \gamma_t\alpha_{t+1}, \ldots).$$

Then β_Λ is invariant under $\langle g \rangle$ for $g = g_1^{n_1} g_2^{n_2} \cdots g_k^{n_k}$.

In Section 4.2, we include elliptic elements in the ordered set. We take A_i to be $\mathrm{axis}(g_i)$ when g_i is pseudo-Anosov while taking A_i to be $\mathrm{fix}(g_i)$ when g_i is either reducible or elliptic. In that setting, we can similarly define

$$\beta_\Lambda = (\ldots, \gamma_{-1-s}A_s, \gamma_{-s}A_{-s+1}, \ldots, \gamma_{-1}A_0, \gamma_0 A_1, \gamma_1 A_2, \gamma_2 A_3, \ldots, \gamma_t A_{t+1}, \ldots).$$

Also for simplicity, we assume $\biguplus_{i=1}^k X_i$ is connected. We will further assume that if $\mathrm{supp}(g_i) = \mathrm{supp}(g_j)$ for some $i \not\equiv j(\mathrm{mod}\ k)$ and for all $i \not\equiv l \not\equiv j(\mathrm{mod}\ k)$, $\mathrm{supp}(g_l)$ and $\mathrm{supp}(g_i)$ are disjoint, then g_i and g_j are not commensurable. For such $(g_i)_1^k$, we say that $(g_i)_1^k$ satisfies **non-cancellation condition**. In addition, if each $|n_i| l_{X_i}(g_i)$ are sufficiently large (see Theorem 3.4), we call word $g_1^{n_1} g_2^{n_2} \cdots g_k^{n_k}$ a **good word**. Name $\mathcal{A}^0 = \{g_i\}_{i \in \mathbb{Z}}$.

We will give $\{i\}_{i \in \mathbb{Z}}$ a new partial order $\prec$ as follow:

1. If X_i and X_{i+1} are disjoint, then i and $i+1$ are not ordered.

2. For $i < j$ if $X_i \cap X_j \neq \varnothing$, then we say $i \prec j$.

3. Then we complete the partial order by transitivity. Notice that the completion will not change the above relations.

Define $W_i^- = \max\{j | j \prec i\} \subset \mathbb{Z}$ to be the set of all maximal elements $j \prec i$ (that is, there is no $l \prec h$ such that $j \prec l \prec i$) and $W_{\beta_i}^- = \cup\{\beta_s | s \in W_i^-\}$; similarly define $W_i^+ = \min\{j | i \prec j\} \subset \mathbb{Z}$ and $W_{\beta_i}^+$.

For any $j \in W_i^\pm$, X_j cannot be disjoint from X_i due to the way we define $\prec$. Also, since we assume that $\biguplus_{i=1}^k X_i$ is connected, if $j \in W_i^-$, then $i - k < j < i$. Similarly, if $j \in W_i^+$, then $i < j < i + k$.

Moreover, by definition, any j and j' in W_i^+ (or W_i^-) are unordered, and thus X_j and $X_{j'}$ are disjoint. Therefore, if $Y_j \supseteq Y_i$ for any j in W_i^+ (or W_i^-), then $W_i^+ = \{j\}$ (or $W_i^- = \{j\}$.)

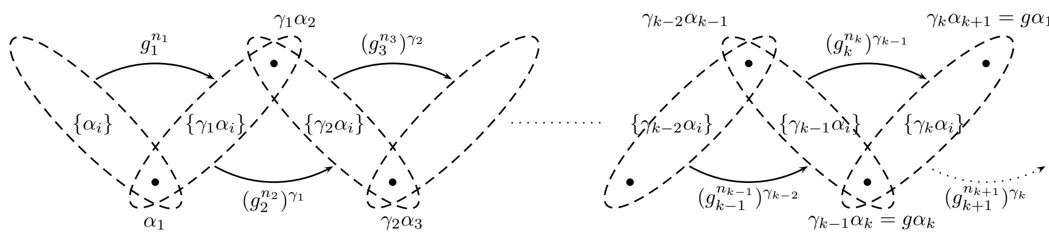

Figure 8: A schematic picture of the invariant sequence β_Λ. The ellipses represent orbits of $\{\alpha_i\}_1^k$. Each $\gamma_i\alpha_i$ is fixed by $(g_i^n)^{\gamma_{i-1}}$. We can extend this sequence of axes bi-infinitely so that it contains $\{g^n\alpha_1, g^n\alpha_k\}_{n=-\infty}^{\infty}$.

Define $\tilde{g}_i = g_i^{\gamma_{i-1}} = \gamma_{i-1}g_i\gamma_{i-1}^{-1}$. Take $\tilde{g}_i = g_i^{\gamma_{i-1}} = \gamma_{i-1}g_i\gamma_{i-1}^{-1}$. So $Y_i = \text{supp}(\tilde{g}_i) = \gamma_{i-1}X_i$ and $\text{axis}(\tilde{g}_i) = \beta_i = \gamma_{i-1}\alpha_i$. Notice that $\beta_\Lambda = \mathcal{A} = \{\beta_i\}_{i\in\mathbb{Z}}$ is invariant under $\langle g\rangle$ where $g = g_1^{n_1}\cdots g_k^{n_k}$ since $g^n(Y_i,\beta_i) = (g^nY_i, g^n\beta_i) = (Y_{i+kn}, \beta_{i+kn})$.

Theorem 3.4 (Good word estimate.)**.** *Given non-elliptic $(g_i)_1^k \in \mathcal{MCG}(S)^k$ satisfying non-cancellation condition, there exist $M_1' > M_3' > 0$ and $C_0 = C_0(\xi(S)) > 0$, if $g = g_1^{n_1}\cdots g_k^{n_k}$ where $|n_i|l_{X_i}(g_i) > M_1'$ for all $1 \le i \le k$,*

$$l_{\mathcal{M}}(g) \ge C_0 \sum_{i=1}^{k} (|n_i| l_{X_i}(g_i) - M_3').$$

Remark: Here $M_1' = M_1 + M_0$ and $M_3' = M_3 + M_0$ where $M_1 > M_3$ and $C_0 > 0$ are constants from Lemma 3.3 (Generalized local to global), and

$$\begin{aligned} M_0 &= \max_i \{\operatorname{diam}\pi_{\alpha_i}(\alpha_s) + d_{\alpha_i}(\alpha_s,\alpha_t) + \operatorname{diam}\pi_{\alpha_i}(\alpha_t) | s \in W_i^-, t \in W_i^+\} \\ &\le \max_i \{\operatorname{diam}\pi_{\alpha_i}(\alpha_s) + d_{\alpha_i}(\alpha_s,\alpha_t) + \operatorname{diam}\pi_{\alpha_i}(\alpha_t) \\ &\qquad | \operatorname{diam}\pi_{\alpha_i}(\alpha_s), \operatorname{diam}\pi_{\alpha_i}(\alpha_t) < \infty\} \end{aligned}$$

is finite provided adjacent elements are not commensurable. Clearly M_0 does not depend on (n_i).

Proof. Take $\mathcal{A} = \{\beta_i\}_{\mathbb{Z},\prec}$. We need to verify that $\prec$ satisfies condition 1 before Lemma 3.3 and that $W_{\beta_i}^{\pm} = \bigcup\{\beta_j | j \in W_i^{\pm}\}$ (or, that $\prec$ defined in this section, is equivalent to the $\prec$ defined in local to global lemma.)

If i and j are not ordered and $i = j+1$, from our set up, $X_i = X_{j+1}$ and X_j are disjoint; hence $Y_j = \gamma_{j-1}.X_j$ and $Y_i = \gamma_{i-1}.X_i = \gamma_{j-1}X_i$ are disjoint. If i and j are not ordered and $i > j+1$, then X_l is disjoint from both X_i and X_j for all $j < l < i$, and $Y_i = \gamma_{i-1}X_i = \gamma_{j-1}.X_i$ disjoint from $Y_j = \gamma_{j-1}X_j$.

If $i \prec j$, then there is a maximal chain $i = i_0 \prec i_1 \prec \cdots \prec i_l = j$,which is maximal in the sense that for any $i_p < i' < i_{p+1}$, $X_{i'}$ is disjoint from both X_{i_p} and $X_{i_{p+1}}$. Then X_{i_p} and $X_{i_{p+1}}$ are not disjoint, and therefore $Y_{i_p} = \gamma_{i_p-1}X_{i_p}$ and $Y_{i_{p+1}} = \gamma_{i_{p+1}-1}X_{i_{p+1}} = \gamma_{i_p-1}g_{i_p}X_{i_{p+1}}$ are not disjoint. Therefore, $\beta_i \prec \beta_j$. The other direction is similar.

Now all we need is to guarantee $d_{\beta_i}(W^-_{\beta_i}, W^+_{\beta_i}) = \min\{d_{\beta_i}(\beta_s,\beta_t) | s \in W^-_i, t \in W^+_i\} > M_1$.

For any $s \in W^-_i$ and $t \in W^+_i$ (hence $i-k < s < i < t < i+k$,)

$$\begin{aligned}
d_{\beta_i}(\beta_s,\beta_t) &= d_{\gamma_{i-1}\alpha_i}(\gamma_{s-1}.\alpha_s, \gamma_{t-1}\alpha_t)\\
&= d_{\alpha_i}(\gamma_{i-1}^{-1}\gamma_{s-1}\alpha_s, \gamma_{i-1}^{-1}\gamma_{t-1}\alpha_t)\\
&= d_{\alpha_i}((g_s^{n_s}\cdots g_{i-1}^{n_{i-1}})^{-1}\alpha_s, g_i^{n_i}(g_{i+1}^{n_{i+1}}\ldots g_{t-1}^{n_{t-1}}\alpha_t)).
\end{aligned}$$

From our previous discussion, X_s is disjoint from any X_j when $s < j < i$ and X_t is disjoint from any X_j when $i < j < t$. Thus,

$$\alpha_s = (g_s^{n_s}\cdots g_{i-1}^{n_{i-1}})^{-1}\alpha_s \subset \mathcal{C}(X_s),$$

and

$$\alpha_t = g_{i+1}^{n_{i+1}}\ldots g_{t-1}^{n_{t-1}}\alpha_t \subset \mathcal{C}X_t.$$

So we have

$$d_{\beta_i}(\beta_s,\beta_t) = d_{\alpha_i}(\alpha_s, g_i^{n_i}\alpha_t).$$

If $l_{X_i}(g_i) > M_0 + M_1$ for all i, then

$$\begin{aligned}
d_{\alpha_i}(\alpha_s, g_i.\alpha_t)) &\geq d_{\alpha_i}(\alpha_t, g_i.\alpha_t) - d_{\alpha_i}(\alpha_s,\alpha_t)\\
&\geq l_{X_i}(g_i) - M_0 > M_1
\end{aligned}$$

for any $s \in W^-_i$, $t \in W^+_i$.

Therefore, by Lemma 3.3 (generalized local to global), for any n and i,

$$\begin{aligned}
d_{\tilde{\mathcal{M}}}(\beta_i, g^n\beta_i) &= d_{\tilde{\mathcal{M}}}(\beta_i, \beta_{i+kn})\\
&\geq C_0(n-2)\sum_{j=1}^{k}(d_{\alpha_j}(\alpha_s, g_j.\alpha_t) - M_3) - C_1\\
&\geq C_0(n-2)\sum_{j=1}^{k}(l_{X_i}(g_i) - M_0 - M_3) - C_1.
\end{aligned}$$

Therefore, because for any $x \in \beta$, $d(x, g^n x) \geq d(\beta, g^n \beta)$,

$$l_{\mathcal{M}}(g) \geq \lim_{n\to\infty} \frac{1}{n} d_{\tilde{\mathcal{M}}}(\beta_i, g^n \beta_i) \geq C_0 \sum_{j=1}^{k} (l_{X_i}(g_i) - M_0 - M_3).$$

□

4 Bounded Generation

In this chapter we show by two proofs that subgroups of mapping class groups with exponential growth are not boundedly generated. First by growth argument through a direct application of Theorem 3.4 (Good word estimate). Next by comparing axes of pseudo-Anosovs with orbits of $g_1^{n_1} \cdots g_k^{n_k}$. The latter is closely related to the proof by Bestvina and Fujiwara [BF02].

4.1 Growth Argument

In this section we will prove $G < \mathcal{MCG}(S)$ with exponential growth is not boundedly generated by a growth argument with Theorem 3.4.

Theorem 4.1. *If $G < \mathcal{MCG}(S)$ has exponential growth, G is not boundedly generated.*

Proof. Given an ordered set of mapping classes $(g_i)_1^k$ and suppose $R > 0$ is sufficiently large, take $\Sigma = \Sigma(R) = \{(n_i) \in \mathbb{Z}^k | l(g_1^{n_1} \cdots g_k^{n_k}) \leq R\}$. We will decide the lower bound of R later.

For any subset $\{g_{i_j}\}_1^{k'}$ of $\{g_i\}_1^k$, if it satisfies non-cancellation condition, then

$$l(g_{i_1}^{n_{i_1}} \cdots g_{i_{k'}}^{n_{i_{k'}}}) \geq C_0 \sum (|n_{i_j}| l_{X_{i_j}}(g_{i_j}) - M')$$

provided each $|n_{i_j}|$ is large enough by Theorem 3.4. Note that M' can be taken to be universal for all subsets of $\{g_i\}_1^k$ with non-cancellation condition.

However, not every word in Σ is a good word. It might have "back-tracking" segments; that is, $g = g_{i_1}^{n_{i_1}} \cdots g_{i_\kappa}^{n_{i_\kappa}}$ while $\{g_{i_j}\}_1^\kappa$ does not satisfies non-cancellation condition. Or, some of the g_i's are elliptic. Or, it might have some small powers $|n_{i_j}|$.

So we will give a finite algorithm which dissect Σ into finitely many subsets. In each of these subsets, we rewrite the words so that it is a good word with a new generating set. We apply Theorem 3.4 in each of these subsets and keep track of the different M's. However, because this procedures is finite, M can be taken to be universal for all of these subsets, and we can count the cardinality of each of them with Theorem 3.4. We will conclude $|\Sigma(R)| \sim o(R^{k+1})$, which implies $\Sigma(R)$ cannot have exponential growth like our given subgroup $G < \mathcal{MCG}(S)$. This works for any given sequence $(g_i)_1^k$ with any cardinality k. Therefore we conclude G is not boundedly generated.

Step 0: Move the elliptic elements to the end. Suppose (g_{i_j}) is the subset of all the elliptic elements. We will deal with subsets Σ_l of Σ in which the powers of g_{i_j} are fixed. (There are only $\prod |\langle g_{i_j}\rangle|$ such subsets.) For each Σ_l, we consider a new list of mapping classes $(g_{i'})_1^{k'}$ which is constructed in the following way:

Starting from the last element, for every elliptic element g_{i_j} in $(g_i)_1^k$, if it has nonzero power in Σ_l, we conjugate all the elements after it by g_{i_j}, and remove g_{i_j} from our list. After exhausting all the elliptic elements, our new list $(g_i')_1^{k'}$ does not have any elliptic element in it.

For example, if g_2, g_6, g_7 are the only elliptic elements in $(g_i)_1^9$. Consider all the words with segments g_2, g_6^3 and g_7^5. We first conjugate g_8, g_9 with g_7^5, and then our sequence become $(g_1, \cdots, g_6, g_8^{g_7^5}, g_9^{g_7^5})$. In this way, we move g_7^5 to the end of the word and because it is given with (g_i), the effect of it on the stable length of the word is bounded. Then we move on to g_6^3. With the same procedure, we obtain a new list $(g_1, \cdots, g_5, g_8^{g_6^3 g_7^5}, g_9^{g_6^3 g_7^5})$. So we move $g_6^3 g_7^5$ to the end of each word with segments g_6^3 and g_7^5. Now we deal with g_2. So we obtain a new sequence without elliptic element

$$(g_1, g_3^{g_2}, \ldots, g_5^{g_2}, g_8^{g_2 g_6^3 g_7^5}, g_9^{g_2 g_6^3 g_7^5}).$$

Now we can write every word with segments g_2, g_6^3, and g_7^5 in terms of the new sequence with $g_2 g_6^3 g_7^5$ at the end. Because our list of elements is finite, and elliptic elements are of finite order, $\max\{|g_2^{n_2} g_6^{n_6} g_7^{n_7}|_w | (n_2, n_6, n_7) \in \mathbb{Z}^3\}$ is finite.

Step 1: Combine adjacent commensurable pairs. Next we will tidy the list by combining consecutive commensurable pairs: if g_i' and g_j' have same support and are commensurable and they commute with any g_l' where $i < l < j$, we can find $h \in \mathcal{MCG}(S)$ such that $g_i' = h^m$ and $g_j' = h^n$. Replace g_i' with h and delete g_j' from our list. Exhaust all possible such pairs in the list. Now our list should satisfy the non-cancellation condition.

Step 2: Applying good word estimate. Define $M_0(\Sigma_l) = \max\{d_{\alpha_i}(\alpha_s, \alpha_t)\}$ as in the proof of Theorem 3.4. We will just rename this new list without elliptic elements and consecutive commensurable pairs (g_i). By Theorem 3.4, if all the powers n_i are sufficiently large,

$$R \geq l_{\mathcal{M}}(g) \geq C_0 \sum_i (|n_i| l_{X_i}(g_i) - M).$$

So there are only at most $(\frac{R}{C_0}+M)^k/(\prod_i l_{X_i}(g_i)) \sim O(R^k)$ of big (n_i) in Σ_l satisfying this inequality.

Step 3: In words with short segments, move the short segments to the end. Now, suppose $(|n_{i_j}|)$ are not large enough; that is, $|n_{i_j}| l_{X_{i_j}}(g_{i_j}) \le M_0(\Sigma_l)+M_1$. We will then again divide Σ_l into finitely many subsets for all combinations of possible small segments. (To be precise, there are $\prod_j \frac{M_0+M_1}{l_{X_{i_j}}(g_{i_j})}$ such sets.) Like how we dealt with elliptic elements, we conjugate elements following the said g_{i_j} with small $|n_{i_j}|$, and move all the small segments to the end.

Step 4: Back to Step 1 and repeat Steps 1 through 4 if necessary. Deal with the consecutive commensurable pairs and apply Theorem 3.4 again. Notice that we might get a bigger M_0 as we obtain our new list of mapping classes. But since all the steps are finite and the length of the list does not increase, M_0 is decided by the original given list of elements.

Repeat until the list of elements becomes empty. This algorithm will stop at finite steps because the cardinality of the list strictly decreases after Steps 0, 1, and 3. Every time after Step 0 and Step 3, we essentially divide Σ_l into finitely many subsets. In each of those subsets, words have some common combination of short segments.

At the end, we reach a maximum M_0 which works for each of these subsets, and at the same time a maximum bound C_2 for the word length of all short segments. Now we use these new constants to estimate the stable length in each subsets uniformly. Note that M_0 is independent of sufficiently large R's. Take $R > M_0 + M_3 + C_0^{-1}C_2$ and each of these subsets have cardinality $o(R^{k+1})$.

So any $G < \mathcal{MCG}(S)$ with exponential growth is not boundedly generated. By Tits alternative, any virtually abelian $G < \mathcal{MCG}(S)$ does not have exponential growth and hence boundedly generated. We conclude that for any subgroup $G < \mathcal{MCG}(S)$, G is boundedly generated if and only if it is virtually abelian. $\square$

4.2 Free Subgroup Argument

In this section we will prove Theorem4.1 by comparing axes of pseudo-Anosovs with sequence β_Λ where $\Lambda = (f_i \in C_i)_1^k$ and (C_i) is a given collection of pseudo-Anosov cyclic subgroups and curve stabilizers.

Theorem 4.2. *[[BF07]] If $G < \mathcal{MCG}(S)$ has exponential growth and has pseudo-Anosov elements, and subgroups $C_1, \dots C_k < G$ are either cyclic or* *contained in some curve stabilizers respectively, then*

$$G \neq C_1 \cdots C_k.$$

Here β_Λ is a sequence of sets similarly defined as in Section 3.3. If C_i is pseudo-Anosov, we define $A_i = \text{axis}(C_i)$ as before. However, if C_i stabilizes some curve(s), we will take $A_i = \text{fix}(C_i)$. Foe take β'_Λ to be the sequence of sets of length $k+1$

$$B_1 = A_1, B_2 = f_1 A_2, B_3 = f_1 f_2 A_3, \ldots, B_k = f_1 \cdots f_{k-1} A_k, B_{k+1} = f_1 \cdots f_k A_1$$

while taking β_Λ to be $(\gamma_i A_i)_{-\infty}^{\infty}$ where $f_i = f_j$ and $A_i = A_j$ if $i \equiv j \pmod{k}$, $\gamma_0 = \text{Id}$, and $\gamma_{i+1} = \gamma_i f_i$. Then $\langle g \rangle$ fixes β_Λ for $g = f_1 f_2 \cdots f_k$.

This proof can be related to the quasi-homomorphism argument by Bestivina and Fujiwara [BF02],[BF07].

To prove this theorem, we first show that for any subgroup containing pseudo-Anosovs with exponential growth, there are infinitely many pseudo-Anosovs that are not commensurable up to conjugacy.

Then for every pseudo-Anosov $g \in G$ and $m \in \mathbb{Z}$, suppose $g^m = g_1 \cdots g_k$, we compare the sequence $\beta_{\Lambda(m)}$ where $\Lambda(m) = (f_{1,m}, \ldots, f_{k,m}) \in C_1 \times \cdots \times C_k$ with $\beta = \text{axis}(g)$.

As m tends to infinity, we can find axis g_i having arbitrarily large projection onto β for some pseudo-Anosov g_i where $C_i = \langle g_i \rangle$ due to Lemma3.3 and [MM00, Lemma 7.7]. By Lemma 1.10, g is commensurable up to conjugacy with g_i. Therefore we achieve the contradiction that we have infinitely many choices of pseudo-Anosovs that are not commensurable up to conjugacy with each other.

Lemma 4.3 (Lemma 7.7, [MM00])**.** *For any $L > 0$, there exists $N > 0$ such that for any $g \in \mathcal{MCG}(S)$, $\beta = \text{axis}(g)$, and $n > N$,*

$$d(\pi_\beta(x), \pi_\beta(g^n(x))) > L.$$

Lemma 4.4. *If $G \in \mathcal{MCG}(S)$ has exponential growth and* $\text{supp}(G) = \uplus\{\text{supp}(g) | g \in G\} = S$*, then there is an infinite set of pseudo-Anosovs $\{g_i\}$ such that for any $i \neq j$, any $l, k \neq 0$, and any $\gamma \in \mathcal{MCG}(S)$,*

$$g_i^l \neq (g_j^\gamma)^k.$$

That is, any pair (g_i, g_j) is not commensurable up to conjugacy.

Proof. With Ping Pong Lemma, one can find a free subgroup $\langle a, b\rangle$ of rank 2 generated by pseudo-Anosovs a and b with $l_S(a) = l_S(b)$. Take $\alpha = \operatorname{axis}(a)$ and $\beta = \operatorname{axis}(b)$. Moreover, by taking n large enough so that 1) $nl_S(a) > Nm^2 l_S(a) + 12\delta + 52\delta$, and $\min\{d_\alpha(x, a^{n'}x), d_\beta(x, b^{n'}x\} > M_1 + 2M_3$ for all $x \in \mathcal{C}(S)$ and $n' > n$. The former bound is from Lemma 1.10 and the later is achievable by [MM00, Lemma 7.7].

We claim that $\{g_i = (a^n)^{2i}(b^n)^{4i+1} | i \in \mathbb{N}\}$ is an infinite set we want. Suppose otherwise that g_i and g_j are commensurable up to some conjugation γ where $i < j$. Then we can pick $\operatorname{axis}(g_i) = \gamma \operatorname{axis}(g_j)$. Compare sequences

$$\beta_i = \{\ldots, g_i^{-1}\alpha, \beta, \alpha, a^{2in}\beta = g_i\beta, g_i\alpha, \ldots\},$$

and

$$\gamma\beta_j = \{\ldots, \gamma g_j^{-1}\alpha, \gamma\beta, \gamma\alpha, \gamma a^{2jn}\beta = \gamma g_j\beta, \gamma g_j\alpha, \ldots\}.$$

By Lemma 3.3 (generalized local to global), each element in both sequences 2δ-fellow travel with $\operatorname{axis}(g_i)$ for at least length $2inl_S(a) - M_3$. Because of the choice of n, we know that each element in the sequence β_i 2δ-fellow travel with some element in $\gamma\beta_j$ for longer than the distance in Lemma 1.10, which implies that they indeed share endpoints at the infinity. For a and b generate a free group, the only possibilities are 1) $\alpha = \gamma'\alpha$ and $\beta = \gamma'\beta$, or 2) $\alpha = \gamma'\beta$ and $\beta = \gamma' g_j^{-1}\alpha$ for some γ'. But in either case, β_i and $\gamma\beta_j$ will actually have to be the same sequence, which contradicts with our choice of g_i's. □

Proof of Theorem 4.1. Suppose otherwise, $G = C_1 \cdots C_k$. Pick any $g \in G$. For any $m \in \mathbb{Z}$, denote $g^m = f_{1,m} f_{2,m} \cdots f_{k,m}$ where $f_{i,m} \in C_i$. Take $\Lambda(m) = (f_{1,m}, \ldots, f_{k,m})$ and $\beta_{\Lambda(m)}$ to be the sequence of axes and/or fixed curves. g^m fixes $\beta_{\Lambda(m)}$. Assume the sets of endpoints at infinity of each pair $(A_{i,m}, A_{i+1,m})$ are disjoint. That is, adjacent pairs of pseudo-Anosovs and adjacent pairs of elliptics are not commensurable; if one is elliptic and the other is pseudo-Anosov, then they do not fix the same geodesic in $\operatorname{Teich}(S)$. Project each element in $\beta_{\Lambda(m)}$ to $\beta = \operatorname{axis}(g)$.

For each pseudo-Anosov C_i, name $\beta_i = B_i = \gamma_{i-1}\alpha_i$. Take $x_i^- = \pi_{\beta_i}(B_{i-1})$ and $x_i^+ = \pi\beta_i(B_{i+1})$ for each $1 < i < k+1$.

For each curve stabilizer C_i, name $x_i^- = x_i^+ = B_i$.

Notice that $d_\beta(x_i^+, x_{i+1}^-) \le d(x_i^+, x_{i+1}^-) + D$ and $d(x_i^+, x_{i+1}^-)$ is independent of $f_i \in C_i$. Now we mainly keep track of the projection of $x_i^\pm$.

Therefore,

$$c|m| \quad \le d_\beta(x_1^-, g^m x_1^-) \le \textstyle\sum_i d_\beta(x_i^-, x_i^+) + d_\beta(x_i^+, x_{i+1}^-),$$

where $c > 0$ is the universal constant from [MM99, Proposition 4.6]. The only term on the right hand size that changes with $(f_{i,m})$ is $d_\beta(x_i^-, x_i^+)$ when $f_{i,m}$ is pseudo-Anosov. $d_\beta(x_i^-, x_i^+)$ is related to $l(f_{i,m})$.

For each sufficiently large m, some $d_\beta(x_i^-, x_i^+)$ is large. That is, the axis of some pseudo-Anosov $f_{i,m}$ has large projection on axis(g). With pigeon hole argument, we can find $1 \le i_0 \le k$ and a subsequence $\{m_s\}$ such that

$$\lim_{s \to \infty} \operatorname{diam}_\beta \operatorname{axis}(f_{i_0, m_s}) = \infty.$$

By Lemma 1.10, g and f_{i_0} are commensurable up to conjugacy.

Lemma 4.4 states that there are infinitely many such g while we are only given at most k choices among $C_1, \dots, C_k$. This is a contradiction.

□

5 Applications of Good Word Estimate

In this chapter we apply Theorem 3.4 (Good word estimate) to generate right-angled Artin groups in $\mathcal{MCG}(S)$. The first example is the result of Clay-Leininger-Mangahas and Koberda [CLM10], [Kob10]. In the second example, some high powers of independent non-elliptic elements of $\mathcal{MCG}(S)$ generate a normal subgroup of $\mathcal{MCG}(S)$ which is an infinitely generated right-angled Artin group.

5.1 Finitely generated right-angled Artin group in $\mathcal{MCG}(S)$

We apply Theorem 3.3 to generalize a result by Clay-Leininger-Mangahas and Koberda.

Theorem 5.1 ([CLM10],[Kob10])**.** *Sufficiently large powers of a finite list of independent non-elliptic elements $(g_i)_1^k$ in $\mathcal{MCG}(S)$ generate a right-angled Artin group which quasi-isometrically embedded in $\mathcal{MCG}(S)$.*

Notice that we require in the collection of non-elliptic elements, no two elements which are supported on the same domain are commensurable. Otherwise, suppose axis(g) and axis(h) are fellow travelers. Then $g^m = h^l$ for some $m, l \neq 0$ by Lemma 1.10. Then for any n, $g^{ln} h^{-mn} = \mathrm{Id}$. Moreover, it is possible to find commensurable g and h so that $g^{m/d} h^{-l/d} = \gamma$ is elliptic where $d = \gcd(m, l) \neq 1$. Then for $n = \frac{lm}{d} + 1$, $g^{\frac{m}{d} n} h^{-\frac{l}{d} n} = \gamma$ is elliptic.

Proof. Denote $\alpha_i = \operatorname{axis}(g_i)$ supported on Y_i. In Theorem 3.4, replace M_0 with

$$M_0' = \max\{\operatorname{diam}_{\alpha_i}(\alpha_s) + d_{\alpha_i}(\alpha_s, \alpha_t) + \operatorname{diam}_{\alpha_i}(\alpha_t)\}.$$

There is $N > 0$ such that for any $n > N$,

$$d_{\alpha_i}(x, g_i^n x) > M_0' + M_1$$

for each $1 \leq i \leq k$ and any $x \in \mathcal{C}(S)$ by Proposition 7.6 (Axis), [MM00].

Therefore, every word $w \in \langle g_i^n | 1 \leq i \leq k \rangle$can be rewritten as a good word, and

$$l_{\mathcal{M}}(w) > 0.$$

$\square$

5.2 Infinitely generated normal right-angled Artin subgroups in $\mathcal{MCG}(S)$

Given non-elliptic elements $(g_i)_1^k$ among which every pair of elements are not commensurable up to conjugacy, we may consider the normal subgroup $G = \langle\langle g_i^n | 1 \leq i \leq k \rangle\rangle$. Similar to the previous section, we want to say that each cyclicly reduced word in G can be rewritten as a good word if n is sufficiently large. Two new issues appear for this particular setup.

1. We may find pseudo-Anosov g such that $g^l = (g^l)^\gamma$ for $l > 1$ and that $g(g^\gamma)^{-1}$ is elliptic. In that case, for some $n \neq 0$, not all the words in $\langle\langle g_i^n | 1 \leq i \leq k \rangle\rangle$ can be rewritten as good words.
2. In Theorem 3.4 (Good word estimate), there are only finitely many "initial data", or the axes of the given non-elliptic elements. So

$$M_0 = \max_i \{\mathrm{diam}_{\alpha_i}(\alpha_s) + d_{\alpha_i}(\alpha_s, \alpha_t) + \mathrm{diam}_{\alpha_i}(\alpha_t) | s \in W_i^-, t \in W_i^+\}$$

   is finite. If we apply (Good word estimate) to $\langle\langle g)i^n \rangle\rangle$, the input will be all translates of axes of (g_i), $\{\gamma\, \mathrm{axis}(g_i)\}$. By Lemma 1.10, for the new set up, we can still control $\mathrm{diam}_{\alpha_i}(\alpha_s)$ and $\mathrm{diam}_{\alpha_i}(\alpha_t)$ by making sure the adjacent segments are not commensurable. But in arbitrary word $g_{i_1}^{n_1\gamma_1} \cdots g_{i_\kappa}^{n_\kappa \gamma_\kappa}$ without commensurable adjacent parts, M_0 can be arbitrarily large.

The first issue can be resolved if we only consider some special large n. By acylindricity, there are at most $m = m(\xi(S))$ bi-infinite tight geodesics 2δ-fellow travel with $\mathrm{axis}(g)$ for non-elliptic g in $\mathcal{C}(\mathrm{supp}(g))$. If we take $m' = m!$, then g^m fixes all bi-infinite tight geodesics which 2δ-fellow travel with $\mathrm{axis}(g)$. Also, if $g^l = (g^l)^\gamma$ for some $l \neq 0$, then $g^{m'} = (g^{m'})^\gamma$.

We can circumvent the second issue by requiring for any $1 \le j \le \kappa$,

$$d_{\gamma_{i_j}\alpha_{i_j}}(\gamma_s\alpha_s, \gamma_t\alpha_t), \qquad s \in W_{i_j}^-, t \in W_{i_j}^+,$$

to be bounded in the word $g_{i_1}^{n_1\gamma_1} \cdots g_{i_\kappa}^{n_\kappa\gamma_\kappa}$. Indeed, we can rewrite parts of any cyclic reduced word in $\langle\langle g) i^n \rangle\rangle$ with different conjugates so that this requirement is met.

Theorem 5.2. *Suppose $(g_i)_1^k$ is a collection of non-elliptic elements in which each pair (g_s, g_t) is not commensurable up to conjugacy if $s \ne t$. There exists an integer $m' = m'(\xi(S))$ such that for sufficiently large n the normal subgroup $\langle\langle g_i^{nm'} \rangle\rangle$ has a representation described as follows. The generators are $\{\gamma g_i^{nm'} \gamma^{-1} | 1 \le i \le k, \gamma \in \mathcal{MCG}(S)\}/\sim$ where $\gamma g_i^{nm'}\gamma^{-1} \sim \gamma' g_i^{nm'} \gamma'^{-1}$ if they fix the same end points at boundary of $\mathcal{C}(S)$; that is, if $\gamma\gamma'^{-1}$ commute with g_i^l for some $l \ne 0$. The relations are in the form of commutators of the generators.*

Proof. Name $\alpha_i = \mathrm{axis}(g_i)$ and $Y_i = \mathrm{supp}(g_i)$. From the discussion above, we take $m' = (m+1)!$. Take $L > N(m^2 l_{Y_i}(g_i))^2 + 64\delta$ as in Corollary 1.11. Take N_0 such that for any $n > N_0$,

$$d_{\alpha_i}(x, g_i^n x) > 100L + 2(M + 8\delta) + M_1$$

for every $x \in \mathcal{C}(S)$. This is possible due to Proposition 7.6 (Axis), [MM00]. Here M_1 is from Theorem 3.3 (Generalized local to global). $4m'L + 2(M + 8\delta)$ is the upper bound of projection diameters and distances as M_0 in Theorem 3.4.

Fix any $n > N_0$, we claim that every cyclicly reduced word in $\langle\langle g_i^{nm'} \rangle\rangle$ can be written as a word $g_{i_1}^{l_1 nm'\gamma_1} \cdots g_{i_\kappa}^{l_s nm'\gamma_\kappa}$ in which the adjacent segments do not fellow travel. Moreover, we may assume

$$d_{\gamma_{i_j}\alpha_{i_j}}(\gamma_s\alpha_s, \gamma_t\alpha_t) \le 20L + 16\delta.$$

where $s \in W_{i_j}^-$ and $t \in W_{i_j}^+$.

Otherwise, starting from α_{i_1}, if

$$d_{\gamma_{i_j}\alpha_{i_j}}(\gamma_s\alpha_s, \gamma_t\alpha_t) > 20L + 16\delta,$$

we then conjugate all elements after $g_{i_j}^{\gamma_j}$ by some power p_j of $g_{i_j}^{\gamma_j}$ so that

$$d_{\gamma_{i_j}\alpha_{i_j}}(\gamma_s\alpha_s, (g_{i_j}^{p_j\gamma_j\gamma_t\alpha_t})) \le 2L + 16\delta.$$

Rewrite w as products of powers of $(g_{i_1}^{\gamma_1}, \ldots, g_{i_j}^{\gamma_{i_j}}, \ldots, (g_t)^{g_{i_j}^{p_j \gamma_{i_j}} \gamma_t}, \ldots)$.

Each time we rewrite w, we only pull the initial axes closer together. New word might have more segments but each segments will still satisfy conditions in Lemma 3.3 (Generalized local to global).

We then apply Lemma 3.3 and conclude that $l_{\mathcal{M}}(w) > 0$. □

References

[AAS07] J. W. Anderson, J. Aramayona, and K. J. Shackleton, *Uniformly exponential growth and mapping class groups of surfaces*, In the tradition of Ahlfors-Bers. IV, Contemp. Math., vol. 432, Amer. Math. Soc., Providence, RI, 2007, pp. 1–6. MR 2342801 (2008g:20089)

[BBF10] M. Bestvina, K. Bromberg, and K. Fujiwara, *The asymptotic dimension of mapping class groups is finite*, June 2010.

[Beh06] Jason A. Behrstock, *Asymptotic geometry of the mapping class group and Teichmüller space*, Geom. Topol. **10** (2006), 1523–1578 (electronic). MR 2255505 (2008f:20108)

[BF02] Mladen Bestvina and Koji Fujiwara, *Bounded cohomology of subgroups of mapping class groups*, Geom. Topol. **6** (2002), 69–89 (electronic). MR 1914565 (2003f:57003)

[BF07] ______, *Quasi-homomorphisms on mapping class groups*, Glas. Mat. Ser. III **42(62)** (2007), no. 1, 213–236. MR 2332668 (2008k:57002)

[BH99] Martin R. Bridson and André Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 319, Springer-Verlag, Berlin, 1999. MR MR1744486 (2000k:53038)

[BKMM08] J. Behrstock, B. Kleiner, Y. Minsky, and L. Mosher, *Geometry and rigidity of mapping class groups*, Jan 2008.

[BMM06] Jeff A. Brock, Howard A Masur, and Yair N. Minsky, *Notes on the "local to global lemma"*, circa 2006.

[Bow06] Brian H. Bowditch, *Intersection numbers and the hyperbolicity of the curve complex*, J. Reine Angew. Math. **598** (2006), 105–129. MR 2270568 (2009b:57034)

[Bow08] ______, *Tight geodesics in the curve complex*, Invent. Math. **171** (2008), no. 2, 281–300. MR 2367021 (2008m:57040)

[Can91] James W. Cannon, *The theory of negatively curved spaces and groups*, Ergodic theory, symbolic dynamics, and hyperbolic spaces (Trieste, 1989), Oxford Sci. Publ., Oxford Univ. Press, New York, 1991, pp. 315–369. MR 1130181

[CK83] David Carter and Gordon Keller, *Bounded elementary generation of* $\mathrm{SL}_n(\mathcal{O})$, Amer. J. Math. **105** (1983), no. 3, 673–687. MR 704220 (85f:11083)

[CLM10] M. Clay, C. J. Leininger, and J. Mangahas, *The geometry of right angled Artin subgroups of mapping class groups*, jul 2010.

[Deh87] Max Dehn, *Papers on group theory and topology*, Springer-Verlag, New York, 1987, Translated from the German and with introductions and an appendix by John Stillwell, With an appendix by Otto Schreier. MR 881797 (88d:01041)

[DG] François Dahmani and Vincent Guirardel, *Rotating families, and the structure of some normal subgroups in groups acting on hyperbolic spaces*, Preprint.

[ECH+92] David B. A. Epstein, James W. Cannon, Derek F. Holt, Silvio V. F. Levy, Michael S. Paterson, and William P. Thurston, *Word processing in groups*, Jones and Bartlett Publishers, Boston, MA, 1992. MR 1161694 (93i:20036)

[EMO05] Alex Eskin, Shahar Mozes, and Hee Oh, *On uniform exponential growth for linear groups*, Invent. Math. **160** (2005), no. 1, 1–30. MR 2129706 (2006a:20081)

[Fek23] M. Fekete, *Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten*, Math. Z. **17** (1923), no. 1, 228–249. MR 1544613

[FLM01] Benson Farb, Alexander Lubotzky, and Yair Minsky, *Rank-1 phenomena for mapping class groups*, Duke Math. J. **106** (2001), no. 3, 581–597. MR MR1813237 (2001k:20076)

[FM98] Benson Farb and Howard Masur, *Superrigidity and mapping class groups*, Topology **37** (1998), no. 6, 1169–1176. MR 1632912 (99f:57017)

[FM11] B. Farb and D. Margalit, *A primer on mapping class groups*, Princeton Mathematical Series, Princeton University Press, 2011.

[Fuj09] Koji Fujiwara, *Quasi-homomorphisms on mapping class groups*, Handbook of Teichmüller theory. Vol. II, IRMA Lect. Math. Theor. Phys., vol. 13, Eur. Math. Soc., Zürich, 2009, pp. 241–269. MR 2497785 (2010j:20059)

[GHV91] É. Ghys, A. Haefliger, and A. Verjovsky (eds.), *Group theory from a geometrical viewpoint*, River Edge, NJ, World Scientific Publishing Co. Inc., 1991. MR MR1170362 (93a:20001)

[Gro87] M. Gromov, *Hyperbolic groups*, Essays in group theory, Math. Sci. Res. Inst. Publ., vol. 8, Springer, New York, 1987, pp. 75–263. MR 919829 (89e:20070)

[Har81] W. J. Harvey, *Boundary structure of the modular group*, Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978), Ann. of Math. Stud., vol. 97, Princeton Univ. Press, Princeton, N.J., 1981, pp. 245–251. MR 624817 (83d:32022)

[Iva84] N. V. Ivanov, *Algebraic properties of the Teichmüller modular group*, Dokl. Akad. Nauk SSSR **275** (1984), no. 4, 786–789. MR 745513 (86d:32023)

[Kla99] Erica Klarreich, *The boundary at infinity of the curve complex and the relative Teichmüller space*, http://www.msri.org/people/members/klarreic/curvecomplex.ps, 1999.

[Kob10] T. Koberda, *Right-angled Artin groups and a generalized isomorphism problem for finitely generated subgroups of mapping class groups*, July 2010.

[Man10] Johanna Mangahas, *Uniform uniform exponential growth of subgroups of the mapping class group*, Geom. Funct. Anal. **19** (2010), no. 5, 1468–1480. MR 2585580 (2011d:57002)

[McC85] John McCarthy, *A "Tits-alternative" for subgroups of surface mapping class groups*, Trans. Amer. Math. Soc. **291** (1985), no. 2, 583–612. MR 800253 (87f:57011)

[MM99] Howard A. Masur and Yair N. Minsky, *Geometry of the complex of curves. I. Hyperbolicity*, Invent. Math. **138** (1999), no. 1, 103–149. MR MR1714338 (2000i:57027)

[MM00] H. A. Masur and Y. N. Minsky, *Geometry of the complex of curves. II. Hierarchical structure*, Geom. Funct. Anal. **10** (2000), no. 4, 902–974. MR MR1791145 (2001k:57020)

[Ol′84] A. Yu. Ol′šanskiĭ, *On a geometric method in the combinatorial group theory*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983) (Warsaw), PWN, 1984, pp. 415–424. MR 804697 (86j:20031)

[RS09] Kasra Rafi and Saul Schleimer, *Covers and the curve complex*, Geom. Topol. **13** (2009), no. 4, 2141–2162. MR 2507116 (2010m:57024)

[Sco83] Peter Scott, *The geometries of* 3*-manifolds*, Bull. London Math. Soc. **15** (1983), no. 5, 401–487. MR 705527 (84m:57009)

[Tav90] O. I. Tavgen′, *Bounded generability of Chevalley groups over rings of S-integer algebraic numbers*, Izv. Akad. Nauk SSSR Ser. Mat. **54** (1990), no. 1, 97–122, 221–222. MR 1044049 (91e:20036)

[Thu80] William P. Thurston, *The geometry and topology of three-manifolds*, 1980 lecture notes, http://library.msri.org/books/gt3m/, 1980.

[Tit72] J. Tits, *Free subgroups in linear groups*, J. Algebra **20** (1972), 250–270. MR 0286898 (44 #4105)