CHAPTER 3

Hyperbolicity and spectral gaps

There are two main sources of quasimorphisms: *hyperbolic geometry* (i.e. negative curvature) and *symplectic geometry* (i.e. partial orders and causal structures). In this chapter we study scl in hyperbolic manifolds, and more generally, in word-hyperbolic groups in the sense of Rips and Gromov [**98**] and groups acting on hyperbolic spaces (we return to symplectic geometry, and quasimorphisms with a dynamical or causal origin in Chapter 5). The construction of explicit quasimorphisms is systematized by Bestvina–Fujiwara ([**13**]), who show that in order to construct (many) quasimorphisms on a group G, it suffices to exhibit an isometric action of G on a δ-hyperbolic space X which is *weakly properly discontinuous* (see Definition 3.51). It is crucial for many important applications that X need not be itself proper.

The relationship between negative curvature and quasimorphisms is already evident in the examples from § 2.3.1. If M is a closed hyperbolic manifold, the space of smooth 1-forms $\Omega^1 M$ injects into $Q(\pi_1(M))$. Evidently, quasimorphisms are sensitive to a great deal of the geometry of M; one of the goals of this chapter is to sharpen this statement, and to say what kind of geometry quasimorphisms are sensitive to.

A fundamental feature of the geometry of hyperbolic manifolds is the *thick-thin decomposition*. In each dimension n there is a universal constant $\epsilon(n)$ (the *Margulis constant*) such that the part of a hyperbolic n-manifold M with injectivity radius less than ϵ (i.e. the "thin" piece) has very simple topology — each component is either a neighborhood of a cusp, or a tubular neighborhood of a single short embedded geodesic. Margulis' observation implies that in each dimension, there is a *universal* notion of what it means for a closed geodesic to be *short*.

In this chapter we prove fundamental inequalities relating length to scl in hyperbolic spaces and to show that there is a *universal* notion of what it means for a conjugacy class in a hyperbolic group to have small scl. We think of this as a kind of *homological Margulis Lemma*. These inequalities generalize to certain groups acting on hyperbolic spaces, such as amalgamated free products and mapping class groups of surfaces.

Much of the content in this chapter is drawn from papers of Bestvina, Calegari, Feighn, and Fujiwara (sometimes in combination), especially [**82, 83, 13, 42, 49, 12**].

3.1. Hyperbolic manifolds

We start with the simplest and most explicit examples of groups acting on hyperbolic spaces, namely fundamental groups of hyperbolic manifolds. In this context, scl can be controlled by directly studying maps of surfaces to manifolds.

When we come to study more general hyperbolic spaces, the use of quasimorphisms becomes more practical.

3.1.1. Margulis' Lemma. The most straightforward formulation of Margulis' Lemma is the following:

THEOREM 3.1 (Margulis' Lemma [**123**]). *For each dimension n there is a positive constant $\epsilon(n)$ (called a* Margulis constant*) with the following property. Let Γ be a discrete subgroup of* $\text{Isom}(\mathbb{H}^n)$. *For any $x \in \mathbb{H}^n$ the subgroup $\Gamma_x(\epsilon)$ of Γ generated by elements which translate x less than ϵ is virtually Abelian.*

Here a group is said to *virtually* satisfy some property P if it contains a subgroup of finite index which satisfies P. If Γ is torsion-free, $\Gamma_x(\epsilon)$ is free Abelian. If Γ is cocompact, $\Gamma_x(\epsilon)$ is either trivial or isomorphic to $\mathbb{Z}$. If M is a hyperbolic manifold, there is a so-called *thick-thin* decomposition of M into the *thin part*, namely the subset $M_{<\epsilon}$ consisting of points where the injectivity radius is less than ϵ, and the *thick part*, namely the subset $M_{\ge\epsilon}$ which is the complement of $M_{<\epsilon}$. Margulis' Lemma implies that if M is complete with finite volume, $M_{<\epsilon}$ is a disjoint union of cusps and solid torus neighborhoods of short simple geodesics.

REMARK 3.2. Good estimates for $\epsilon(n)$ are notoriously difficult to obtain. In dimension 2 there is an elementary estimate $\epsilon(2) \ge \text{arcsinh}(1) = 0.8813\cdots$ due to Buser [**37**]. Meyerhoff [**152**] showed $\epsilon(3) \ge 0.104$, and Kellerhals [**124, 125**] showed $\epsilon(n) \ge \sqrt{3}/9\pi = 0.0612\cdots$ for $n = 4, 5$, and obtained an explicit estimate [**126**] for arbitrary n:

$$\epsilon(n) \ge \frac{2}{3^{\nu+1}\pi^\nu}\int_0^{\pi/2} \sin^{\nu+1} t dt$$

where $\nu = [\frac{n-1}{2}]$. The same paper gives explicit lower bounds on the diameter of an embedded tube around a sufficiently short geodesic.

3.1.2. Drilling and Filling. In 3-dimensions, Margulis' Lemma implies that a sufficiently short geodesic is simple, and it can be *drilled* to produce a cusped hyperbolic 3-manifold. That is, the open manifold $M - \gamma$ admits a complete finite-volume hyperbolic structure, defining a hyperbolic manifold M_γ. We denote this suggestively by

$$M \xrightarrow{\text{drill}} M_\gamma$$

Conversely, M can be obtained from M_γ by adding a solid torus under hyperbolic Dehn surgery

$$M_\gamma \xrightarrow{\text{fill}} M$$

The geodesic γ is the core of the added solid torus.

Let $T = \partial N(\gamma)$ be the torus cusp of M_γ. Choose meridian-longitude generators m, l for $H_1(T;\mathbb{Z})$ so that the longitude is trivial in $H_1(M_\gamma;\mathbb{Q})$, and the meridian intersects the longitude once. Note that the meridian is ambiguous, and different choices differ by multiples of the longitude.

Some multiple n of the longitude l is trivial in $H_1(M_\gamma;\mathbb{Z})$ and bounds a surface S. Let p and q be coprime integers, and let $M_{p/q}$ be the result of p/q Dehn surgery on M_γ; i.e. topologically, $M_{p/q}$ is obtained from M_γ by adding a solid torus in such a way that the meridian of the added solid torus represents a primitive class $pm+ql$ in $H_1(T;\mathbb{Z})$. The p, q co-ordinates depend on the choice of meridian m. A change of basis $m \to m + l$ induces $p \to p$ and $q \to q - p$.

Thurston's hyperbolic Dehn surgery Theorem ([**198, 10**]) says that except for finitely many choices of p/q, the manifold $M_{p/q}$ is hyperbolic. Moreover, as p or q or both go to infinity, $\text{length}(\gamma) \to 0$, and the geometry of $M_{p/q}$ converges on compact subsets (in the Gromov–Hausdorff sense) to that of M_γ.

The longitude wraps p times around the core γ of the added solid torus in $M_{p/q}$. Hence ∂S wraps np times. If a denotes the conjugacy class in $\pi_1(M_{p/q})$ corresponding to the free homotopy class of γ, we obtain an estimate

$$\text{scl}(a) \le \frac{-\chi(S)}{2np}$$

In particular, for fixed M_γ, and for any positive constant δ, away from finitely many lines in Dehn surgery space (corresponding to choices p, q for which $|p|$ is small) the core of the added solid torus has scl $< \delta$. Heuristically, *most sufficiently short geodesics in hyperbolic* 3*-manifolds have arbitrarily small* scl.

Conversely, we will see that conjugacy classes in $\pi_1(M)$ with *sufficiently small* scl are represented by arbitrarily short geodesics.

3.1.3. Pleated surfaces. To study scl in $\pi_1(M)$, we need to probe M topologically by maps of surfaces $S \to M$. Under suitable geometric hypotheses, it makes sense to take representative maps of surfaces which are tailored to the geometry of M. For M hyperbolic, a very useful class of maps of surfaces into M are so-called *pleated surfaces*.

Pleated surfaces were introduced by Thurston [**198**].

DEFINITION 3.3. Let M be a hyperbolic manifold. A *pleated surface* is a complete hyperbolic surface S of finite area, together with an isometric map $f : S \to M$ which takes cusps to cusps, and such that every $p \in S$ is in the interior of a straight line segment which is mapped by f to a straight line segment.

Note that the term "isometric map" here means that f takes rectifiable curves on S to rectifiable curves in M of the same length.

The set of points $L \subset S$ where the line segment through p is unique is called the *pleating locus*. It turns out that L is a *geodesic lamination* on S; i.e. a closed union of disjoint simple geodesics. Moreover, the restriction of f to each component of $S - L$ is totally geodesic.

Since S has finite area, L is nowhere dense, and $S - L$ has full measure in S.

EXAMPLE 3.4 (Thurston's spinning construction; § 8.8 and § 8.10 [**198**]). The most important and useful method of producing pleated surfaces is Thurston's *spinning* construction.

Let P be a pair of pants; i.e. a hyperbolic surface with three boundary components. Let $f : P \to M$ be a relative homotopy class of map sending the three boundary components by maps of nonzero degree to three (not necessarily simple) geodesics in M. The class of f determines a homomorphism from $\pi_1(P)$ to $\pi_1(M)$ up to conjugacy.

We give P a hyperbolic structure, and let Δ be a geodesic triangle in P with one vertex on each boundary component. As we move the vertices around on ∂P, the geodesic triangle deforms continuously. *Spinning* Δ involves dragging the vertices around and around the components of ∂P. The sides of Δ get longer and longer, and accumulate on ∂P. The Hausdorff limit of $\partial \Delta$ is a geodesic lamination L in P

with three infinite leaves spiraling around ∂P, and whose complement consists of two (open) ideal triangles. See Figure 3.1.

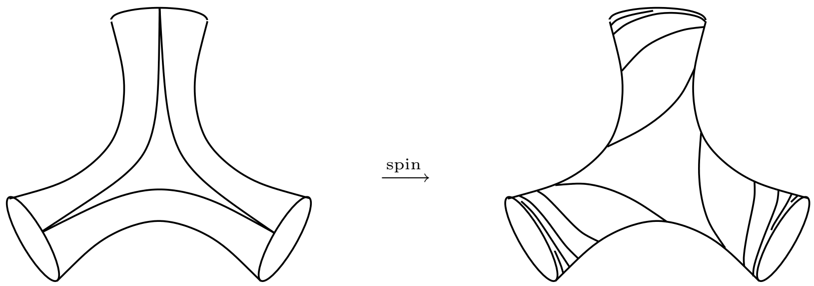

FIGURE 3.1. Spinning produces an ideal triangulation of a pair of pants

We can build a pleated surface $\overline{f} : P \to M$ in the homotopy class of f with pleating locus contained in L as follows. First, $\overline{f}$ takes components of ∂P to the unique closed geodesics in the homotopy class of $f(\partial P)$. For each infinite geodesic $l \in L$, the ends of $f(l)$ spiral around $f(\partial P)$. Except in degenerate cases, the image of $f(l)$ is a quasigeodesic which can be straightened to a unique geodesic $\overline{l}$ which spirals around two components of $\overline{f}(\partial P)$. This defines the map $\overline{f}$ on L. Each component of $P - L$ is an ideal triangle, and we define $\overline{f}$ on each such triangle Δ to be the unique totally geodesic map which extends $\overline{f}$ (after possibly reparameterizing by a translation on each edge) on $\partial\Delta \subset L$.

REMARK 3.5. If M has parabolic elements, the construction in Example 3.4 must be modified very slightly.

Suppose $f : P \to M$ takes some boundary component ∂_0 of P to a free homotopy class in M corresponding to a parabolic conjugacy class α in $\pi_1(M)$. After lifting $\tilde{f} : \tilde{P} \to \tilde{M}$, each conjugate of α fixes a unique point in the sphere at infinity S^2_∞. If Δ is a triangle with a vertex v on ∂_0, and $\tilde{\Delta}$ is a lift of Δ to $\tilde{P}$, straighten $\tilde{f}$ on $\tilde{v}$ by sending this vertex to the unique fixed point of the corresponding conjugate of α. The rest of the construction is as before.

LEMMA 3.6 (Thurston, § 8.10 [**198**]). *A map $f : P \to M$ from a pair of pants into a hyperbolic manifold M can be straightened to a pleated surface unless it factors through a map to a circle.*

PROOF. The map $f : P \to M$ determines an equivariant map $\tilde{P} \to \mathbb{H}^n$ from the universal cover $\tilde{P}$ of P. A lift of the triangle Δ has vertices on three distinct edges e_1, e_2, e_3 of $\tilde{P}$. Spinning drags the vertices of Δ to endpoints of the e_i, so f can be straightened on Δ providing the endpoints of the e_i are distinct for different i. If $\alpha, \beta \in \pi_1(M)$ don't commute, their axes have disjoint endpoints at infinity. Commuting elements in a closed hyperbolic manifold group generate a cyclic group. So the straightening can be achieved if and only if the image of $\pi_1(P)$ in $\pi_1(M)$ does not factor through a cyclic group. $\square$

Using this lemma, we show that a map $f : S \to M$ either admits an obvious simplification which reduces the genus, or has a pleated representative.

LEMMA 3.7. *Let M be a hyperbolic manifold, and let a be a nontrivial conjugacy class in $\pi_1(M)$. Let S be a compact oriented surface with exactly one boundary* *component. If $f : S \to M$ is a map sending the class of ∂S in $\pi_1(S)$ to a then there is another map $f' : S' \to M$ where the genus of S' is no more than that of S, and where f' sends the class of $\partial S'$ in $\pi_1(S')$ to a, which is homotopic to a pleated representative.*

PROOF. We decompose S into subsurfaces $S = S_1 \cup S_2 \cup \cdots \cup S_g$ where each S_i is a twice-punctured torus for $i < g$, and S_g is a once-punctured torus. For each i denote the two boundary components of S_i by $\gamma_i^\pm$ where γ_i^+ is glued to γ_{i+1}^- in S, and γ_1^- maps to a by f.

If any $\gamma_i^\pm$ maps by f to an inessential loop in M, we can compress f, S along the image of this curve, sewing in two disks, to produce a map $f' : S' \to M$ where S' is of smaller genus than S, and for which $f'(\partial S')$ is in the class of a. After finitely many compressions of this kind, we assume that every $\gamma_i^\pm$ maps by f to an essential loop in M.

For each $i < g$, let $\alpha_i^\pm, \beta_i$ be embedded essential loops in S_i as in Figure 3.2:

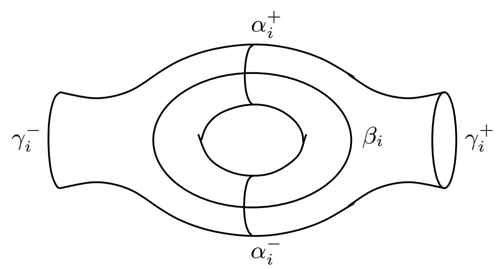

FIGURE 3.2. The curves $\alpha_i^\pm$ and β in S_i

The loops α_i^+, β_i intersect in one point p_i^+. Their images under f define elements a^+, b of $\pi_1(M)$ based at $f(p_i^+)$. If $[a^+, b] = \text{id}$ then we can compress S, cutting out a neighborhood of $\alpha_i^+ \cup \beta_i$ in S_i, and sewing in a disk, thereby reducing the genus of S. So without loss of generality, we assume the elements a^+, b do not commute.

The curves α_i^- and β_i intersect at a different point. By sliding this point half way around β_i and mapping by f, we obtain an element $a^- \in \pi_1(M)$ based at $f(p_i^+)$ if we slide in one direction, and ba^-b^{-1} if we slide in the other direction. Also without loss of generality, we assume a^-, b do not commute.

Let P_i^-, P_i^+ be the two pairs of pants obtained from S_i by cutting along $\alpha_i^+ \cup \alpha_i^-$. Suppose $f : P_i^- \to M$ factors up to homotopy through a map to a circle. This happens if a^+, a^- as above generate a cyclic subgroup of $\pi_1(M)$. In this case, we replace $\alpha_i^\pm$ by their images under a Dehn twist around β_i (see Example 3.59 for a definition). At the level of π_1, this replaces a^+, a^- by $a^+b, b^{-1}a^-$, and defines a new pair of pants decomposition in which $f : P_i^- \to M$ does not factor up to homotopy through a map to a circle. Of course, now $f : P_i^+ \to M$ might factor through a circle, in which case we do another Dehn twist, replacing the original a^+, a^- by $a^+b^2, b^{-2}a^-$. In this way we obtain a decomposition of S_i into two pairs of pants such that the restriction of f to either does not factor up to homotopy through a map to a circle. By Lemma 3.7 f can be replaced by a pleated representative on each such pair of pants, and we are done.

The construction of a pleated representative on S_g is similar but simpler, with $\alpha_i^\pm$ being replaced by a single α_g. $\square$

Remark 3.8. Other surfaces which perform a similar function include *harmonic* (maps of) surfaces and *minimal surfaces*. The use of one kind of surface or the other is often a matter of taste. One technical advantage of pleated surfaces is that they generalize in some sense to arbitrary δ-hyperbolic groups; see Definition 3.39.

3.2. Spectral Gap Theorem

Let M be a closed hyperbolic manifold. There is a natural bijection between the set of conjugacy classes in $\pi_1(M)$ and the set of closed geodesics in M. It is a fundamental fact that the function

$$\{\text{closed geodesics}\} \xrightarrow{\text{length}} \mathbb{R}$$

which assigns to a closed geodesic its length, is *proper*; i.e. there are only finitely many closed geodesics with length bounded above by any constant. By contrast, if $G = \pi_1(M)$, the function

$$\{\text{conjugacy classes in } [G,G]\} \xrightarrow{\text{scl}} \mathbb{R}$$

which assigns to a (homologically trivial) conjugacy class its stable commutator length, is *not* proper: i.e. there are always infinitely many distinct conjugacy classes with uniformly bounded stable commutator length. However, some vestige of properness holds in this context. If the stable commutator length of a conjugacy class is *sufficiently* small, the length of the corresponding geodesic must also be (comparably) small. This implies that at least for sufficiently small ϵ, the preimage $\text{scl}^{-1}([0,\epsilon])$ is finite. One can define δ_∞ to be the supremum of the set of ϵ with this property; it turns out that there is a universal estimate $\frac{1}{12} \le \delta_\infty \le \frac{1}{2}$.

3.2.1. Length inequality. We now show that in a hyperbolic manifold group, a conjugacy class with sufficiently small stable commutator length is represented by an arbitrarily short geodesic. The material in this section is largely drawn from § 6 of [**42**].

Theorem 3.9 (Length inequality). *For every dimension m and any $\epsilon > 0$ there is a positive constant $\delta(\epsilon, m)$ such that if M is a complete hyperbolic m-manifold, and a is a conjugacy class in $\pi_1(M)$ with $\text{scl}(a) \le \delta(\epsilon, m)$ then if a is represented by a geodesic γ, we have*

$$\text{length}(\gamma) \le \epsilon$$

Proof. Let S be a surface of genus g with one boundary component, and $f : S \to M$ a map wrapping ∂S homotopically n times around γ. By Lemma 3.7, after possibly reducing the genus of S if necessary, we can assume without loss of generality that f, S is a pleated surface. This determines a hyperbolic structure on S with geodesic boundary for which the map f is an isometry on paths. In particular, $\text{length}(\partial S) = n \cdot \text{length}(\gamma)$ and $\text{area}(S) = -2\pi\chi(S) = (4g-2)\pi$ by Gauss–Bonnet.

Choose ϵ which is small compared to the 2-dimensional Margulis constant $\epsilon(2)$. We defer the precise choice of ϵ for the moment. Consider the thick-thin decomposition of S with respect to 2ϵ in the sense of § 3.1.1. More precisely, let DS denote the double of S (which is a closed hyperbolic surface), Let DS_{thick} and DS_{thin} denote the subsets of DS where the injectivity radius is $\ge 2\epsilon$ and $< 2\epsilon$ respectively, and define S_{thick} and S_{thin} to be equal to $DS_{\text{thick}} \cap S$ and $D_{\text{thin}} \cap S$ respectively.

The set S_{thin} is a union of open embedded annuli around very short simple geodesics, together with a union of open embedded rectangles which run between pairs of segments of ∂S which are distance $< \epsilon$ apart at every point. Each rectangle doubles to an annulus in DS_{thin}. If there are s annuli and r rectangles in S_{thin}, then there are $2s + r$ annuli in DS_{thin}. Components of DS_{thin} are disjoint and pairwise non-isotopic. Any maximal collection of disjoint pairwise non-isotopic simple closed curves in a closed orientable surface of negative Euler characteristic must decompose the surface into pairs of pants. Since the genus of DS is $2g$, we estimate $2s+r \le -\frac{3}{2}\chi(DS) = 6g-3$. Hence r, the number of rectangle components of S_{thin}, is at most $6g-3$.

By abuse of notation, we add to S_{thick} the annulus components of S_{thin} (if any), so that S_{thin} consists exactly of the set of thin rectangles running between pairs of arcs in ∂S. With this new definition, a point $p \in \partial S$ is in S_{thick} if and only if the length of an essential arc in S from p to ∂S is at least ϵ. In particular, the $\epsilon/2$ neighborhood of $\partial S \cap S_{\text{thick}}$ is *embedded*, and there is an estimate

$$(4g-2)\pi = \text{area}(S) \ge \text{area}(S_{\text{thick}}) \ge \frac{\epsilon}{2}\text{length}(\partial S \cap S_{\text{thick}})$$

Since there are at most $6g-3$ components of S_{thin}, and each component intersects ∂S in two arcs, there are at most $12g-6$ components of $\partial S \cap S_{\text{thin}}$. But

$$\text{length}(\partial S \cap S_{\text{thin}}) = \text{length}(\partial S) - \text{length}(\partial S \cap S_{\text{thick}}) \ge n \cdot \text{length}(\gamma) - (8g-4)\frac{\pi}{\epsilon}$$

where we used $\text{length}(\partial S) = n \cdot \text{length}(\gamma)$ and the previous inequality. It follows that there is at least one arc σ of $\partial S \cap S_{\text{thin}}$ satisfying

$$\text{length}(\sigma) \ge \frac{n \cdot \text{length}(\gamma) - (8g-4)\pi/\epsilon}{12g-6} = \frac{n \cdot \text{length}(\gamma)}{12g-6} - \frac{2\pi}{3\epsilon}$$

Hence S_{thin} contains a component R which is a rectangular strip of thickness $\le \epsilon$ with σ on one side. We denote the side opposite to σ by σ'. We call σ and σ' the *long* sides of R. Because S is oriented, the orientations on opposite sides of R are "anti-aligned". We lift R to the universal cover $\mathbb{H}^n$, and by abuse of notation refer to the lifted rectangle as R. The sides σ, σ' of R are contained in geodesics l, l' that cover γ. Without loss of generality, we can suppose that l is an axis for a, and l' is an axis for bab^{-1} where $b(l) = l'$. Moreover, the action of a on l and a' on l' move points is (nearly) opposite directions.

Let p be the midpoint of σ, and let q be a point on the opposite side of R with $d(p,q) < \epsilon$. Suppose further that

$$\text{length}(\sigma) = \text{length}(\sigma') > 2 \cdot \text{length}(\gamma) + 4\epsilon$$

It follows that $bab^{-1}(q) \in \sigma'$ and there is $r \in \sigma$ with $d(bab^{-1}(q), r) \le \epsilon$ and therefore $d(bab^{-1}(p), r) \le 2\epsilon$. Since $d(q, bab^{-1}(q)) = \text{length}(\gamma)$,

$$|d(p,r) - \text{length}(\gamma)| \le 2\epsilon$$

and therefore $d(p, a(r)) \le 2\epsilon$ and we can estimate

$$d(p, abab^{-1}(p)) \le 4\epsilon$$

Similarly we estimate $d(p, bab^{-1}a(p)) \le 4\epsilon$. See Figure 3.3. In the figure, the axes l and l' are both roughly vertical. The element a translates points roughly downwards along l, and bab^{-1} translates points roughly upwards along l'.

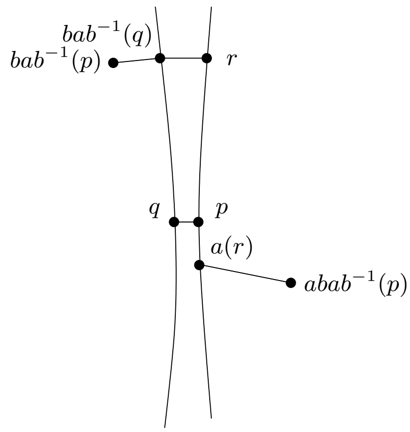

FIGURE 3.3. The composition $abab^{-1}$ translates the midpoint p a small distance

If we choose 4ϵ less than an m-dimensional Margulis constant $\epsilon(m)$ then $abab^{-1}$ and $bab^{-1}a$ must commute. There are two possibilities, which break up into subcases.

CASE ($abab^{-1}$ and $bab^{-1}a$ are hyperbolic with the same axis). In this case, since they are conjugate, they are either equal or inverse.

SUBCASE ($abab^{-1} = bab^{-1}a$). In this case a and bab^{-1} commute, and since they are conjugate, they are equal or inverse. But a and bab^{-1} translate their respective axes in almost opposite directions, so they cannot be equal; hence we must have $bab^{-1} = a^{-1}$ and therefore b has order 2, which is impossible in a hyperbolic manifold group.

SUBCASE ($abab^{-1} = a^{-1}ba^{-1}b^{-1}$). In this case $a^2 = ba^{-2}b^{-1}$ and therefore b has order 2, which is impossible as we already remarked.

CASE ($abab^{-1}$ and $bab^{-1}a$ parabolic with the same fixed point). $z \in S^{n-1}_\infty$. In this case, $a^{-1}(abab^{-1})a$ is parabolic with fixed point $a^{-1}(z)$. But $a^{-1}(abab^{-1})a = bab^{-1}a$ which has fixed point z, so $a^{-1}(z) = z$. Since a translates along an axis, it is hyperbolic, and we have obtained a hyperbolic and a parabolic element in $\pi_1(M)$ with a common fixed point at infinity. This is well-known to violate discreteness, see for instance Maskit [**147**], p. 19 for details.

In every case we obtain a contradiction, and therefore we must have

$$2 \cdot \text{length}(\gamma) + 4\epsilon \geq \text{length}(\sigma)$$

Putting this together with our earlier inequality, we obtain

$$2 \cdot \text{length}(\gamma) + 4\epsilon \geq \frac{n \cdot \text{length}(\gamma)}{12g - 6} - \frac{2\pi}{3\epsilon}$$

Rearranging this gives

$$\text{length}(\gamma) \cdot \left(\frac{n}{12g - 6} - 2 \right) \leq 4\epsilon + \frac{2\pi}{3\epsilon}$$

The right hand side is a constant which depends only on the size of a Margulis constant in dimension n. If scl is very small, we can make n/g very large, and therefore obtain an upper bound on length(γ) which goes to 0 as scl $\to 0$ as claimed. □

REMARK 3.10. Note that when $m < n$ a hyperbolic m-manifold group is also a hyperbolic n-manifold group, and therefore $\delta(\epsilon, m) \geq \delta(\epsilon, n)$. In § 3.3 we will see that for small ϵ there are estimates

$$\delta(\epsilon, 3) = O(\epsilon^{1/2})$$

and

$$\delta(\epsilon, 3) \geq \delta(\epsilon, n) \geq O(\epsilon^{(n-1)/(n+1)})$$

in any fixed dimension n.

On the other hand, the dependence of δ on ϵ is not proper. In particular, as $\epsilon \to \infty$, the constant $\delta(\epsilon, n)$ is bounded above by some finite bound, independent of dimension n. This universal upper bound should be thought of as a kind of homological Margulis constant. In the next subsection, we will give an explicit estimate for this constant.

3.2.2. Spectral Gap.

THEOREM 3.11 (Spectral Gap Theorem). *Let M be a closed hyperbolic manifold, of any dimension ≥ 2. Let $\delta_\infty(M)$ be the first accumulation point for stable commutator length on conjugacy classes in $\pi_1(M)$. That is, $\delta_\infty(M)$ is the smallest number such that for any $\delta < \delta_\infty(M)$, there are only finitely many conjugacy classes a in $\pi_1(M)$ with* $\mathrm{scl}(a) \leq \delta$. *Then*

$$\frac{1}{12} \leq \delta_\infty(M) \leq \frac{1}{2}$$

PROOF. We use the same setup and notation as in the proof of Theorem 3.9. Since M is a closed hyperbolic manifold, there are only finitely many conjugacy classes represented by geodesics shorter than any given length. So we suppose a is a conjugacy class represented by a geodesic γ which is "sufficiently long" (in a sense to be made precise in a moment). We choose ϵ and find a segment σ, as in the proof of Theorem 3.9, and suppose we have

$$\mathrm{length}(\gamma) + 4\epsilon < \mathrm{length}(\sigma)$$

(note the missing factor of 2). We choose p to be one of the endpoints of σ, so that

$$d(p, abab^{-1}(p)) \leq 4\epsilon$$

Since M is fixed, there is some ϵ such that 4ϵ is smaller than the translation length of any nontrivial element in $\pi_1(M)$. Hence $abab^{-1} = \mathrm{id}$. But this means $bab^{-1} = a^{-1}$, and b has order 2, which is impossible in a manifold group.

Contrapositively, this means that we must have

$$\mathrm{length}(\gamma) + 4\epsilon \geq \mathrm{length}(\sigma)$$

and therefore, just as in the proof of Theorem 3.9, we obtain

$$\mathrm{length}(\gamma) \cdot \left(\frac{n}{12g-6} - 1\right) \leq 4\epsilon + \frac{2\pi}{3\epsilon}$$

In contrast to the case of Theorem 3.9, the right hand side definitely depends on the manifold M. Nevertheless, for fixed M, it is a constant, and we see that for γ

sufficiently long, g/n cannot be much smaller than $1/12$. This establishes the lower bound in the theorem.

We now establish the upper bound. M is a closed hyperbolic manifold, and therefore $\pi_1(M)$ contains many nonabelian free groups. In fact, if a, b are arbitrary noncommuting elements of $\pi_1(M)$, sufficiently high powers of a and b generate a free group, by the ping-pong lemma. This copy of F_2 is quasi-isometrically embedded, and by passing to a subgroup, one obtains quasi-isometrically embedded copies of free groups of any rank.

For each n, the element $[a^n, b^n]$ is in the commutator subgroup. In fact, it is a commutator, and therefore satisfies $\mathrm{scl}([a^n, b^n]) \leq 1/2$. In a free group, the words $x^n y^n x^{-n} y^{-n}$ are cyclically reduced of length $4n$. Since the embedding is quasi-isometric, the geodesic representatives of $[a^n, b^n]$ have length which goes to infinity linearly in n. It follows that these elements fall into infinitely many conjugacy classes, and the upper bound is established. $\square$

Remark 3.12. From the method of proof one sees for sufficiently long γ that if no translate l' of l is ϵ-close and anti-aligned with l along segments σ, σ' whose length is at least $(\lambda + \epsilon) \cdot \mathrm{length}(\gamma)$ then $\mathrm{scl}(a) \geq \frac{1}{12\lambda}$.

For example, in a free group, a cyclically reduced word w and a conjugate of its inverse cannot share a subword of length longer than $\frac{1}{2}\mathrm{length}(w)$. This leads to an estimate $\mathrm{scl}(a) \geq 1/6$ in a free group, which is not yet optimal, but is still an improvement (a sharp bound $\mathrm{scl}(a) \geq 1/2$ in a free group will be established in Theorem 4.111).

An estimate on the size of anti-aligned translates is essentially a kind of macroscopic small cancellation property. One can give an alternative proof of Theorem 3.11 along these lines using generalized small cancellation theory (see [**65**] for more details). For certain groups, ordinary small cancellation theory can be applied, leading to sharp results; we will discuss this approach in § 4.3.

3.3. Examples

3.3.1. Hyperbolic Dehn surgery. We elaborate on the discussion in § 3.1.2.

Lemma 3.13. *Let M be a hyperbolic 3-manifold, and let γ be a geodesic loop which is the core of an embedded solid torus of radius T. Then there is a 1-form α supported in the tube of radius T about γ, with $\int_\gamma \alpha = \mathrm{length}(\gamma)\sinh(T)$ and $\|d\alpha\| \leq 1 + 1/(T - \epsilon)$ for any $\epsilon > 0$.*

Proof. Let S be the solid torus of radius T about γ. On S, let $r : S \to \mathbb{R}$ be the function which measures distance to γ. Denote radial projection to γ by

$$p : S \to \gamma$$

Parameterize γ by θ, so that $d\theta$ is the length form on γ, and $\int_\gamma d\theta = \mathrm{length}(\gamma)$. Pulling back by p extends θ and $d\theta$ to all of S. We define

$$\alpha = d\theta \cdot (\sinh(T) - \sinh(r))$$

on S, and extend it by 0 outside S. Notice that

$$\|d\theta\| = 1/\cosh(r)$$

on S. By direct calculation, $d\alpha = \cosh(r) d\theta \wedge dr$ on S, so $\|d\alpha\| = 1$ at every point of S.

The form α is not smooth along ∂S, but it is Lipschitz. Let $\beta_\epsilon(r)$ be a C^∞ function on $[0, T]$ taking the value 1 in a neighborhood of 0 and the value 0 in a neighborhood of T, and with $|\beta'_\epsilon| < 1/(T-\epsilon)$ throughout, for some small ϵ. The product $\alpha_\epsilon := \beta_\epsilon(r)\alpha$ is C^∞ and satisfies

$$d\alpha_\epsilon = d\theta \wedge dr(\beta_\epsilon(r)\cosh(r) + \beta'_\epsilon(r)\sinh(r))$$

so $\|d\alpha_\epsilon\| \le 1 + 1/(T-\epsilon)$. □

As in § 2.3.1 there is a de Rham quasimorphism q_α associated to α by integration over based geodesic representatives of elements, after choosing a basepoint. The homogenization of q_α is obtained by integrating α over free geodesic loops. A limit of such quasimorphisms as $\epsilon \to 0$ has defect at most $2\pi(T+1)/T$ by Lemma 2.58.

In order for Lemma 3.13 to be useful, we need a good estimate of T in terms of length(γ).

Lemma 3.14 (Hodgson–Kerckhoff, p. 403 [**111**]). *Let S be a Margulis tube in a hyperbolic 3-manifold. Let T be the radius of S and* length(γ) *the length of the core geodesic. Then there is an estimate*

$$\text{length}(\gamma) \ge 0.5404\frac{\tanh(T)}{\cosh(2T)}$$

Note for γ sufficiently small this implies $e^T \ge 1.03\,\text{length}^{-1/2}(\gamma)$.

Remark 3.15. In any dimension n a much cruder argument due to Reznikov [**177**] shows that for sufficiently small γ there is a constant C_n such that $e^T \ge C_n\,\text{length}^{-2/(n+1)}(\gamma)$.

Now fix M, a 1-cusped hyperbolic 3-manifold. Fix generators m, l for $H_1(\partial M)$ for which l generates the kernel of $H_1(\partial M;\mathbb{Q}) \to H_1(M;\mathbb{Q})$. Let $M_{p/q}$ denote the result of p/q Dehn surgery on M in these co-ordinates, and let $\gamma(p/q)$, or just γ for short, denote the core geodesic of the filled solid torus.

Theorem 3.16. *Let $M_{p/q}$ be the result of p/q surgery on M. Suppose $M_{p/q}$ is hyperbolic. When the core geodesic γ is contained in a Margulis tube of radius at least T then*

$$\text{length}(\gamma) \le \left(\frac{7.986\pi\,\text{scl}(l)(T+1)}{Tp}\right)^2$$

Proof. By Lemma 3.13 there is a homogeneous quasimorphism q_α on $\pi_1(M_{p/q})$ with defect at most 2π, and satisfying

$$q_\alpha(\gamma) \ge \text{length}(\gamma)\sinh(T)\frac{T}{T+1}$$

On the other hand, the conjugacy class of γ^p contains the image of l under the surjective homomorphism $\pi_1(M_\gamma) \to \pi_1(M_{p/q})$ induced by Dehn surgery, so by the easy direction of Bavard's Duality Theorem 2.70, we estimate

$$\frac{q_\alpha(\gamma)}{4\pi} \le \text{scl}(\gamma) \le \frac{\text{scl}(l)}{p}$$

Using the estimate from Lemma 3.14, a straightforward calculation gives the desired conclusion. □

Neumann–Zagier [**161**] introduce the following quadratic form Q:

$$Q(p,q) = \frac{(\text{length of } pm + ql)^2}{\text{area}(\partial S)}$$

Here ∂S is the horotorus boundary of the cusp of M, and $pm + ql$ is a straight curve on the horotorus (in the intrinsic Euclidean metric) representing $m^p l^q$. Equivalently, if we scale the Euclidean cusp to have area 1, the form just becomes $Q(p, q) = \text{length}^2(pm + ql)$.

LEMMA 3.17 (Neumann–Zagier, Prop. 4.3 [**161**]). *With notation as above, in the manifold $M_{p/q}$ there is an estimate*

$$\text{length}(\gamma) = 2\pi Q(p,q)^{-1} + O\left(\frac{1}{p^4 + q^4}\right)$$

In particular, for q fixed, there is an estimate

$$\lim_{p\to\infty} (pm)^2 \text{length}(\gamma)/2\pi = 1$$

where m is the length of the meridian in the Euclidean cusp, normalized to have area 1.

REMARK 3.18. We see from Lemma 3.17 that the estimates obtained in Theorem 3.16 are sharp, up to an order of magnitude. Together with Remark 3.15, this justifies the claims made in Remark 3.10.

THEOREM 3.19. *Let M be a 1-cusped hyperbolic manifold, with notation as above. Normalize the Euclidean structure on the cusp ∂S to have area 1, and let m be the length of the shortest curve on ∂S which is homologically essential in M. If* $\text{length}(m) < 1$ *then*

$$\text{scl}(l) \geq \frac{1}{4\pi\, \text{length}(m)^2}$$

PROOF. For brevity, we denote (normalized) length(m) by m. We expand S to a maximal horotorus. For a maximal horotorus, every essential slope on ∂S has length at least 1, by Jørgensen's inequality [**147**]. It follows that if $m < 1$, then $\text{area}(\partial S) \geq 1/m^2$. Under p/q surgery for very large p, the area of the boundary of a maximal embedded tube around γ is almost equal to that of $\text{area}(\partial S)$. The boundary of such a tube is intrinsically Euclidean in its induced metric, and is isometric to a torus obtained from a product annulus by gluing the two end components with a twist. The boundary components of the annulus have have length equal to the circumference of a circle in the hyperbolic plane of radius T, which is $2\pi\sinh(T)$. By elementary hyperbolic trigonometry, the height of the annulus is equal to $\text{length}(\gamma)\cosh(T)$. Hence the area of the boundary of the tube is $2\pi\,\text{length}(\gamma)\sinh(T)\cosh(T)$.

So we can estimate

$$\text{area}(\partial S) = \lim_{p\to\infty} 2\pi\, \text{length}(\gamma)\sinh(T)\cosh(T)$$

and therefore

$$e^T \geq \frac{\sqrt{2}}{m\sqrt{\pi}}\, \text{length}^{-1/2}(\gamma)$$

Using this estimate in the place of Lemma 3.14 in Theorem 3.16, and applying Lemma 3.17, we obtain

$$\frac{2\pi}{(pm)^2} = \lim_{p\to\infty} \text{length}(\gamma) \leq \left(\frac{4\text{scl}(l)m\pi\sqrt{2\pi}}{p}\right)^2$$

and therefore
$$\mathrm{scl}(l) \geq \frac{1}{4\pi m^2}$$
as claimed ∎

In other words, one can estimate $\mathrm{scl}(l)$ from below from the geometry of the cusp.

3.3.2. Manifolds with small δ_∞. Note that the proof of Theorem 3.11 actually shows that if M is any closed hyperbolic manifold, and a is a conjugacy class in $\pi_1(M)$ represented by a geodesic γ, then if $\mathrm{length}(\gamma)$ is sufficiently long, $\mathrm{scl}(a) \geq \delta$ for any $\delta < 1/12$.

EXAMPLE 3.20. Let S be a closed nonorientable surface with $\chi(S) = -1$. A presentation for $\pi_1(S)$ is
$$\langle a, b, c \mid [a, b] = c^2\rangle$$
so the conjugacy class of c satisfies $\mathrm{scl}(c) \leq 1/4$. On the other hand, for a suitable choice of hyperbolic structure on S, the geodesic in the free homotopy class of c can be arbitrarily long.

QUESTION 3.21. *What are the optimal constants in Theorem 3.11?*

We will see in § 4.3.4 that the upper bound of $1/2$ is sharp, and is realized in free and orientable surface groups.

EXAMPLE 3.22. For any group G and any elements $a, b \in G$ the element $[a, b]$ satisfies $\mathrm{scl}([a, b]) \leq 1/2$. Moreover, by Proposition 2.104, if a and b do not generate a free rank 2 subgroup of G, we must have $\mathrm{scl}([a, b]) < 1/2$.

However, a theorem of Delzant [**64**] shows that in any word-hyperbolic group G (see § 3.4 for a definition) there are only finitely many conjugacy classes of non-free 2-generator subgroups. Note that this class of groups includes fundamental groups of closed hyperbolic manifolds of any dimension. Therefore only finitely many conjugacy classes of elements $[a, b]$ with $\mathrm{scl}([a, b]) < 1/2$ can be constructed in a fixed hyperbolic group G this way.

3.3.3. Complex length. If M is a closed hyperbolic 3-manifold, a conjugacy class $a \in \pi_1(M)$ determines a geodesic γ which has a complex length, denoted $\mathrm{length}_\mathbb{C}(\gamma)$, defined as follows. The hyperbolic structure on M determines a representation $\rho : \pi_1(M) \to \mathrm{PSL}(2, \mathbb{C})$. The trace $\mathrm{tr}(a)$ is well-defined up to multiplication by ± 1. We set
$$\mathrm{length}_\mathbb{C}(\gamma) = \cosh^{-1}(\mathrm{tr}(a)/2)$$
which is well-defined up to integral multiples of $2\pi i$. The real part of $\mathrm{length}_\mathbb{C}$ is the usual length of γ, and the imaginary part is the angle of rotation on the normal bundle $\nu\gamma$ to γ induced by parallel transport around γ.

If γ is trivial in $H_1(M; \mathbb{Q})$ there is a slightly different $\mathbb{C}$-valued complex length, denoted $\mathrm{length}_H(\gamma)$, and defined as follows. Let $N(\gamma)$ be an open solid torus neighborhood of γ, and let T be the torus boundary of $M - N(\gamma)$. Let l be the slope on T which generates the kernel of the map $H_1(T; \mathbb{Q}) \to H_1(M - N(\gamma); \mathbb{Q})$. If M is obtained from $M - N(\gamma)$ by p/q filling with respect to some basis (q is arbitrary, depending on a choice of meridian m, but p is well-defined) then l determines a framing of $\nu\gamma^p$, the normal bundle of the p-fold cover of γ. This framing determines the imaginary part of $\mathrm{length}_H(\gamma)$; in words, minus the imaginary part is the angle that the framing l twists relative to parallel transport around γ. Note that the imaginary part of $\text{length}_H(\gamma)$ and the imaginary part of $\text{length}_{\mathbb{C}}(\gamma)$ will only agree up to integral multiples of $2\pi i/p$. Note that for a fixed choice of meridian m one can estimate

$$\text{imaginary part of length}_H(\gamma) = \text{some function of } p, q + O(1)$$

In terms of differential forms: near γ we can define cylindrical co-ordinates θ, ϕ, r where θ parameterizes the length along γ and r is distance to γ, as in Lemma 3.13, and where ϕ is the angular co-ordinate, taking values (locally) in $\mathbb{R}/2\pi\mathbb{Z}$. The co-ordinate ϕ is not globally well-defined unless $\text{length}_{\mathbb{C}}(\gamma)$ has imaginary part which is a multiple of $2\pi i$, but the forms $d\phi$ and $d\theta$ are well-defined. With respect to this co-ordinate system,

$$\text{length}_H(\gamma) = \frac{1}{p}\int_l d\theta + i d\phi$$

In analogy to the construction in Lemma 3.13, define $\beta = d\phi(\cosh(T) - \cosh(r))$ and observe that

$$\|d(\alpha + i\beta)\| = 1$$

on $T - \gamma$. One must be careful, since $d\phi$ does not extend over γ. Nevertheless, if S is a surface in $M - N(\gamma)$ of genus g whose boundary wraps m times around l, we can represent S by a pleated surface in M. For sufficiently large p or q the length (in the usual sense) of γ will be very short, and any surface S which intersects γ transversely will have area at least e^T. In particular, for all but finitely many surgeries, a pleated representative of S in M is disjoint from γ, and we obtain an estimate of the form

$$|\text{length}_H(\gamma)| \le \text{some function of } \text{scl}(l), p, q$$

valid for large p or q, which refines the inequality in Theorem 3.16.

3.4. Hyperbolic groups

We would like to generalize Theorem 3.9 and Theorem 3.11 beyond fundamental groups of hyperbolic manifolds to more general (word) hyperbolic groups. There are two essential ingredients in the proof of these theorems:

(1) the existence of a pleated surface representative in each homotopy class
(2) the existence of a Margulis constant in each dimension n

In fact, the proof of Theorem 3.11 only uses the existence of a Margulis constant in dimension 2, and the fact that in a given closed hyperbolic manifold there is a uniform positive lower bound on the translation length of any element.

We will see that both of these ingredients have acceptable generalizations to the context of hyperbolic groups, and therefore we obtain generalizations of these theorems with similar (geometric) proofs.

Alternatively, these theorems can be proved by explicitly constructing quasimorphisms with suitable properties and appealing to (the easy direction of) Bavard duality. The construction of quasimorphisms on hyperbolic groups extends to groups acting (weakly properly discontinuously) on hyperbolic spaces, such as mapping class groups, groups acting on trees, $\text{Out}(F_n)$ and so on, as we shall see in subsequent sections.

Where it pertains to quasimorphisms and stable commutator length, the material in the remainder of this chapter draws substantially on [**13, 12, 49**]. We also appeal to [**98, 24, 156**] for facts about hyperbolic spaces and groups, and [**21, 148**] for facts about the geometry of the curve complex.

3.4.1. Definitions and basic properties. Let G be a group with a finite symmetric generating set A. Let $C_A(G)$ be the Cayley graph of G with respect to A. In other words, $C_A(G)$ is the graph with one vertex for each element of G, and one edge from vertices g to g' for each pair of elements $g, g' \in G$ and each $a \in A$ for which $g' = ga$. We make $C_A(G)$ into a path metric space by declaring that the length of every edge is 1. The left action of G on itself extends to a simplicial (and therefore isometric) action of G on $C_A(G)$. Providing A contains no elements of order 2, the action is free and cocompact, with quotient a wedge of $|A|$ circles.

DEFINITION 3.23. A path metric space is *δ-hyperbolic* for some $\delta \geq 0$ if for every geodesic triangle abc, every point in the edge ab is contained in the union of the δ-neighborhoods of the other two edges:

$$ab \subset N_\delta(bc) \cup N_\delta(ca)$$

A group G with a finite symmetric generating set A is *δ-hyperbolic* if $C_A(G)$ is δ-hyperbolic as a path metric space.

G is *word-hyperbolic* (or simply *hyperbolic*) if there is a $\delta \geq 0$ and a finite symmetric generating set A for which $C_A(G)$ is δ-hyperbolic.

REMARK 3.24. Note that our definition of a δ-hyperbolic space requires it to be a path metric space; other definitions (e.g. in terms of the Gromov product) do not require this.

EXAMPLE 3.25. Finitely generated free groups are hyperbolic. Fundamental groups of compact surfaces with $\chi < 0$ are hyperbolic.

EXAMPLE 3.26. Let M be a closed Riemannian manifold with sectional curvature uniformly bounded above by a negative number. Then $\pi_1(M)$ is hyperbolic.

EXAMPLE 3.27. A group with a presentation satisfying the small cancellation condition $C(7)$ (see § 4.3) is hyperbolic.

EXAMPLE 3.28. A group $G = \langle X_m \mid R \rangle$ on a finite generating set X_m with a "random" set of relations R, drawn according to a suitable probability law (see [**163**]) is hyperbolic with probability 1.

In some sense, "most" groups are hyperbolic. On the other hand, many naturally occurring classes of groups (e.g. amenable groups, $\mathrm{SL}(n, \mathbb{Z})$ for $n \geq 3$, fundamental groups of cusped hyperbolic manifolds of dimension at least 3) are not hyperbolic. Nevertheless, hyperbolic groups are central in geometric group theory.

REMARK 3.29. If G is δ-hyperbolic with respect to a generating set A, there is an n depending only on δ so that if A_n denotes the set of elements in G of word length at most n, then G is 7-hyperbolic with respect to the generating set A_n. Hence δ may be taken to be some fixed small number at the expense of possibly increasing $|A|$. On the other hand, δ cannot be made arbitrarily small: a graph is 0-hyperbolic if and only if it is a tree. If $C_A(G)$ is a tree, then G is free, and A is a free generating set for G.

We assume the reader is familiar with basic elements of coarse geometry: (k, ϵ)-quasi-isometries, quasigeodesics, etc. We summarize some of the main properties of δ-hyperbolic spaces below (see [**98**] or [**24**] for details):

THEOREM 3.30 (Basic properties of hyperbolic spaces). *Let X be a δ-hyperbolic path metric space.*

(1) **Morse Lemma**. *For every k, ϵ there is a universal constant $C(\delta, k, \epsilon)$ such that every (k, ϵ)-quasigeodesic segment with endpoints $p, q \in X$ lies in the C-neighborhood of any geodesic joining p to q.*
(2) **Quasigeodesity is local**. *For every k, ϵ there is a universal constant $C(\delta, k, \epsilon)$ such that every map $\phi : \mathbb{R} \to X$ which restricts on each segment of length C to a (k, ϵ)-quasigeodesic is (globally) $(2k, 2\epsilon)$-quasigeodesic.*
(3) **Ideal boundary**. *There is an* ideal boundary ∂X *functorially associated to X, whose points consist of quasigeodesic rays up to the equivalence relation of being a finite Hausdorff distance apart. There is a natural topology on ∂X for which it is metrizable. If X is proper, ∂X is compact. Moreover, any quasi-isometric embedding $X \to Y$ between hyperbolic spaces induces a continuous map $\partial X \to \partial Y$.*

If G is hyperbolic, we denote the ideal boundary of its Cayley graph by ∂G. As a topological space, this does not depend on the choice of a generating set, so we call it the *ideal boundary* (or just the *boundary*) of G. The left action of G on itself induces an action of G on ∂G by homeomorphisms. Every element $g \in G$ is either finite order (i.e. is *elliptic*), or fixes two points $p^{\pm}$ in ∂G with "source-sink" dynamics (i.e is *hyperbolic*). That is, for any $q \in \partial G - p^{\pm}$ and any neighborhood U of p^+, the translate $g^n(q)$ lies in U for all sufficiently large positive n. The point p^+ is called the *attracting fixed point* of g, and p^- is called the *repelling fixed point.* Note that p^- is the attracting fixed point and p^+ the repelling fixed point for g^{-1}.

In fact, hyperbolic groups are completely characterized by the dynamics of their action on the boundary. The following characterization is due to Bowditch.

Theorem 3.31 (Bowditch, [**20**]). *Let M be a perfect metrizable compact Hausdorff space. Let G be a group acting faithfully on M by homeomorphisms. Let M^3 denote the space of* distinct *ordered triples of elements of M; i.e. the open subset of $M \times M \times M$ consisting of triples which are pairwise distinct. If the induced action of G on M^3 is properly discontinuous and cocompact, then G is hyperbolic, and there is a G-equivariant homeomorphism from M to ∂G.*

It is straightforward to show that a hyperbolic group acts on its boundary as in Theorem 3.31 and therefore this theorem gives a complete characterization of hyperbolic groups. If G is hyperbolic and ∂G contains more than two points, Klein's ping-pong argument applied to the action of G on ∂G shows that G contains many (quasi-isometrically embedded) nonabelian free groups of arbitrary finite rank. A hyperbolic group for which ∂G contains at most two points is said to be *elementary*; a group is elementary hyperbolic if and only if it is virtually cyclic.

Definition 3.32. If X is a metric space, and $g \in \mathrm{Isom}(X)$, the *translation length* of g, denoted $\tau(g)$, is the limit

$$\tau(g) = \lim_{n\to\infty} \frac{d_X(p, g^n(p))}{n}$$

where $p \in X$ is arbitrary.

The triangle inequality implies that this limit exists and is independent of p (and is therefore a conjugacy invariant). If X is a path metric space, and g fixes some geodesic l and acts on it as a translation, then $\tau(g) = d_X(q, g(q))$ for any $q \in l$. If G is a word-hyperbolic group and A is a generating set, then for any $g \in G$ the translation length $\tau_A(g)$, or just $\tau(g)$ if A is understood, is the translation length of g thought of as an element of $\mathrm{Isom}(C_A(G))$ under the natural left action of G on itself. Algebraically, $\tau(g) = \lim_{n\to\infty} \|g^n\|_A/n$ where $\|\cdot\|$ denotes word length with respect to the generating set A.

EXAMPLE 3.33. Let G be any group, and let S denote the set of commutators in G. Then the commutator subgroup $[G,G]$ acts on $C_S([G,G])$ by isometries, and for every $g \in [G,G]$ there is an equality $\mathrm{scl}(g) = \tau(g)$.

The following Lemma is an easy consequence of the local finiteness of $C_A(G)$, the fact that quasigeodesity is local, and the Morse Lemma.

LEMMA 3.34 (Axes in hyperbolic Cayley graphs). *Let G be δ-hyperbolic with respect to the generating set A. Then there is a positive constant $C(\delta,|A|)$ such that every $g \in G$ either has finite order, or there is some $n \le C$ such that g^n fixes some bi-infinite geodesic axis l_g and acts on it by translation.*

For a proof, see Theorem 5.1 from [**78**], or [**24**].

COROLLARY 3.35. *Let G be δ-hyperbolic with respect to the generating set A. Then there is a positive constant $C'(\delta,|A|)$ such that every $g \in G$ either has finite order, or satisfies $\tau(g) \ge C'$.*

PROOF. Since $C_A(G)$ is a graph in which every edge has length 1, elements of $\mathrm{Isom}(C_A(G))$ act on $C_A(G)$ simplicially. It follows that if an element $\gamma \in \mathrm{Isom}(C_A(G))$ acts on some geodesic l by translation, then $\tau(\gamma)$ is an integer. Now apply Lemma 3.34. $\square$

REMARK 3.36. The same argument shows that for a fixed hyperbolic group G, there is a constant $n(\delta,|A|)$ so that $\tau(g) \in \frac{1}{n}\mathbb{Z}$ for all $g \in G$.

3.4.2. Mineyev's flow space. The main difference between hyperbolic manifolds and Cayley groups of hyperbolic groups is *synchronous exponential convergence* of asymptotic geodesics. Two asymptotic geodesic rays in the hyperbolic plane have parameterizations by length such that the distance between corresponding points goes to 0 like e^{-t}. In a word-hyperbolic group, asymptotic geodesic rays eventually come within distance δ of each other, but may not get any closer. It is this synchronous exponential convergence which lets one estimate area from topology in hyperbolic surfaces, and it is crucial for our applications.

It is a fundamental insight due originally to Gromov that the geometry of a δ-hyperbolic space becomes much more tractable when one considers as primitive elements not *points*, but (bi-infinite) *geodesics*. Mineyev gave a precise codification of this insight, and constructed a geometric *flow space* associated to a δ-hyperbolic metric space, in which synchronous exponential convergence of asymptotic geodesics is restored.

A bi-infinite geodesic in a δ-hyperbolic space X contains two distinct geodesic rays, which are asymptotic to distinct points in ∂X. Conversely, if X is a *proper* metric space (i.e. the closed balls of any radius are compact) then any two distinct points in ∂X are the endpoints of some infinite geodesic.

We use the abbreviation $\partial^2 X$ to denote the space of ordered pairs of distinct points in ∂X:

$$\partial^2 X = \{(a,b) \in \partial X \times \partial X \text{ for which } a \neq b\}$$

Mineyev's flow space is not quite a metric space but rather a *pseudo-metric space*, i.e. a space together with a non-negative function $d(\cdot,\cdot)$ on pairs of points which satisfies all the axioms of a metric space except that $d(p,q)$ should be strictly positive for distinct points p and q. The reason is that Mineyev's space is a union (in a suitable sense) of *oriented* geodesics. Two geodesics with opposite orientation corresponding to the same (equivalence class of) geodesic in X cannot be distinguished by the distance function. However, there is a natural quotient of Mineyev's flow space in which these distinct oriented geodesics are identified, and the function d descends to a genuine metric on the quotient.

THEOREM 3.37 (Mineyev's flow space [**156**]). *Let X, d_X be a δ-hyperbolic graph whose vertices all have valence $\le n$. Then there exists a pseudo-metric space $\mathcal{F}(X), d^\times$ called the* flow space of X *with the following properties:*

(1) *$\mathcal{F}(X)$ is homeomorphic to $\partial^2 X \times \mathbb{R}$. The factors $(p,q,\cdot)$ under this homeomorphism are called the* flowlines.
(2) *There is an $\mathbb{R}$-action on $\mathcal{F}(X)$ (the* geodesic flow*) which acts as an isometric translation on each flowline $(p,q,\cdot)$.*
(3) *There is a $\mathbb{Z}/2\mathbb{Z}$ action $x \to x^*$ which anti-commutes with the $\mathbb{R}$ action, which satisfies $d^\times(x,x^*) = 0$, and which interchanges the flowlines $(p,q,\cdot)$ and $(q,p,\cdot)$.*
(4) *There is a natural action of* $\mathrm{Isom}(X)$ *on $\mathcal{F}(X)$ by isometries. If $g \in$* $\mathrm{Isom}(X)$ *is hyperbolic with fixed points $p^\pm$ in ∂X then g fixes the flowline $(p^-,p^+,\cdot)$ of $\mathcal{F}(X)$ and acts on it as a translation by a distance which we denote $\tau(g)$. This action of* $\mathrm{Isom}(X)$ *commutes with the $\mathbb{R}$ and $\mathbb{Z}/2\mathbb{Z}$ actions.*
(5) *There are constants $C \ge 0$ and $0 \le \lambda < 1$ such that for all triples $a,b,c \in \partial X$, there is a natural isometric parameterization of the flowlines $(a,c,\cdot),(b,c,\cdot)$ for which there is exponential convergence*
$$d^\times((a,c,t),(b,c,t)) \le C\lambda^t$$
Explicitly, $(a,c,0)$ is the point on $(a,c,\cdot)$ closest to b, and similarly for $(b,c,0)$ and a (as measured by suitable horofunctions).
(6) *If X admits a cocompact isometric action, then up to an additive error, there is an* $\mathrm{Isom}(X)$ *equivariant (k,ϵ) quasi-isometry between $\mathcal{F}(X), d^\times$ and X, d_X.*

Moreover, all constants as above depend only on δ and n.

This theorem conflates several results and constructions in [**156**]. The pseudo-metric $d^\times$ is defined in § 3.2 and § 8.6 on a slightly larger space which Mineyev calls the *symmetric join*. The flow space, defined in § 13, is a natural subset of this space. The basic properties of the $\mathbb{R}, \mathbb{Z}/2\mathbb{Z}$ and $\mathrm{Isom}(X)$ action are established in § 2. The remaining properties are subsets of Theorem 44 (p. 459) and Theorem 57 (p. 468).

There are a number of subtle details in the statement of this theorem, which require some discussion.

Bullet (3) implies that after quotienting $\mathcal{F}(X)$ by $\mathbb{Z}/2\mathbb{Z}$, the flowlines $(p,q,\cdot)$ and $(q,p,\cdot)$ become identified, and we can speak of *the* (unparameterized) geodesic joining p and q in the quotient which we denote $\overline{\mathcal{F}}(X)$. By abuse of notation, for each pair of distinct points $p,q \in \partial X$, let $(p,q,\cdot)$ denote a particular isometric parameterization of the unique geodesic in $\overline{\mathcal{F}}(X)$ joining p to q.

Bullet (5) is precisely the *synchronous* exponential convergence of flowlines which is achieved in hyperbolic space, but which is not achieved in hyperbolic groups. We refer to the special isometric synchronous parameterizations of asymptotic geodesics in this bullet as *nearest point parameterizations*. Note that nearest point parameterizations also make sense for ideal triangles in hyperbolic space, or hyperbolic space scaled to have constant curvature K for any $K < 0$. We define an *ideal triangle* in $\overline{\mathcal{F}}(X)$ to be the union of three (unparameterized) geodesics joining distinct $a, b, c \in \partial X$ in pairs.

The additive error in Bullet (6) spoiling genuine equivariance is necessary in case $\mathrm{Isom}(X)$ is indiscrete or does not act freely on X. If X is the Cayley graph of a torsion-free hyperbolic group G, then G acts freely on both $\mathcal{F}(X)$ and on X, and therefore the quasi-isometry can be chosen to be truly G-equivariant.

LEMMA 3.38. *Let Δ be an ideal triangle in $\overline{\mathcal{F}}(X)$. For each $K < 0$, let Δ_K be the edges of an ideal triangle in the complete simply-connected 2-manifold of constant curvature K. For each K, let ι be the map $\iota : \Delta_K \to \Delta$, unique up to permutation of vertices, which is an isometry on each edge, and which is compatible with the nearest point parameterizations. Then for suitable K depending only on n and δ, the map ι_K is Lipschitz, with Lipschitz constant depending only on n and δ.*

PROOF. Multiplying distances by $K^{-1/2}$ scales curvature by K. On an ideal triangle in $\mathbb{H}^2$, with the nearest point parameterization, there is an estimate

$$d((a,c,t),(b,c,t)) \le e^{-t}$$

So it suffices to make K big enough so that $e^{-|K|^{-1/2}} \le \lambda$. Since λ depends only on n and δ, so does K. Since C depends only on n and δ, so does the Lipschitz constant. □

3.4.3. Spectral gap theorem. With a suitably modified definition, we can construct pleated surfaces in $\overline{\mathcal{F}}(X)$ just as we did in hyperbolic manifolds.

DEFINITION 3.39. A pleated surface (possibly with boundary) in $\overline{\mathcal{F}}(X)$ consists of the following data:

(1) a hyperbolic surface S containing a geodesic lamination L whose complementary regions are all ideal triangles, and for which $\partial S \subset L$
(2) a homomorphism $\rho : \pi_1(S) \to \mathrm{Isom}(\overline{\mathcal{F}}(X))$
(3) if $\tilde{L}$ denotes the preimage of L in the universal cover $\tilde{S}$, a map $\iota : \tilde{L} \to \overline{\mathcal{F}}(X)$, equivariant with respect to the covering space action of $\pi_1(S)$ on $\tilde{L}$ and the action of $\pi_1(S)$ on $\overline{\mathcal{F}}(X)$ by ρ, which multiplies distances by a fixed constant on each edge, and is compatible with the nearest point parameterizations.

Notice that with this definition, the image of an element of $\pi_1(\partial S)$ under ρ has infinite order, and fixes two points in ∂X. Notice too that the map ι is not typically an isometry on leaves of $\tilde{L}$, but is rather an isometry after the metric on S has been scaled by some factor. The reason is so that we can insist that the map ι is *Lipschitz*, as in Lemma 3.38.

For a given $\rho : \pi_1(S) \to \mathrm{Isom}(\overline{\mathcal{F}}(X))$ it is by no means clear which laminations L on S are realized by pleated surfaces for some hyperbolic structure on S. However, if L is *proper* (i.e. every leaf accumulates only on the boundary) then the natural analogues of Lemma 3.6 and Lemma 3.7 are valid, with essentially the same proof, at least in the case where $\rho(\pi_1(S))$ does not contain any elliptic or parabolic elements. For the sake of simplicity therefore, we state our theorems below for torsion free hyperbolic groups.

LEMMA 3.40. *Suppose* $\mathrm{Isom}(X)$ *is torsion-free and cocompact, and* $\rho : \pi_1(S) \to \mathrm{Isom}(X)$ *is incompressible (i.e. injective on essential simple loops). Then there is a pleated surface in the sense of Definition 3.39 compatible with* ρ.

PROOF. We show how to choose a hyperbolic metric on S so that ι as in Definition 3.39 exists. Explicitly, choose K as in Lemma 3.38. We will construct a metric on S of constant curvature K; scaling this metric completes the proof.

Let $g \in \pi_1(S)$ be in the conjugacy class of the loop ∂S. If l is a geodesic whose ends spiral around ∂S, the ends of a lift $\tilde{l}$ are asymptotic to two fixed points of conjugates of g. Using ρ and equivariance, the images of these fixed points in ∂X are well-defined and distinct. As in Lemma 3.7, we can choose a proper full lamination L on S (i.e. one for which every complementary region is an ideal triangle, and each geodesic spirals around ∂S at both ends) for which the three points in $\partial\pi_1(S)$ associated to each ideal triangle are mapped to three distinct points in ∂X.

For each edge of $\tilde{L}$ there is a corresponding flowline of $\overline{\mathcal{F}}(X)$ we would like to map it to. If we fix an ideal triangle of constant curvature K, there is a unique map ι from its boundary to $\overline{\mathcal{F}}(X)$ which is isometric on each edge and compatible with the nearest point parameterizations at each of the three endpoints.

An edge in $\tilde{L}$ contained in two distinct triangles in $\tilde{S}$ inherits two different parameterizations; glue the corresponding ideal triangles in $\tilde{S}$ with a shear which is the difference of these two parameterizations. Then ι as defined on the two triangles is compatible on this edge. Since L is proper, the result of this gluing is connected, and determines a (scaled) hyperbolic structure on $\tilde{S}$ and a Lipschitz map $\iota : \tilde{L} \to \overline{\mathcal{F}}(X)$ which is an isometry on each edge. This construction is equivariant, and therefore the scaled hyperbolic structure on $\tilde{S}$ covers a scaled hyperbolic structure on S. $\square$

From this fact we can deduce analogues of Theorem 3.9 and Theorem 3.11.

THEOREM 3.41 (Calegari–Fujiwara [**49**], Thm. A). *Let G be a torsion-free group which is δ-hyperbolic with respect to a symmetric generating set $|A|$. Then there is a positive constant $C(\delta, |A|) > 0$ such that for all nontrivial $a \in G$ there is an inequality* $\mathrm{scl}(a) \geq C$.

PROOF. Let $a \in G$ be given. Let X denote the Cayley graph $C_A(G)$, and construct $\mathcal{F}(X)$ and $\overline{\mathcal{F}}(X)$. Let S be a surface of genus g with one boundary component, and $\rho : \pi_1(S) \to G$ a homomorphism taking the generator of $\pi_1(\partial S)$ to a^n. By Lemma 3.40, after reducing the genus of S if necessary, we can find a pleated surface (S, L) and $\iota : \tilde{L} \to \overline{\mathcal{F}}(X)$ with notation as in Definition 3.39. Let C be such that ι is C-Lipschitz.

As in the proof of Theorem 3.9, for any $\epsilon > 0$, we can find a component σ of $\partial S \cap S_{<\epsilon}$ of length at least

$$\mathrm{length}(\sigma) \geq \frac{\mathrm{length}(\partial S)}{12g - 6} - \frac{2\pi}{3\epsilon}$$

and a rectangular strip R of thickness $\leq \epsilon$ with σ on one side. For the sake of notation, and by analogy with Theorem 3.9, we define $\mathrm{length}(\gamma) = \mathrm{length}(\partial S)/n$.

If τ denotes the translation length of $\rho(a)$ on a flowline of $\overline{\mathcal{F}}(X)$, then $\tau \leq C \cdot \text{length}(\gamma)$. Assume $\text{length}(\sigma) > \text{length}(\gamma) + 4\epsilon$ and let p be one endpoint of σ. Let $\tilde{\sigma} \subset \tilde{L}$ be a lift of σ to $\tilde{S}$ and let $\tilde{p}$ be the corresponding lift of p. Let $a, bab^{-1} \in \pi_1(S)$ be as in the proof of Theorem 3.9. Then we have

$$d^{\times}(\iota(\tilde{p}), \rho(abab^{-1})(\iota(\tilde{p}))) \leq 4C\epsilon$$

This implies that the translation length of $\rho(abab^{-1})$ on $\overline{\mathcal{F}}(X)$ is at most $4C\epsilon$, and therefore, by bullet (6) of Theorem 3.37 the translation length of $\rho(abab^{-1})$ on X is at most $4Ck\epsilon$. On the other hand, by Corollary 3.35, since G is torsion free, there is a positive lower bound C' on the translation length of any nontrivial element of G. So if we choose ϵ so that $4Ck\epsilon < C'$ we can conclude that $\rho(bab^{-1}) = \rho(a)^{-1}$; which implies $\rho(b)$ has finite order in G, contrary to the hypothesis that G is torsion free.

This contradiction implies that $\text{length}(\sigma) \leq \text{length}(\gamma) + 4\epsilon$ and therefore

$$\text{length}(\gamma) \cdot \left(\frac{n}{12g - 6} - 1 \right) \leq 4\epsilon + \frac{2\pi}{3\epsilon}$$

On the other hand, since a is nontrivial, $Ck \cdot \text{length}(\gamma) \geq \tau(a) \geq C'$ (note that additive constants in quasi-isometries disappear when comparing translation lengths). Putting this together with our earlier estimate, and rearranging gives

$$\text{scl}(a) \geq \frac{1}{12} \left(\frac{C'}{C' + Ck \cdot \left(k\epsilon + \frac{2\pi}{3\epsilon}\right)} \right)$$

Finally, all constants which appear depend only on δ and $|A|$. □

THEOREM 3.42 (Calegari–Fujiwara [**49**], Thm. B). *Let G be a torsion-free nonelementary word hyperbolic group. Let $\delta_\infty(G)$ be the first accumulation point for stable commutator length on conjugacy classes in G. Then*

$$\frac{1}{12} \leq \delta_\infty(G) \leq \frac{1}{2}$$

PROOF. With setup and notation as in Theorem 3.41 we obtain the estimate

$$\text{length}(\gamma) \cdot \left(\frac{n}{12g - 6} - 1 \right) \leq 4\epsilon + \frac{2\pi}{3\epsilon}$$

If γ is sufficiently long, this implies $n/(12g - 6)$ is arbitrarily close to 1, so $\text{scl}(a)$ cannot be much smaller than $1/12$. This establishes the lower bound.

The upper bound follows exactly as in the proof of Theorem 3.11 by finding a quasi-isometrically embedded copy of F_2, the free group of rank 2, in G (which exists because G is nonelementary). □

3.5. Counting quasimorphisms

The geometric methods we have used to this point can be pushed only so far. The construction of Mineyev's flow space and the fine properties of its metric are very delicate and involved, and there are no realistic prospects of extending them more generally (e.g. to non-proper δ-hyperbolic spaces). Instead we turn to a generalization of Brooks' counting quasimorphisms (see § 2.3.2) due to Epstein–Fujiwara [**78**] for hyperbolic groups, and Fujiwara [**82**] in general.

3.5.1. Definition and properties. Let G be a group acting simplicially on a δ-hyperbolic complex X (not assumed to be locally finite).

DEFINITION 3.43. Let σ be a finite oriented simplicial path in X, and let σ^{-1} denote the same path with the opposite orientation. A *copy* of σ is a translate $a \cdot \sigma$ where $a \in G$.

If we fix a basepoint $p \in X$, then for any $a \in G$ there is a geodesic γ from p to $a(p)$. It is no good to try to define a counting function by counting (disjoint) copies of σ in γ, since γ is in general not unique. Instead, one considers a function which is sensitive to *all possible paths* from p to $a(p)$.

DEFINITION 3.44. Let σ be a finite oriented simplicial path in X, and let $p \in X$ be a base vertex. For any oriented simplicial path γ in X, let $|\gamma|_\sigma$ denote the maximal number of disjoint copies of σ contained in γ. Given $a \in G$, define

$$c_\sigma(a) = d(p, a(p)) - \inf_\gamma(\text{length}(\gamma) - |\gamma|_\sigma)$$

where the infimum is taken over *all* oriented simplicial paths γ in X from p to $a(p)$. Define the (small) *counting quasimorphism* h_σ by the formula

$$h_\sigma(a) = c_\sigma(a) - c_{\sigma^{-1}}(a)$$

Since length and $|\cdot|_\sigma$ take integer values on simplicial paths, the infimum of $\text{length}(\gamma) - |\gamma|_\sigma$ is *achieved* on some path γ. Any path with this property is called a *realizing path* for c_σ.

One may similarly define a "big" counting function C_σ which counts all copies of σ in each path γ, and a "big" counting quasimorphism H_σ. For the moment these are just names; we will show that h_σ is a quasimorphism, and estimate its defect in terms of δ.

If p, q are any two vertices in X, one can define

$$c_\sigma([p,q]) = d(p,q) - \inf_\gamma(\text{length}(\gamma) - |\gamma|_\sigma)$$

where the infimum is taken over all paths from p to q.

LEMMA 3.45. *One has the following elementary facts:*

(1) $c_\sigma([p,q]) = c_{\sigma^{-1}}([q,p])$
(2) $|c_\sigma([p,q]) - c_\sigma([p,q'])| \le d(q,q')$
(3) *If q is on a realizing path for σ from p to r, then*

$$c_\sigma([p,r]) \ge c_\sigma([p,q]) + c_\sigma([q,r]) \ge c_\sigma([p,r]) - 1$$

PROOF. Reversing a realizing path for c_σ gives a realizing path for $c_{\sigma^{-1}}$. A realizing path from p to q can be concatenated with a path of length $d(q,q')$ to produce some path from p to q', and *vice versa*. If q is on a realizing path from p to r, then it can intersect at most one copy of σ in that path. $\square$

In the sequel, we always assume that the length of σ is at least 2. It follows that $\text{length}(\gamma) - |\gamma|_\sigma \ge \text{length}(\gamma)/2$ for any path γ, and we obtain an *a priori* upper bound on $d(p, a(p)) - \text{length}(\gamma) + |\gamma|_\sigma$.

Realizing paths have the following universal geometric property:

LEMMA 3.46 (Fujiwara, Lemma 3.3 [**82**]). *Suppose* $\text{length}(\sigma) \ge 2$. *Any realizing path for c_σ is a $(2,4)$-quasigeodesic.*

PROOF. Let γ be a realizing path, and q, r points on γ. Let α be the subpath of γ from q to r, and let β be a geodesic with the same endpoints. Then β intersects at most two disjoint copies of σ in γ. Let γ' be obtained from γ by cutting out the subpath from q to r and replacing it with β. We have

$$|\gamma'|_\sigma \geq |\gamma|_\sigma - 2 - |\alpha|_\sigma \geq |\gamma|_\sigma - 2 - \text{length}(\alpha)/2$$

since each copy of σ in α has length at least 2 by assumption. On the other hand, since γ is a realizing path,

$$\text{length}(\gamma') - |\gamma'|_\sigma \geq \text{length}(\gamma) - |\gamma|_\sigma$$

Since $\text{length}(\gamma') - \text{length}(\gamma) = \text{length}(\beta) - \text{length}(\alpha)$, putting these estimates together gives

$$\text{length}(\beta) \geq \text{length}(\alpha)/2 - 2$$

□

REMARK 3.47. More generally, one can obtain better constants

$$K = \frac{\text{length}(\sigma)}{\text{length}(\sigma) - 1}, \ \epsilon = \frac{2 \cdot \text{length}(\sigma)}{\text{length}(\sigma) - 1}$$

which depend explicitly on the length of σ. The argument is essentially the same as that of Lemma 3.46.

By bullet (1) from Theorem 3.30 (i.e. the "Morse Lemma"), there is a constant $C(\delta)$ such that any realizing path for c_σ from p to $a(p)$ must be contained in the C-neighborhood of any geodesic between these two points. In particular, we have the following consequence:

LEMMA 3.48. *There is a constant $C(\delta)$ such that for any path σ in X of length at least 2, and for any $a \in G$, if the C-neighborhood of any geodesic from p to $a(p)$ does not contain a copy of σ, then $c_\sigma(a) = 0$.*

Finally, the defect of h_σ can be controlled *independently* of $\text{length}(\sigma)$:

LEMMA 3.49 (Fujiwara, Prop. 3.10 [**82**]). *Let σ be a path of length at least 2. Then there is a constant $C(\delta)$ such that $D(h_\sigma) \leq C$.*

PROOF. It is evident from the definitions that h_σ is antisymmetric, so it suffices to bound $|h_\sigma(a) + h_\sigma(b) + h_\sigma(b^{-1}a^{-1})|$. More generally, let p_1, p_2, p_3 be any three points in X. We will bound $|\sum_i h_\sigma([p_i, p_{i+1}])|$ where here and in the sequel, indices are taken mod 3.

Let α_i and α_i' be realizing paths for c_σ and $c_{\sigma^{-1}}$ respectively from p_i to p_{i+1}. By δ-thinness and Lemma 3.46, we can find points q_i, q_i' in each α_i, α_i' so that all 6 points are mutually within distance $N = N(\delta)$ of each other.

By definition, $|\sum_i h_\sigma([p_i, p_{i+1}])| = |\sum_i c_\sigma(\alpha_i) - c_{\sigma^{-1}}(\alpha_i')|$. By Lemma 3.45

$$c_\sigma(\alpha_i) \geq c_\sigma([p_i, q_i]) + c_\sigma([q_i, p_{i+1}]) \geq c_\sigma(\alpha_i) - 1$$

and

$$c_{\sigma^{-1}}(\alpha_i') \geq c_{\sigma^{-1}}([p_i, q_i']) + c_{\sigma^{-1}}([q_i', p_{i+1}]) \geq c_{\sigma^{-1}}(\alpha_i') - 1$$

By Lemma 3.45 again, $|c_\sigma([q_i, p_{i+1}]) - c_\sigma^{-1}([p_{i+1}, q_{i+1}'])| \leq N$ for each i, and therefore $D \leq 6 + 6N$ by the triangle inequality and the estimates above. □

3.5.2. Weak proper discontinuity. Lemma 3.49 is *not* by itself enough to deduce the existence of nontrivial quasimorphisms on a group G acting simplicially on a δ-hyperbolic complex X, as the following example shows.

EXAMPLE 3.50. Let $G = \mathrm{SL}(2, \mathbb{Z}[1/2])$. The ring $\mathbb{Z}[1/2]$ admits a discrete 2-adic valuation, with valuation ring $\mathbb{Z}$. Let $A = \mathrm{SL}(2, \mathbb{Z})$ thought of as a subgroup of G, and let B be the group of matrices of the form

$$B = \left\{ \begin{pmatrix} a & 2^{-1}b \\ 2c & d \end{pmatrix} \right\}$$

where $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is in $\mathrm{SL}(2, \mathbb{Z})$. The intersection $C = A \cap B$ is the subgroup of $\mathrm{SL}(2, \mathbb{Z})$ consisting of matrices of the form $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ where c is an even integer, and the group G is abstractly isomorphic to $A *_C B$.

There is a natural simplicial action of G on the Bass–Serre tree associated to its description as an amalgamated free product. Note that C has index 3 in both A and B, and therefore the Bass–Serre tree is regular and 3-valent. The action of G on this tree is simplicial and minimal. Nevertheless, $Q(G) = 0$, as follows from a Theorem of Liehl (see Example 5.38).

In order to show that h_σ is nontrivial, one wants to use Lemma 3.48. To apply this lemma, it is only necessary to find elements g such that if l is an axis for g on X, there are no translates of l^{-1} (i.e. l with the opposite orientation) which stay almost parallel to l on a scale large compared to the translation length of g, say of size $N\tau(g)$ where $\tau(g)$ is the translation length of g. Informally, such a pair of translates are said to be "anti-aligned". If p is in the midpoint of such a pair of anti-aligned axes l and $l' = h(l)$, the element $hg^nh^{-1}g^n$ translates $g^m(p)$ a uniformly short distance, for all n, m with $|n| + |m|$ small compared to N.

This discussion motivates the following definition, introduced in the paper [**13**] by Bestvina–Fujiwara:

DEFINITION 3.51. Suppose a group G acts simplicially on a δ-hyperbolic complex X. The action of an element $g \in G$ is *weakly properly discontinuous* if for every x and every $C > 0$ there is a constant $N > 0$ such that the set of elements $f \in G$ for which

$$d_X(x, fx) \le C \text{ and } d_X(g^N x, fg^N x) \le C$$

is finite.

To see how this addresses the issue of anti-aligned axes, suppose h_i are a sequence of elements for which $h_i(l)$ is anti-aligned with l (i.e. the axes are C apart with opposite orientation) on bigger and bigger segments centered at $x \in l$, and where C is as in Lemma 3.48. Then every f_i of the form $f_i = h_i g^n h_i^{-1} g^n$ satisfies $d_X(x, f_i x) \le C$ and $d_X(g^N x, f_i g^N x) \le C$ for any fixed N, n providing i is sufficiently big. If there are only finitely many distinct f_i, then for some i and for some distinct n, m we conclude

$$h_i g^n h_i^{-1} g^n = h_i g^m h_i^{-1} g^m$$

and therefore

$$h_i g^{n-m} h_i^{-1} = g^{m-n}$$

In other words, some nontrivial power of g is conjugate to its inverse.

Conversely, suppose no nontrivial power of g is conjugate to its inverse, and suppose that the action of g on X is weakly properly discontinuous. Let σ be a fundamental domain for the action of g on an axis, and let σ^N be a fundamental domain for g^N. Then for sufficiently big n, no translate of σ^{-N} is contained in any realizing path for g^n, and therefore the homogenization of h_{σ^N} is nontrivial on g.

REMARK 3.52. To make this discussion rigorous, one must replace "axis" throughout by "quasi-axis". This extension is routine and does not lead to any more substantial difficulties. See [**13**] for details.

3.5.3. Crossing number and growth in surface groups. We briefly mention a nontrivial application of counting quasimorphisms. Let S be a closed, orientable surface of genus at least 2.

DEFINITION 3.53. If $a \in \pi_1(S)$ is primitive, the *crossing number of* a, denoted $\mathrm{cr}(a)$, is the number of self-intersections of the geodesic representative of the free homotopy class of a in S. If $b = a^n$ then define $\mathrm{cr}(a^n) = n^2\mathrm{cr}(a)$.

Actually, crossing number is a somewhat subtle notion. For precise definitions, see § 4.2.1.

REMARK 3.54. Note that the function $\mathrm{cr}(\cdot)$ is characteristic (i.e. constant on orbits of $\mathrm{Aut}(G)$).

For each non-negative integer n let $S_n \subset \pi_1(S)$ denote the set consisting of elements with $\mathrm{cr}(a) \le n$. Note that S_n generates $\pi_1(S)$ for all $n \ge 0$. For each $a \in \pi_1(S)$, let $w_n(a)$ denote the word length of a in the generators S_n.

The following is the main theorem of [**44**]:

THEOREM 3.55 (Calegari [**44**], Thm. A). *Let S be a closed, orientable surface of genus at least 2. Then there are constants $C_1(S), C_2(S), C_3(S)$ such that for any non-negative integers n, m and any $a \in \pi_1(S)$ with $\mathrm{cr}(a) > 0$ there is an inequality*

$$w_n(a^m) \ge \frac{C_1 m}{\sqrt{n} + C_2} - C_3$$

A rough outline of the proof is as follows. Fix a finite generating set A for $\pi_1(S)$, and consider the Cayley graph $C_A(S)$. For each a, we build a counting quasimorphism h associated to a multiple of a which has an axis in $C_A(S)$. If $b \in \pi_1(S)$ satisfies $h(b) \ne 0$, then an axis for b contains a long segment which is close to the axis of a. This implies that the geodesic representative of b has a long segment which is close to the geodesic representative of a, and therefore b has a definite number of self-intersections. More precisely, if a realizing path for b contains p copies of a fundamental domain for the axis of a, the geodesic representative of b contains at least p^2 self-intersections. In particular, one obtains an estimate $|h(b)| \lesssim O(\sqrt{n}) + O(1)$. Since the defect of h is independent of a, b, the proof follows.

3.5.4. Separation theorem. If G is δ-hyperbolic with finite generating set $|A|$, the action of G on the Cayley graph $C_A(G)$ is properly discontinuous (and therefore certainly weakly properly discontinuous). It follows that there are many nontrivial counting quasimorphisms on G. In fact, one has the following theorem, which generalizes Theorem 3.41:

THEOREM 3.56 (Calegari–Fujiwara [**49**], Thm. A′). *Let G be a group which is δ-hyperbolic with respect to some symmetric generating set A. Let a be nontorsion, with no positive power conjugate to its inverse. Let $a_i \in G$ be a collection of* *elements with $T := \sup_i \tau(a_i)$ finite. Suppose that for all nonzero integers n, m and all $b \in G$ and indices i we have an inequality*

$$a_i^m \neq ba^n b^{-1}$$

Then there is a homogeneous quasimorphism $\phi \in Q(G)$ such that

(1) *$\phi(a) = 1$ and $\phi(a_i) = 0$ for all i*
(2) *The defect satisfies $D(\phi) \leq C(\delta, |A|)\left(\frac{T}{\tau(a)} + 1\right)$*

PROOF. By Lemma 3.34, after replacing each a_i by a fixed power whose size depends only on δ and $|A|$, we can assume that each a_i acts as translation on some geodesic axis l_i. Similarly, let l be a geodesic axis for a. Choose some big N (to be determined), and let σ be a fundamental domain for the action of a^N on l. The quasimorphism ϕ will be a multiple of the homogenization of h_σ, normalized to satisfy $\phi(a) = 1$. We need to show that if N is chosen sufficiently large, there are no copies of σ or σ^{-1} contained in the C-neighborhood of any l_i or l^{-1}, where C is as in Lemma 3.48.

Suppose for the sake of argument that there is such a copy, and let p be the midpoint of σ. The segment σ is contained in a translate $b(l)$. The translation length of a_i on l_i is $\tau(a_i) \leq T$, and the translation length of bab^{-1} on $b(l)$ is $\tau(a)$ (the case of l^{-1} is similar and is omitted). For big N, we can assume the length of σ is large compared to $\tau(a)$ and $\tau(a_i)$. Then for each n which is small compared to N, the element $w_n := a_i ba^n b^{-1} a_i^{-1} ba^{-n} b^{-1}$ satisfies $d(p, w_n(p)) \leq 4C$. Since there are less than $|A|^{4C}$ elements in the ball of radius $4C$ about any point, eventually we must have $w_n = w_m$ for distinct n, m. But this implies

$$a_i ba^n b^{-1} a_i^{-1} ba^{-n} b^{-1} = a_i ba^m b^{-1} a_i^{-1} ba^{-m} b^{-1}$$

and therefore a_i^{-1} and $ba^{n-m}b^{-1}$ commute. Since G is hyperbolic, commuting elements have powers which are equal, contrary to the hypothesis that no conjugate of a has a power equal to a power of a_i.

This contradiction implies that $\tau(a_i) + |A|^{4C}\tau(a) \geq N\tau(a)$. On the other hand, $D(h_\sigma)$ is uniformly bounded, by Lemma 3.49, and satisfies $h_\sigma(a^{Nn}) \geq n$. Homogenizing and scaling by the appropriate factor, we obtain the desired result. □

In fact, let $\sum n_i a_i$ be any integral chain which is nonzero in B_1^H. Without loss of generality, we may replace this chain by a rational chain with bounded denominators, with the same scl, and such that no distinct a_i, a_j have conjugate powers. After reordering, suppose $\tau(a_1) \geq \tau(a_i)$ for all i, and let ϕ be as in Theorem 3.56, so that $\phi(a_1) = 1$ and $\phi(a_i) = 0$ for $i \neq 1$. The defect $D(\phi)$ is bounded above by a constant depending only on δ and $|A|$. The coefficient of a_1 is bounded below by a positive constant depending only on δ and $|A|$. Hence by Bavard duality, $\text{scl}(\sum n_i a_i)$ is bounded below by a positive constant depending only on δ and $|A|$. In other words we have proved:

COROLLARY 3.57. *Let G be hyperbolic. Then* scl *is a norm on $B_1^H(G)$. Moreover, the value of* scl *on any nonzero integral chain in $B_1^H(G)$ is bounded below by a positive constant that depends only on δ and $|A|$.*

3.6. Mapping class groups

In this section we survey some of what is known about scl in mapping class groups. Our survey is very incomplete, since our main goal is to state an analogue of Theorem 3.41 for mapping class groups, and to give the idea of the proof.

DEFINITION 3.58. Let S be an oriented surface (possibly punctured). The *mapping class group* of S, denoted $\mathrm{MCG}(S)$, is the group of isotopy classes of orientation-preserving self-homeomorphisms of S.

EXAMPLE 3.59 (Dehn twist). Let γ be an essential simple curve in S. A *right-handed Dehn twist* in γ is the map $t_\gamma : S \to S$ supported on an annulus neighborhood $\gamma \times [0,1]$ which takes each curve $\gamma \times t$ to itself by a positive twist through a fraction t of its length. If the annulus is parameterized as $\mathbb{R}/\mathbb{Z} \times [0,1]$, then in co-ordinates, the map is given by $(\theta, t) \to (\theta + t, t)$.

By abuse of notation, we typically refer to both a specific homeomorphism and its image in $\mathrm{MCG}(S)$ as a Dehn twist.

REMARK 3.60. The inverse of a right-handed Dehn twist is a left-handed Dehn twist. Sometimes, right-handed Dehn twists are called *positive* Dehn twists. Notice that the handedness of a Dehn twist depends on an orientation for S but *not* on an orientation for γ.

The mapping class group is a fundamental object in 2-dimensional topology, and the literature on it is vast. Our treatment of it in this section is very cursory, and intended mainly just to introduce definitions and notation. For simplicity, we restrict attention throughout this section to mapping class groups of closed, orientable surfaces, although most of the results generalize to mapping class groups of surfaces with boundary or punctured surfaces. We refer the interested reader to [**16**] or [**80**] for background and details.

An element of $\mathrm{MCG}(S)$ induces an outer automorphism of $\pi_1(S)$ (outer, because homeomorphisms are not required to keep the basepoint fixed). This fact connects geometry with algebra. In fact, the connection is more intimate than it may appear at first glance, because of

THEOREM 3.61 (Dehn–Nielsen). *The natural map* $\mathrm{MCG}(S) \to \mathrm{Out}(\pi_1(S))$ *is an injection, with image equal to subgroup consisting of automorphisms which permute the peripheral subgroups.*

In particular, for S closed, $\mathrm{MCG}(S)$ is isomorphic to $\mathrm{Out}(\pi_1(S))$. For each S, the group $\mathrm{MCG}(S)$ is finitely presented. A finite generating set consisting of Dehn twists was first given by Dehn; a description of a finite set of relations (with respect to a slightly different generating set) was first given by Hatcher–Thurston [**106**].

The mapping class group of any closed orientable S is generated by finite order elements; in particular, its Abelianization is finite. If the genus of S is at least 3, the Abelianization is trivial:

THEOREM 3.62 (Powell, [**171**]). *Let S be a closed, orientable surface of genus at least* 3*. Then* $\mathrm{MCG}(S)$ *is perfect.*

A short proof of this theorem is due to Harer:

PROOF. It is well-known that $\mathrm{MCG}(S)$ is generated by Dehn twists about nonseparating curves (e.g. Lickorish's generating set [**134**]). By the classification of surfaces, any two nonseparating curves may be interchanged by a homeomorphism of S; it follows that $H_1(\mathrm{MCG}(S);\mathbb{Z})$ is generated by the image t of a twist about any nonseparating curve.

If the genus of S is at least 3, then S contains a non-separating four-holed sphere. The *lantern relation*, in MCG(4-holed sphere), says that the product of Dehn twists in the boundary components of a 4-holed sphere is equal to the product of twists in three curves in the sphere which separate the boundary components in pairs, and intersect each other in two points. The image of this relation in $H_1(\mathrm{MCG}(S);\mathbb{Z})$ is $t^4 = t^3$, so $\mathrm{MCG}(S)$ is perfect. □

3.6.1. Right-handed Dehn twists. Interesting lower bounds in scl can be obtained using gauge theory. This is a subject which has been pioneered by Kotschick, in [**130, 131**] and Endo–Kotschick [**73, 74**].

The following theorem is essentially due to Endo–Kotschick [**73**] although for technical reasons, the result is stated in that paper only for powers of a single separating Dehn twist. This technical assumption is removed in [**23**], and the result extended to products of positive twists in disjoint simple curves in [**131**].

THEOREM 3.63 (Endo–Kotschick [**73**], Kotschick [**131**]). *Let S be a closed orientable surface of genus $g \ge 2$. If $a \in \mathrm{MCG}(S)$ is the product of k right-handed Dehn twists along essential disjoint simple closed curves $\gamma_1, \cdots, \gamma_k$ then*

$$\mathrm{scl}(a) \ge \frac{k}{6(3g-1)}$$

It is beyond the scope of this survey to give a complete proof, but the way in which gauge theory enters the picture is the following. The product $a = t_{\gamma_1} t_{\gamma_2} \cdots t_{\gamma_k}$ lets one build a Lefschetz fibration E over the disk with fiber S which is singular over k distinct points, and such that the restriction of E to ∂D is a surface bundle with monodromy a. Over each singular point p_i, the fiber is a copy of S "pinched" along the curve γ_i, and such that the monodromy of a small loop around p_i is the twist t_{γ_i}. Since the curves γ_i are all disjoint, we can adjust the fiber structure on E so there is only *one* singular fiber, which degenerates along all the γ_i simultaneously. Since the twists are all right-handed, E admits a symplectic structure. Take an n-fold branched cover of the disk over the singular point, and pull back the fibration. After a suitable resolution, we get a new symplectic Lefschetz fibration E' over D with one singular fiber, such that the monodromy around the boundary is a^n, and such that the singular fiber has kn vanishing cycles, which come in parallel families of the γ_i.

An expression of a^n as a product of commutators in $\mathrm{MCG}(S)$ defines a nonsingular S bundle E'' over a once-punctured surface F, and by gluing E'' to E' along their boundaries in a fiber-respecting way, one obtains a closed symplectic manifold W. Then the engine of the proof is the well-known theorem of Taubes [**194**] in Seiberg–Witten theory which shows that for a minimal symplectic 4-manifold with $b_2^+ > 1$ the canonical class is represented by a symplectically embedded surface without spherical components. From this one derives inequalities on intersection numbers of certain surfaces in W and the result follows.

It is crucial in Theorem 3.63 that the twists in the different curves should all have the same handedness.

EXAMPLE 3.64 (Kotschick, Endo–Kotschick [**131, 74**]). Let α be an essential simple closed curve, and let $g \in \mathrm{MCG}(S)$ be such that $g(\alpha) \cap \alpha = \emptyset$, and $g(\alpha)$ is not isotopic to α. Let $h = t_\alpha t_{g(\alpha)}^{-1}$. Since α and $g(\alpha)$ are disjoint,

$$h^n = t_\alpha^n t_{g(\alpha)}^{-n} = t_\alpha^n g t_\alpha^{-n} g^{-1} = [t_\alpha^n, g]$$

so $\mathrm{scl}(h) = 0$. Note in this case that there is always some $f \in \mathrm{MCG}(S)$ which interchanges α and $g(\alpha)$. For such an f we have $fhf^{-1} = h^{-1}$, so that $h = 0$ in B_1^H.

As another example, let α, β, γ be disjoint nonseparating non-isotopic simple closed curves, and define $h = t_\alpha^{-1} t_\beta^{-1} t_\gamma^2$. If g interchanges α and γ, and g' interchanges β and γ, then $h^k = [t_\gamma^k, g][t_\gamma^k, g']$. In this case, all powers of h are in distinct conjugacy classes, and h is not in B_1^H. This example shows that H is not closed in $B_1(\mathrm{MCG}(S))$.

Interesting upper bounds on scl can be obtained by explicit examples.

EXAMPLE 3.65 (Korkmaz [**129**]). Let $a \in \mathrm{MCG}(S)$ be a Dehn twist in a nonseparating closed curve. Then a^{10} can be written as a product of two commutators.

Let $a_1, \cdots, a_5$ be curves on S as in Figure 3.4, where a_4, a_5 are nonseparating. For each i, let t_i denote a positive Dehn twist in a_i. Notice that t_1, t_3, t_4, t_5 all

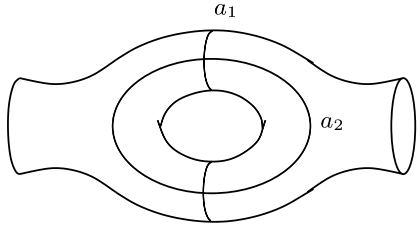

FIGURE 3.4. The curves $a_1, \cdots, a_5$ in S

commute. Moreover, a neighborhood of $a_1 \cup a_2$ is a once-punctured torus. An element of the mapping class group of a punctured torus is determined by its action on homology, and one may verify that the relation $t_1 t_2 t_1 = t_2 t_1 t_2$ holds by an elementary calculation, and similarly with t_3 in place of t_1. The relation $t_4 t_5 = (t_1 t_2 t_3)^4$ is a little harder to see, but still elementary.

Following [**129**], we calculate

$$\begin{aligned} t_4 t_5 &= (t_1 t_2 t_3)(t_1 t_2 t_3)(t_1 t_2 t_3)(t_1 t_2 t_3) \\ &= (t_1 t_2 t_1)(t_3 t_2 t_3)(t_1 t_2 t_1)(t_3 t_2 t_3) \\ &= (t_2 t_1 t_2)(t_2 t_3 t_2)(t_2 t_1 t_2)(t_2 t_3 t_2) \end{aligned}$$

Since t_2 commutes with both t_4 and t_5, this gives

$$t_4 t_5 = t_1 (t_2^2 t_3 t_2^{-2}) t_2^4 t_1 t_2^{-1} (t_2^3 t_3 t_2^{-3}) t_2^{-1} t_2^6$$

If $\alpha = t_2^2(a_3)$ and $\beta = t_2^3(a_3)$ then this yields

$$(t_4 t_\alpha^{-1} t_5 t_1^{-1}) = t_2^4 (t_1 t_2^{-1} t_\beta t_2^{-1}) t_2^6$$

Each bracketed expression is of the form $t_a t_b^{-1} t_c t_d^{-1}$ for simple curves a, b, c, d. It can be verified in each case that the curves $a \cup b$ lie in S in the same combinatorial pattern as $d \cup c$. Therefore there is $g \in \mathrm{MCG}(S)$ for which $g(a) = d$ and $g(b) = c$. But this means

$$t_a t_b^{-1} t_c t_d^{-1} = t_a t_b^{-1} g t_b t_a^{-1} g^{-1} = [t_a t_b^{-1}, g]$$

That is, each of the two bracketed expressions are commutators, and the proof follows.

In case $\mathrm{genus}(S) = 2$, one obtains

$$\frac{1}{30} \le \mathrm{scl}(t) \le \frac{3}{20}$$

where the first inequality comes from Theorem 3.63.

Let t denote a Dehn twist in a nonseparating curve (for concreteness). One may ask to what extent $\mathrm{scl}(t)$ depends on $\mathrm{genus}(S)$. In fact, it turns out that Theorem 3.63 gives the correct order of magnitude. This follows from a general phenomenon, especially endemic in transformation groups, which we describe.

Example 3.66 (Münchhausen trick). Suppose we are given a group G acting on a set Y. Suppose further that there is an identity $a = \prod_{i=1}^n [b_i, c_i]$ in G where a, b_i, c_i all have support in some subset $X \subset Y$. Suppose finally that there is $g \in G$ such that $X \cap g^i(X) = \emptyset$ for $0 < i \le m$.

If H is a subgroup of G consisting of elements with support in X, define $\Delta : H \to G$ by

$$\Delta(h) = h h^g h^{g^2} \cdots h^{g^{m-1}}$$

where the superscript notation denotes conjugation. The condition on g ensures Δ is a homomorphism, and therefore $\mathrm{cl}(\Delta(a)) \le n$. Now define an element $j = a(a^2)^g (a^3)^{g^2} \cdots (a^m)^{g^{m-1}}$. We have the identity $[j, g] = \Delta(a)(a^{-m})^{g^m}$, which exhibits a^m as a product of at most $n+1$ commutators. If m is large compared to n, then $\mathrm{scl}(a)$ is small.

Corollary 3.67 (Kotschick). *If t is a Dehn twist in a non-separating curve, there is an estimate* $\mathrm{scl}(t) = O(1/g)$.

Proof. The lower bound is Theorem 3.63. The upper bound follows by exhibiting t as a product of commutators of elements b_i, c_i supported in some fixed surface T with boundary, which can be included into a surface S of arbitrary genus. Then apply the Münchhausen trick. $\square$

Remark 3.68. Many variations on Corollary 3.67 are proved in [**132**] and [**29**], and the same trick appears in the proof of Theorem 5.13. The trick works whenever there is "enough room" in Y for many disjoint copies of X; in many important applications, X is (in some sense) a copy of Y. The terminology "Münchhausen trick" is taken from [**118**], and "refers to the story about how the legendary baron allegedly succeeded in pulling himself out of a quagmire by his own hair". This trick goes back at least to [**81**] (in fact one could argue it goes back to Zeno of Elea).

3.6.2. The complex of curves. For most surfaces S, the group $\mathrm{MCG}(S)$ is not word-hyperbolic. Nevertheless, it acts naturally on a certain δ-hyperbolic simplicial complex, called the *complex of curves*. This complex was first introduced by Harvey [**103**], but it was Masur–Minsky [**148**] who established some of its most important basic properties. A similar complex was also introduced by Hatcher–Thurston [**106**]. A good introductory reference to the complex of curves is [**182**].

DEFINITION 3.69 (Harvey [**103**]). Let S be a closed, orientable surface of genus at least 2. The *complex of curves*, denoted $\mathcal{C}(S)$, is the simplicial complex whose k simplices consist of isotopy classes of pairwise disjoint non-parallel essential simple closed curves on S.

With this definition, $\mathcal{C}(S)$ is a simplicial complex of dimension $3g-4$. The natural permutation action of $\mathrm{MCG}(S)$ on the set of isotopy classes of essential simple curves on S induces a simplicial action of $\mathrm{MCG}(S)$ on $\mathcal{C}(S)$.

REMARK 3.70. Similar definitions can be made when S has smaller genus, or has punctures or boundary components. See [**148**].

We can think of $\mathcal{C}(S)$ as a metric space, by taking every edge to have length 1 and every simplex to be equilateral. In the sequel, we are typically interested not in $\mathcal{C}(S)$ itself, but in its 1-skeleton. Usually, by abuse of notation, when we talk about $\mathcal{C}(S)$ we really mean its 1-skeleton. It should be clear from context which sense is meant in each case.

The main property of $\mathcal{C}(S)$ from our point of view is the following theorem:

THEOREM 3.71 (Masur–Minsky [**148**]). *Let S be as above. Then $\mathcal{C}(S)$ is δ-hyperbolic for some $\delta(S)$.*

An element $a \in \mathrm{MCG}(S)$ is *reducible* if it permutes some finite set of isotopy classes of disjoint, non-parallel essential simple closed curves. It turns out that an element $a \in \mathrm{MCG}(S)$ has a finite orbit in $\mathcal{C}(S)$ if and only if a is either finite order or reducible. An element which is neither finite order nor reducible is said to be *pseudo-Anosov*.

THEOREM 3.72 (Masur–Minsky [**148**]). *Let $a \in \mathrm{MCG}(S)$ be pseudo-Anosov. Then every orbit of a on $\mathcal{C}(S)$ is a quasigeodesic.*

In particular, every pseudo-Anosov element has a positive translation length $\tau(a)$. In fact, Bowditch [**21**] proves the following analogue of Lemma 3.34:

THEOREM 3.73 (Bowditch, Theorem 1.4 [**21**]). *Let S be a closed, orientable surface of genus at least 2. Then there is a constant $C(S)$ such that for every pseudo-Anosov $a \in \mathrm{MCG}(S)$, there is $n \le C$ such that a^n fixes some bi-infinite geodesic axis l_a and acts on it by translation.*

3.6.3. Acylindricity. The action of $\mathrm{MCG}(S)$ on $\mathcal{C}(S)$ is not proper; the stabilizer of a vertex is isomorphic to a copy of $\mathrm{MCG}(S')$ for some smaller surface S'. Nevertheless, Bestvina–Fujiwara ([**13**]) show that every pseudo-Anosov element of $\mathrm{MCG}(S)$ acts weakly properly discontinuously on $\mathcal{C}(S)$. As a corollary, they deduce the following theorem:

THEOREM 3.74 (Bestvina–Fujiwara, [**13**], Theorem 12). *Let G be a subgroup of $\mathrm{MCG}(S)$ which is not virtually Abelian. Then the dimension of $Q(G)$ is infinite.*

In particular, if ϕ is pseudo-Anosov, and p, q are sufficiently far apart on an axis for ϕ, only finitely many elements of $\mathrm{MCG}(S)$ move both p and q a bounded distance. This is enough to show that every pseudo-Anosov element either has a (bounded) power conjugate to its inverse, or has positive scl. To obtain *uniform* estimates on scl, one needs a slightly stronger statement, captured in the following theorem of Bowditch:

THEOREM 3.75 (Bowditch, Acylindricity Theorem [**21**]). *Let S be a closed orientable surface of genus $g \geq 2$. For any $t > 0$ there exist positive constants $C_1(t, S), C_2(t, S)$ such that given any two points $x, y \in \mathcal{C}(S)$ with $d(x, y) \geq C_1$ there are at most C_2 elements $a \in \mathrm{MCG}(S)$ such that $d(x, ax) \leq t$ and $d(y, ay) \leq t$.*

REMARK 3.76. A similar theorem is also proved by Masur–Minsky [**148**].

We are now in a position to state the analogue of Theorem 3.41 and Theorem 3.56 for mapping class groups.

THEOREM 3.77 (Calegari–Fujiwara [**49**], Thm. C). *Let S be a closed orientable surface of genus at least 2. Then there are constants $C_1(S), C_2(S) > 0$ such that for any pseudo-Anosov element $a \in \mathrm{MCG}(S)$ either there is a positive integer $n \leq C_1$ for which a^n is conjugate to its inverse, or else there is a homogeneous quasimorphism $\phi \in Q(\mathrm{MCG}(S))$ with $\phi(a) = 1$ and $D(\phi) \leq C_2$.*

Moreover, suppose $a_i \in \mathrm{MCG}(S)$ are a (possibly infinite) collection of elements with $T := \sup_i \tau(a_i)$ finite. Suppose that for all nonzero integers n, m and all $b \in \mathrm{MCG}(S)$ and indices i we have an inequality

$$a_i^m \neq b a^n b^{-1}$$

Then there is a homogeneous quasimorphism $\phi \in Q(\mathrm{MCG}(S))$ such that

(1) *$\phi(a) = 1$ and $\phi(a_i) = 0$ for all i*
(2) *The defect satisfies $D(\phi) \leq C_2(S)\left(\frac{T}{\tau(a)} + 1\right)$*

PROOF. The proof is essentially the same as the proof of Theorem 3.56, with Theorem 3.73 used in place of Lemma 3.34. After replacing a and a_i by (bounded) powers, one assumes that they stabilize axes l and l_i respectively. For each n, let $w_n := a_i b a^n b^{-1} a_i^{-1} b a^{-n} b^{-1}$. If $b(l)$ is close to l_i on a segment σ which is long compared to $\tau(a), \tau(a_i)$ and C_1 (as in Theorem 3.75), then one can find points p and p' on l_i with $d(p, p') \geq C_1$ such that $d(p, w_n(p)) \leq t$ and $d(p', w_n(p')) \leq t$ for all n small compared to length(σ).

One needs to know that two pseudo-Anosov elements in $\mathrm{MCG}(S)$ which commute have powers which are proportional (the pseudo-Anosov hypothesis cannot be omitted here); see e.g [**198**]. Otherwise, the remainder of the proof is copied verbatim from the proof of Theorem 3.56. □

REMARK 3.78. In contrast with the case of word-hyperbolic groups, it should be noted that there are infinitely many conjugacy classes of pseudo-Anosov elements in $\mathrm{MCG}(S)$ with bounded translation length. In fact, the first accumulation point for translation length in $\mathrm{MCG}(S)$ is $O(1/g \log(g))$, where g is the genus of S; see Theorem 1.5 of [**79**].

The separation property of the quasimorphisms produced by Theorem 3.77 is very powerful, and has a number of consequences, including the following.

COROLLARY 3.79. *Let Σ be a subset of $\mathrm{MCG}(S)$ consisting only of reducible elements, and let G be the subgroup it generates. Suppose G contains a pseudo-Anosov element a with no power conjugate to its inverse. Then the Cayley graph of G with respect to the generating set Σ has infinite diameter.*

PROOF. By Theorem 3.77 there is a homogeneous quasimorphism ϕ defined on $\mathrm{MCG}(S)$ with $\phi(a^n) = n$ which vanishes on Σ. If b is an element of G with length at most m in the generators Σ, then $\phi(b) \leq (m - 1)D(\phi)$. □

The hypotheses of this Corollary are satisfied whenever G is not reducible or virtually cyclic.

Example 3.80 (Broaddus–Farb–Putman [**25**]). The *Torelli group*, denoted $\mathcal{I}(S)$, is the kernel of the natural map $\mathrm{MCG}(S) \to \mathrm{Aut}(H_1(S)) = \mathrm{Sp}(2g, \mathbb{Z})$. It is not a perfect group; the kernel of the map

$$\mathcal{I}(S) \to H_1(\mathcal{I}(S); \mathbb{Z})/\text{torsion}$$

is denoted $\mathcal{K}(S)$, and is generated by Dehn twists about separating simple closed curves. This is an infinite (in fact, characteristic) generating set. On the other hand, by Corollary 3.79 the diameter of the Cayley graph of $\mathcal{K}$ with respect to this generating set is infinite.

3.7. Out(F_n)

Very recently, the methods discussed in this chapter have been used to construct many nontrivial quasimorphisms on $\mathrm{Out}(F_n)$, the group of outer automorphisms of a free group. The main results described in § 3.7.2 were announced by Hamenstädt in May 2008, and first appeared in (pre-)print in Bestvina–Feighn [**12**]. In what follows we restrict ourselves to describing the construction of suitable δ-hyperbolic simplicial complexes on which $\mathrm{Out}(F_n)$ acts, and summarizing the important properties of these complexes and the action without justification.

3.7.1. Outer space.

In what follows, $\mathrm{Out}(F_n)$ denotes the outer automorphism group of the free group F_n of rank $n \geq 2$. The modern theory of $\mathrm{Out}(F_n)$ is dominated by several deep analogies between this group and mapping class groups. The cornerstone of these analogies is Culler–Vogtmann's construction [**61**] of *Outer space*, which serves as an analogue of Teichmüller space.

Definition 3.81. Fix F_n, a free group of rank n. An action of F_n on an $\mathbb{R}$-tree T is *minimal* if there is no proper F_n-invariant subtree of T. Let $\rho : F_n \to \mathrm{Isom}(T)$ be an action which is minimal, free and discrete. Associated to any such ρ there is a *length function* $\ell_\rho \in \mathbb{R}^{F_n}$ where $\ell_\rho(g)$ is the translation length of $\rho(g)$ on T.

Outer space, denoted in the sequel $\mathcal{PT}$, is the projectivization of the space of length functions of minimal, free, discrete actions of F_n on $\mathbb{R}$-trees, with the weak topology. Its compactification $\overline{\mathcal{PT}}$ is obtained by adding weak limits of projective classes of length functions.

If $\rho : F_n \to \mathrm{Isom}(T)$ is a minimal, free, discrete action of F_n on an $\mathbb{R}$-tree, and $\varphi : F_n \to F_n$ is an automorphism, then $\rho \circ \varphi^{-1} : F_n \to \mathrm{Isom}(T)$ is another action. If φ is inner, the length functions ℓ_ρ and $\ell_{\rho\circ\varphi^{-1}}$ are equal. Hence the group $\mathrm{Out}(F_n)$ acts in a natural way on $\mathcal{PT}$, and this action extends continuously to its compactification.

Outer space has a natural cellular structure, which can be described as follows. For each action $\rho : F_n \to \mathrm{Isom}(T)$, let Γ_ρ be the quotient of T by $\rho(F_n)$, thought of as a metric graph together with an isomorphism of its fundamental group with F_n (i.e. a *marking*), which is well-defined up to conjugacy. The cells of $\mathcal{PT}$ are the actions which correspond to a fixed combinatorial type of Γ_ρ, together with a choice of marking. This cellular structure extends naturally to $\overline{\mathcal{PT}}$.

3.7.2. Fully irreducible automorphisms.

DEFINITION 3.82. An element $\varphi \in \mathrm{Out}(F_n)$ is *fully irreducible* if for all proper free factors F of F_n and all $k > 0$ the subgroup $\varphi^k(F)$ is not conjugate to F.

The main result of [**12**] is as follows:

THEOREM 3.83 (Bestvina–Feighn [**12**], p.11). *For any finite set $\varphi_1, \cdots, \varphi_k$ of fully irreducible elements of $\mathrm{Out}(F_n)$ there is a connected δ-hyperbolic graph X (depending on the φ_i) together with an isometric action of $\mathrm{Out}(F_n)$ on X such that*

1. *the stabilizer in $\mathrm{Out}(F_n)$ of a simplicial tree in $\overline{\mathcal{PT}}$ has bounded orbits*
2. *the stabilizer in $\mathrm{Out}(F_n)$ of a proper free factor $F \subset F_n$ has bounded orbits*
3. *the φ_i all have nonzero translation lengths*

The construction of the graph X is somewhat complicated, and follows a template developed by Bowditch [**20**] to study convergence group actions. A fully irreducible automorphism ψ has one stable and one unstable fixed point in the boundary of $\overline{\mathcal{PT}}$, which we denote $T^{\pm}_{\psi}$. A tree T is *irreducible* if it is of the form T^+_{ψ} for some fully irreducible ψ.

Choose sufficiently small closed neighborhoods $D^{\pm}_i$ of $T^{\pm}_{\varphi_i}$. In $\overline{\mathcal{PT}}$, let $\mathcal{M}$ be the subspace of *all* irreducible trees. Define an *annulus* to be an ordered pair of closed subsets of $\mathcal{M}$ either of the form $(\psi(D^-_i)\cap\mathcal{M}, \psi(D^+_i)\cap\mathcal{M})$ or $(\psi(D^+_i)\cap\mathcal{M}, \psi(D^-_i)\cap\mathcal{M})$, where $\psi \in \mathrm{Out}(F_n)$ and $D^{\pm}_i$ are as above. Denote the set of annuli (defined as above) by $\mathcal{A}$. The pair $(\mathcal{M}, \mathcal{A})$ depends on the choice of the φ_i, and both $\mathcal{M}$ and $\mathcal{A}$ admit natural actions by $\mathrm{Out}(F_n)$.

For any subset $K \subset \mathcal{M}$ and any annulus $A = (A^-, A^+)$ write $K < A$ if $K \subset \mathrm{int}A^-$, and write $A < K$ if $K \subset \mathrm{int}A^+$. If $A = (A^-, A^+)$ and $B = (B^-, B^+)$ are two annuli, write $A < B$ if $\mathrm{int}A^+ \cup \mathrm{int}B^- = \mathcal{M}$. Then for any pair of subsets K, L of $\mathcal{M}$, define $(K|L) \in [0, \infty]$ to be the biggest number of annuli A_i in $\mathcal{A}$ such that

$$K < A_1 < A_2 < \cdots < A_n < L$$

Let $\mathcal{Q}$ denote the set of ordered triples of distinct points in $\mathcal{M}$. If $A = (a_1, a_2, a_3)$ and $B = (b_1, b_2, b_3)$ are elements of $\mathcal{Q}$, then define

$$\rho(A, B) = \max(\{a_i, a_j\}|\{b_k, b_l\})$$

where the maximum is taken over all $i \neq j$ and $k \neq l$. The *Bowditch complex* of the pair $(\mathcal{M}, \mathcal{A})$, is the graph whose vertices are the elements of $\mathcal{Q}$ and whose edges are the pairs of elements $A, B \in \mathcal{Q}$ with $\rho(A, B) \leq r$ for some sufficiently big r. Bowditch [**20**] gives certain axioms for an abstract pair $(\mathcal{M}, \mathcal{A})$ which ensure that the associated Bowditch complex is δ-hyperbolic. The substance of Theorem 3.83 is the proof that $(\mathcal{M}, \mathcal{A})$ as above satisfies Bowditch's axioms.

In order to construct quasimorphisms, one must also know that many elements of $\mathrm{Out}(F_n)$ act weakly properly discontinuously. This is the following proposition, also from [**12**]:

PROPOSITION 3.84 (Bestvina–Feighn [**12**], p.24). *For φ_i and X as in Theorem 3.83, the action of each φ_i on X is weakly properly discontinuous; i.e. for every $x \in X$ and every $C > 0$ there is a constant $N > 0$ such that the set of $\psi \in \mathrm{Out}(F_n)$ for which*

$$d_X(x, \psi x) \leq C \text{ and } d_X(\varphi_i^N x, \psi\varphi_i^N x) \leq C$$

is finite.

Theorem 3.83 allows one to construct many quasimorphisms on $\mathrm{Out}(F_n)$ by the method of § 3.5. Proposition 3.84 implies that these quasimorphisms are nontrivial and independent. Consequently, one concludes that $Q(\mathrm{Out}(F_n))$ is infinite dimensional; in fact ([**12**] Corollary 4.28), for any subgroup Γ of $\mathrm{Out}(F_n)$ which contains two independent fully irreducible automorphisms, $Q(\Gamma)$ is infinite dimensional.